

SAS ユーザー総会 2025 論文集

会場：東京大学・弥生講堂

会期：2025 年 9 月 24 日(水) ～25 日(木)

主催：SAS ユーザー会 世話人会

目次

論文

一般線型モデルとGLMプロシジャの基礎
田中 真史（イーピーエス株式会社） -------10

SASの中でRを動かす
関根 暁史（藤本製薬株式会社）-------28

世界初の科学技術で重要なLSDの判別理論―統計や生成AIでできない組み合わせ最適化LDF―
新村秀一[1]、新村秀樹[2]、市川均[3]（[1]成蹊大学名誉教授、[2]イノベストメント(株)、[3]有限会社マイス(LINDO Japan 代表) -------37

RTFファイルのページレイアウトを制御するマクロとコンペアマクロ
平野 勝也 株式会社インテリム データサイエンスユニット） -------49

ROBUSTREGプロシジャによるロバスト回帰とその経時測定データ解析への応用
伊庭 克拓、浅野 豊、鈴木 聡、毛利 誠、横田 大輔 （大塚製薬株式会社 新薬開発本部 バイオメトリックス部 統計解析室） -------62

兵庫県下の消防本部における救急搬送データのJMPを活用した分析
川向 肇 [1]・野田晴菜[2]（[1] 兵庫県立大学社会情報科学部・大学院情報科学研究科・ソーシャルデータサイエンス研究所, [2] 兵庫県立大学社会情報科学部） -------80

SASによる症例数再推定を伴うアダプティブデザインの実装と統計的評価～CPZ（Constrained Promising Zone）デザインの実践～
内藤 寿輝，森田 祐介，太田 裕二，浜田 泉（ノーベルファーマ株式会社 データサイエンス部） -------91

非臨床統計解析ソフトウェアEXSUSとStatPreclinicaの比較検討
本田主税（小野薬品工業株式会社 研究プロジェクト統括部 研究推進部、 ファーマコヴィジランス統括部 PV システム部 PV データ戦略室） -------110

階層型クラスタリングを活用したヒートマップのSAS実装
高田 浩成 （イーピーエス株式会社） -------120

ベイズ流試験デザインにおけるシミュレーションに基づく動作特性のポイントとその具体例
吉本 拓矢 [1,3]，田中 大樹 [2,3] （[1] 中外製薬株式会社バイオメトリクス部，[2] 鳥居薬品株式会社調査役，[3] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会） -------138

目次

MCMCプロシジャによるcommensurateprior法を用いた外部対照データの動的借用法
吉本 拓矢 [1,4]，田中 優一 [2,4]，高戸 俊一 [3,4] （[1] 中外製薬株式会社 バイオメトリクス部，[2] 大鵬薬品株式会社 データサイエンス部，[3] 杏林製薬株式会社 開発推進部，[4] 日本製薬工業協会医薬品評価委員会 データサイエンス部会） -------152

Q-Learningに基づく最適な動的治療レジメの推定
徳田芳稀 [1] （[1] エイツーヘルスケア株式会社 データサイエンスソリューション部） ------170

対数正規分布に従う変数に対するBland-AltmanPlotにおけるLimitsofAgreementの描画
徳田芳稀、田村沙弥（エイツーヘルスケア株式会社 データサイエンスソリューション部）
-----188

NLPプロシージャを用いた入れ子型ロジットモデルにおけるパラメータ推定に関する一考察
坂巻英一（国立大学法人宇都宮大学） -------201

SAS周辺で使用される文字コードの取り扱いについて
兵頭 照将（イーピーエス株式会社） -------210

SASプログラムの可能性を広げるFILENAMEPIPEオプション-Linuxコマンド連携による拡張性の向上とSASログでの実行管理
三木 悠吾（グラクソ・スミスクライン株式会社） ------217

ベイズ流臨床試験デザインの頻度論的較正：事前情報と中間解析の影響
三島遼（京都府立医科大学附属病院） -------226

線形モデルのプロシジャ選択を考える
大山暁史（イーピーエス株式会社） -------236

用量漸増デザインにおけるBHLRM+EXNEXの動作特性の評価
吉田和正 [1,4]，石田弘輔 [2,4]，相澤愛奈 [3,4] （[1] 日本ベーリンガーインゲルハイム株式会社，[2] ノバルティス ファーマ株式会社，[3] ファイザーR&D 合同会社，[4] 日本製薬工業協会医薬品評価委員会 データサイエンス部会） -------251

IMLプロシジャで理解するCox回帰
堀本隆一 （株式会社エス・シー・エー） -------265

超多時点のラテン方格 QT 試験データに関する統計解析の新たな方向性－ICH-E14/S7B Q&A（2022）への対応－
高橋 行雄 （BioStat 研究所（株）） -------277

目次

Wilcoxon-Mann-Whitney（WMW）検定の問題点とBrunnerMunzel検定への展望：SASを用いたBrunnerMunzel検定の実装
橋本 隆正、舟尾 暢男、水間 浩太郎（武田薬品工業株式会社 日本開発センター 生物統計室）
-------287

ベイズ流ワイブル回帰に基づく Commensurate Prior 法のシミュレーションによる動作特性の検討
高戸 俊一 [1,4], 田中 優一 [2,4], 吉本 拓矢 [3,4]
（[1] 杏林製薬株式会社 開発推進部、[2] 大鵬薬品株式会社 データサイエンス部、[3] 中外製薬株式会社 バイオメトリクス部、[4] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会）
-------308

ワクチン試験での必要イベント数の設定について
浦狩保則（シミック（株）データサイエンス事業本部） -------325

PMx 解析に向けた統合データセット構築の実務的工夫 ～SAS プログラマーの視点から～
トロンナムチャイ ブンスィニー、白武 要（イーピーエス株式会社 臨床開発事業本部 データサイエンスセンター 統計解析 1 部） -------335

SASから呼び出すPowerShell
伊住 寿雄（ソーバル株式会社） -------340

ワクチン開発における非劣性試験のデザインについて
江口凌平、浦狩保則 (シミック(株)データサイエンス事業本部) -------349

層別Newcombe信頼区間による割合の群間差の非劣性検定における欠測データに対する多重代入法について
多田 由希夫，長谷川 貴大 （塩野義製薬株式会社） --------360

ジョブ理論を組み込んだAIプロフィール生成とDataMaker活用の新展開～ SAS Hackathon 2024の実証実験に基づく～
村田 粋堂、下山 晃平、八木橋 和也、田中 蒼斗、渡慶次 一哲、小野 潔 (コムチュア株式会社 デジタルイノベーション本部) -------374

ベイズ統計における有効サンプルサイズ(ESS)のSASによる実装
玉乃井 崚太（シミック（株）データサイエンス事業本部） -------389

目次

SAS による条件付き検出力の算出と性能比較
栗田凪，魚住龍史（東京科学大学工学院経営工学系）------405

トレンドの変化点を推定可能な区分線形回帰の SAS による実装と応用事例のご紹介
ビービ 彩加、菊池 宏和（シミック（株）データサイエンス事業本部） -------419

2 値分類変数に対するベイズ流例数設計の SAS による実装
高橋 朋寛，古賀 優一，平野 真太郎（エイツーヘルスケア株式会社 データサイエンス本部 データサイエンスソリューション部） -------430

用量反応試験における normal dynamic linear model 及びベイズ流階層モデルのシミュレーションによる動作特性の検討
畑山知慶[1]、松田裕也[1,2]、栗城正朗[1,3]、伊庭克拓[1,4]（[1] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会、[2] 中外製薬株式会社 デジタル戦略企画部、[3] 協和キリン株式会社 開発本部 バイオメトリックス部、[4] 大塚製薬株式会社 バイオメトリックス部） -------447

～実例と演習で学ぶ～ Transpose Procedure / Array Statement
牛尾 英孝（株式会社ピープロ・ジャパン） -------456

がん第 II 相バスケット試験でのベイズ流階層モデルの適用と動作特性評価
豊泉 樹一郎[1]、松田 裕也[1,2]（[1] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会、[2] 中外製薬株式会社） -------466

BASE SAS 環境における k 近傍法を用いた欠測値補完の実装と精度評価
小林 邦世（イーピーエス株式会社） -------473

～ノンパラメトリック検定～ 代表的手法及びパラメトリック検定との比較
牛尾 英孝（株式会社ピープロ・ジャパン） ------492

外部データから生成した Mixture prior を用いた Bayesian MMRM のシミュレーションの実装
東川 正晃[1,3]、遠藤 麻衣[2,3]（[1]EA ファーマ株式会社、[2] 田辺三菱製薬株式会社、[3] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会）-------502

TRAJ プロシジャを用いた縦断測定データ解析方法
石渡量太、橋口正行、志賀剛（東京慈恵医科大学 医学部臨床薬理講座） -------514

目次

RMSTREG プロシジャの回帰分析を理解する
久貝航太郎，魚住龍史（東京科学大学工学院経営工学系）------529

境界内平均生存時間による生存時間解析の実践
魚住龍史（東京科学大学工学院経営工学系）-------542

プレゼンテーション

IML ライセンスを利用せずに Proc FCMP で基本的な行列計算を行う方法について
森岡 裕 （イーピーエス株式会社） -------552

SAS システムでリアルワールドデータを解析する際の注意点
岡本 史子、平野 勝也 （株式会社インテリム データサイエンスユニット メディカルデータアナリシスグループ リアルワールドデータ解析チーム） -------553

SAS Packages - the way to share
Jablonski, Bartosz（Takeda Pharmaceutical / yabwon / Warsaw University of Technology）
-------554

SASPAC - SAS package の作り方と事例紹介 -
仲家 諒、鳴澤 輝子（武田薬品工業株式会社） -------555

小さく始める PROC TRANSPOSE ～ 転置無用？ ～
森田 祐介、太田 裕二、内藤 寿輝、浜田 泉 （ノーベルファーマ株式会社 データサイエンス部） -------556

クリックで見せる！Gemini を活用したハイライトグラフ作成法
橋本菜乃、小野洋平（聖路加国際大学公衆衛生学研究科） -------557

SAS Viya を中核とする統合解析環境のさらなる進化
坂井 聡、松下 和希、北西 由武（塩野義製薬株式会社 DX 推進本部 データサイエンス部）
-------558

PharmaSUG 2025 U.S. レポート ―海外学会で得た最新の知見と気づき―
家田 維哉 （日本たばこ産業株式会社） -------559

AI-SAS for RWE：STEP1 実装と STEP2 開発の展望
米田 卓司，藤田 智紀，北西 由武（塩野義製薬株式会社） -------560

目次

SAS による RWD 研究のデータ分析の落とし穴 ～SAS/ACCESS Interface to Amazon Redshift の機能を通して～
土屋 裕章[1]、高見 光広[2]、北村 嵩英[2] （[1]Johnson & Johnson Innovative Medicine、[2]イーピーエス株式会社） -------561

観察研究を外部対照とする際の症例数と検出力の計算～それでも Power procedure を使いますか？～
馬場 崇充[1,2]，堀江 悠生[1]，藤田 智紀[2]，坂口 弘樹[1] （[1]塩野義製薬株式会社 解析センター，[2]塩野義製薬株式会社 データサイエンス部） -------562

外部データの取り込み・加工に関する拡張機能井上 貴博
（ノバルティス ファーマ株式会社） -------563

MMRM における Cohen's d の考察
高橋 祐策、吉田 祐樹、北西 由武 （塩野義製薬株式会社 DX 推進本部 データサイエンス部） -------564

ログの確認と対処法
後藤 駿輔、井上 貴博（ノバルティス ファーマ株式会社） -------565

ADS spec を基にした SAS コード生成の検討
藤田 智紀、任 点、米田 卓司、吉田 祐樹、北西 由武（塩野義製薬株式会社） -------566

有害事象データの可視化とその有用性 -------567
笹原 大敬（イーピーエス株式会社）

SAS における PROC SQL ～DATA ステップとの違いとサブクエリ活用まで～
武藤 隆成（エイツーヘルスケア株式会社） -------568

ダミーデータを用いた疫学研究の手法の確認
矢田 徹（株式会社セブントゥワン） -------569

SAS における MCP 利用の検討
中松 建 -------570

目次

EPS-RBaseによるR活用を支えるサービスの紹介
魚田 孝志、森岡 裕（イーピーエス株式会社）　-------571

GTL（GraphTemplateLanguage）未経験者のためのグラフ作成の基礎とSGPLOTとの比較を通したGTLの利点に関する紹介
安 正鎬、藤川 桂、宮辻 淳（株式会社インテリム データサイエンスユニット）
-------572

SAS・R・Pythonにおける条件式の比較
和田 尚磨、谷 里志（株式会社タクミインフォメーションテクノロジー）　-------573

SAS・R・Pythonにおける変数処理の比較
谷 里志、和田 尚磨（株式会社タクミインフォメーションテクノロジー）　-------574

SASの“あるある”をスマートに解決実務に効く3つのマクロをパッケージ化
鈴木 慎吾、西井 雅之、毎田 豊、土生 敏明、小原 裕美子（SAS Institute Japan）---575

PharmaForestでのパッケージ紹介

森岡 裕（イーピーエス株式会社）　-------576

次世代に向けたSASの取り組みについて
土生 敏明、吉原 也真人、西井 雅之、成松 怜央（SAS Institute Japan）　-------577

PharmaForestについて
山野辺 浩己（マルホ株式会社）　-------578

Estimand初心者集まれ！ Estimandの復習と解析業務への影響点の紹介
菊池 宏和、中西 豊支（シミック株式会社）　-------579

JMPによるCDISC標準サンプルデータ(SDTM)の可視化手法の紹介
中野 薫、坂本 佳代子、上野 恵子（株式会社タクミインフォメーションテクノロジー）-------580

Enhancing Statistical Analysis Workflows: LLM-Driven Approach with SAS HTTP Procedure
高浪 洋平、仲家 諒（武田薬品工業株式会社）　-------581

SASによるMMRM（反復測定のための混合効果モデル）の実装と臨床試験への応用（加齢黄斑変性（AMD）に伴う地図状萎縮（GA）進行抑制効果の解析を例として）
ZU YANHUI、ZHAO YUNXI（Clinchoice 株式会社）　-------582

Pharmacometrics 解析のためのデータセット作成プロセスの改善 SAS とR のCollaboration の提案
碇 新里[1]、Dimple Patel[2]、永田吉範[1]、Douglas Marsteller[2] （[1] 協和キリン株式会社 開発本部 医科学薬理部、[2]Kyowa Kirin, Inc.） -------583

AI エージェントと SAS Intelligent Decisioning の融合 〜 CRM と与信を超えて広がる新しいパラダイム〜
小野 潔（コムチュア株式会社 デジタルイノベーション本部） -------584

現場で役立つ SAS 活用術：リアルワールドデータ解析の実践
小川武則[1]、金沢 信一[2]（[1] メディカル・データ・ビジョン株式会社、[2] エイツーヘルスケア株式会社） -------585

リバーシチャンピオンシップ

出場者： -------586

関根暁史(藤本製薬株式会社)
伊住寿雄(ソーバル 株式会社)
川上貴弘(DXC Technology Japan)
太田裕二(ノーベルファーマ株式会社)
平井健太(株式会社タクミインフォメーションテクノロジー)
飯島康浩（個人登録）
関根伸吾(ネイチャーインサイト株式会社)
飯島泰平(ネイチャーインサイト株式会社)

一般線型モデルとGLMプロシジャの基礎

田中 真史

（イーピーエス株式会社）

Understanding the Basis of the General Linear Model and SAS GLM Procedure

Masashi Tanaka

EPS Corporation

要旨

一般線型モデルと SAS の GLM プロシジャは、推定可能関数の観点から考えると分かりやすい。推定可能関数は GLM プロシジャ全般に関わっており、各ステートメントで推定可能関数を出力しながら解析を行うと、出力結果について、より確実な理解や判断ができるはずである。本稿では、一般線型モデルと GLM プロシジャの基礎をできるだけ明確に説明した。

キーワード：一般線型モデル（General Linear Model）、GLM プロシジャ（GLM Procedure）、推定可能関数（Estimable Function）、最小二乗平均（Least Squares Means）、平方和のタイプ（Type of Sum of Squares）

1. はじめに

一般線型モデル（General Linear Model）を解析するために SAS で GLM プロシジャが開発されたのは、ほぼ半世紀前の 1976 年である [1] [2]。今日では、少なくとも臨床試験の統計解析ではより新しい解析プロシジャが常用されており、GLM プロシジャを使用する機会はおそらく以前よりも減っている。

しかしながら、データにシンプルな仮定を置ける場合の GLM プロシジャの有用性は変わらず、難しいモデル解析を検討する際も、準備として GLM プロシジャで様子を調べることがある。これらの有用性は、一般線型モデルが線型代数と多変量正規分布の知識のみで分析でき、モデル解析の本質的な仕組みが分かりやすいことにも由来するのであろう。GLM プロシジャのステートメントの多くは MIXED プロシジャなどに対応するステートメントがあり [3]、モデル解析用のプロシジャの基本的な知識を提供する。GLM プロシジャとそれが扱う「古典的」な一般線型モデルは今なお重要である。

以下、2 章で一般線型モデルの仕組みを「推定可能関数」（Estimable Function）なる装置に依拠して筆者の理解の範囲で説明する。3 章では、2 章の内容に基づき GLM プロシジャの出力結果を調べる。この過程を通して、より確実に一般線型モデル及び GLM プロシジャを扱い、SAS によるモデル解析全般を理解するための基礎を得ることが本稿の目的である。

2. 一般線型モデルの基礎

2.1. 一般線型モデル

本稿ではGLMプロシジャで解析できる一般線型モデルのみを扱う。$\boldsymbol{y}$（n次元ベクトル）をデータ、$\boldsymbol{X}$（$n \times m$行列）を計画行列、$\boldsymbol{\beta}$（m次元ベクトル）をパラメータ、$\boldsymbol{\varepsilon}$（n次元ベクトル）を平均$\boldsymbol{0}$、分散$\sigma^2\boldsymbol{I}$（$\boldsymbol{I}$は単位行列）の正規誤差とする。確率モデルは

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.1.1}$$

とする。$\boldsymbol{y}$と$\boldsymbol{\varepsilon}$は確率変数である。$\boldsymbol{\beta}$については、推定値を$\widehat{\boldsymbol{\beta}}$と表して$\boldsymbol{\beta}$と区別する。推定値$\widehat{\boldsymbol{\beta}}$は

$$\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{2.1.2}$$

の正規方程式を解くことで得られる。$\boldsymbol{X}'\boldsymbol{X}$が逆行列を持つ場合は、$\widehat{\boldsymbol{\beta}}$が一意に定まり悩みは無い。しかしながら、一般に正規方程式は一意に解けず無数の解が存在するので、一意に定まる推定量を見つけたい。そのために次節で「推定可能関数」を定義する。

3章のために正規方程式を正射影の射影行列で書き直しておく。$\boldsymbol{X} = (\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_m)$の列空間（$c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m$の集合（$c_1, \ldots, c_m$は実数））を$\mathrm{Im}\,\boldsymbol{X}$で表す。式2.1.2より$\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$は$\mathrm{Im}\,\boldsymbol{X}$に直交する。$\mathrm{Im}\,\boldsymbol{X}$への正射影を射影行列$\boldsymbol{P}$で表すと、$\boldsymbol{P}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{0}$で、$\boldsymbol{PX} = \boldsymbol{X}$だから

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{2.1.3}$$

が得られる。射影行列は、$\boldsymbol{X}$の列ベクトルから基底を選んで$\tilde{\boldsymbol{X}}$と置いて、$\boldsymbol{P} = \tilde{\boldsymbol{X}}\left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'$で計算できる。

2.2. 推定可能関数

列空間の双対ベクトルから始める。この場合の双対ベクトルとは、内積を取ることである。

定義 1　列空間の双対ベクトル

計画行列$\boldsymbol{X}$の列空間$\mathrm{Im}\,\boldsymbol{X}$から実数への線型写像を列空間の双対ベクトルと定義する。

$\boldsymbol{v} \in \mathrm{Im}\,\boldsymbol{X}$を転置した横ベクトル$\boldsymbol{v}'$を列空間の双対ベクトルと見做すことができる。実際、$\boldsymbol{x} \in \mathrm{Im}\,\boldsymbol{X}$に対して$\boldsymbol{v}'\boldsymbol{x}$を与える線型写像は列空間の双対ベクトルである。逆に列空間の任意の双対ベクトル$\mathcal{E}$は、$\mathrm{Im}\,\boldsymbol{X}$の正規直交基底$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$を取ると、$\boldsymbol{v} = \mathcal{E}(\boldsymbol{e}_1)\boldsymbol{e}_1 + \cdots + \mathcal{E}(\boldsymbol{e}_k)\boldsymbol{e}_k$を用いて$\mathcal{E}(\boldsymbol{x}) = \boldsymbol{v}'\boldsymbol{x}$と表せる。以下、$\boldsymbol{v}'\boldsymbol{x} = \boldsymbol{v} \cdot \boldsymbol{x}$と内積の記号も適当に使う。

定義 2　推定可能関数

列空間の双対ベクトル$\mathcal{E}$から$\tilde{\mathcal{E}}(\boldsymbol{\beta}) = \mathcal{E}(\boldsymbol{X\beta})$で定義した$\boldsymbol{\beta}$の関数$\tilde{\mathcal{E}}$を推定可能関数と呼ぶ。

$\boldsymbol{\beta}$の推定値$\widehat{\boldsymbol{\beta}}$が一意に定まらなくても、$\boldsymbol{X}\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_2$であれば$\tilde{\mathcal{E}}\left(\widehat{\boldsymbol{\beta}}_1\right) = \tilde{\mathcal{E}}\left(\widehat{\boldsymbol{\beta}}_2\right)$となり、推定可能関数は同じ値を与える。次の性質1が重要である。

性質 1　$\boldsymbol{X\beta}_1 = \boldsymbol{X\beta}_2$であれば$\tilde{\mathcal{E}}(\boldsymbol{\beta}_1) = \tilde{\mathcal{E}}(\boldsymbol{\beta}_2)$であるような$\boldsymbol{\beta}$から実数への線型写像$\tilde{\mathcal{E}}$は、ある$\boldsymbol{v} \in \mathrm{Im}\,\boldsymbol{X}$によって$\tilde{\mathcal{E}}(\boldsymbol{\beta}) = \boldsymbol{v}'\boldsymbol{X\beta}$と表すことができる。つまり$\tilde{\mathcal{E}}$は推定可能関数である。

証明　$\tilde{\mathcal{E}}$は線型写像だから、ある$\boldsymbol{l}$によって$\tilde{\mathcal{E}}(\boldsymbol{\beta}) = \boldsymbol{l}'\boldsymbol{\beta}$と表せる。仮定の式で$\boldsymbol{\beta}_2 = \boldsymbol{0}$と置くと、$\boldsymbol{X\beta}_1 = \boldsymbol{0}$であれば$\boldsymbol{l}'\boldsymbol{\beta}_1 = 0$であることが分かる。$\boldsymbol{X\beta} = \boldsymbol{0}$であるような$\boldsymbol{\beta}$の集合を$\mathrm{Ker}\,\boldsymbol{X}$、その直交補空間を$(\mathrm{Ker}\,\boldsymbol{X})^{\perp}$、

$\boldsymbol{l}'\boldsymbol{\beta} = 0$であるような$\boldsymbol{\beta}$の集合をKer $\boldsymbol{l}'$と書く。Ker $\boldsymbol{X}$とKer $\boldsymbol{l}'$の関係は、Ker $\boldsymbol{X} \subset$ Ker $\boldsymbol{l}'$である。$\boldsymbol{\beta}$を

$$\boldsymbol{\beta} = \boldsymbol{\gamma} + \boldsymbol{\delta}, \qquad \boldsymbol{\gamma} \in (\text{Ker } \boldsymbol{X})^{\perp}, \qquad \boldsymbol{\delta} \in \text{Ker } \boldsymbol{X}$$

と直交分解する。ここで$\boldsymbol{X\beta}$を$\boldsymbol{\gamma}$に対応付ける写像が定義できる。$\boldsymbol{X\beta} = \boldsymbol{X\tilde{\beta}}$であれば$\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}} \in$ Ker $\boldsymbol{X}$で、$\tilde{\boldsymbol{\beta}}$の対応先もやはり$\boldsymbol{\gamma}$だからである。この写像は線型写像で、$\boldsymbol{\gamma} = \boldsymbol{AX\beta}$と行列$\boldsymbol{A}$で表せる。Ker $\boldsymbol{X} \subset$ Ker $\boldsymbol{l}'$だから$\boldsymbol{l}'\boldsymbol{\beta} = \boldsymbol{l}'\boldsymbol{\gamma} = \boldsymbol{l}'\boldsymbol{AX\beta}$で、$\boldsymbol{\beta}$は任意だから$\boldsymbol{l}' = (\boldsymbol{A}'\boldsymbol{l})'\boldsymbol{X}$が得られる。$\boldsymbol{A}'\boldsymbol{l} = \boldsymbol{v} + \boldsymbol{w}$、$\boldsymbol{v} \in$ Im $\boldsymbol{X}$、$\boldsymbol{w} \in (\text{Im } \boldsymbol{X})^{\perp}$と直交分解すれば

$$\boldsymbol{l}' = \boldsymbol{v}'\boldsymbol{X} \tag{2.2.1}$$

である。$\tilde{\mathcal{E}}$については

$$\tilde{\mathcal{E}}(\boldsymbol{\beta}) = \boldsymbol{l}'\boldsymbol{\beta} = \boldsymbol{v}'\boldsymbol{X\beta}$$

となる。 ■

式2.2.1を導くところは次のように考えても良い。$\boldsymbol{X}^{+} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{l}' \end{pmatrix}$と置くと、Ker $\boldsymbol{X}$とKer $\boldsymbol{X}^{+}$は同じである。線型代数の教科書の「次元公式」辺りを参照すると、$\boldsymbol{X}$の行空間の次元と$\boldsymbol{X}^{+}$の行空間の次元は等しいことが分かる。$\boldsymbol{X}$の行空間の基底がそのまま$\boldsymbol{X}^{+}$の行空間の基底になるから、$\boldsymbol{l}'$は$\boldsymbol{X}$の行ベクトルの線型結合で$\boldsymbol{l}' = \boldsymbol{u}'\boldsymbol{X}$と表せる。$\boldsymbol{u}$を直交分解してIm $\boldsymbol{X}$の成分を取り出せば後は同じである。

$\boldsymbol{X\beta}$を$\boldsymbol{y}$の期待値と見ると$\boldsymbol{l}'\boldsymbol{\beta} = E[\boldsymbol{v}'\boldsymbol{y}]$で、これが SAS のユーザーズガイド冒頭の$\boldsymbol{l}'\boldsymbol{\beta}$が「推定可能」であることの定義である [4]。

性質 1 に基づいて、「$\boldsymbol{X\beta}_1 = \boldsymbol{X\beta}_2$であれば同じ値を与える$\boldsymbol{\beta}$から実数への線型写像」を「推定可能関数」と定義することもできる。しかし本稿では、データの空間で内積を取る図形的なイメージを重視し、Im $\boldsymbol{X}$の双対ベクトルを経由した。データの空間で式を解釈できることは、推定値$\widehat{\boldsymbol{\beta}}$の不定性に計算結果が影響を受けないことや、列空間が同じなら$\boldsymbol{X}$以外の計画行列を選んでも計算結果が変わらないことを意味する。

式2.2.1の右辺の$\boldsymbol{v}$を$\boldsymbol{X}$の列ベクトルの線型結合で$\boldsymbol{Xc}$と表せば

$$\boldsymbol{l}' = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X} \tag{2.2.2}$$

となる。この式はパラメータの数と同じ本数の1次式なので扱いやすい。適当に$\boldsymbol{l}'$を定めたときに上手く$\boldsymbol{c}'$が見つかれば、$\tilde{\mathcal{E}}(\boldsymbol{\beta}) = \boldsymbol{l}'\boldsymbol{\beta} = (\boldsymbol{Xc})'\boldsymbol{X\beta}$が推定可能関数になる。内積で考えれば、$\boldsymbol{Xc}$が$\boldsymbol{X}$の列ベクトルに直交しているときに$\boldsymbol{l}'$の成分が0になる。必要な成分以外が0であるような推定可能関数を見つけることが一般線型モデルの分析のポイントである。$\boldsymbol{c}$の不定性は、$\boldsymbol{Xc}$として用いる限り結果に影響しない（性質 2）。式2.2.2の解はひとつ求めれば良い。

性質 2　$\boldsymbol{c}_1'\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{c}_2'\boldsymbol{X}'\boldsymbol{X}$ならば$\boldsymbol{Xc}_1 = \boldsymbol{Xc}_2$である。

証明　$\boldsymbol{v} = \boldsymbol{X}(\boldsymbol{c}_1 - \boldsymbol{c}_2)$と置く。もとより$\boldsymbol{v} \in$ Im $\boldsymbol{X}$で、$\boldsymbol{X}'\boldsymbol{v} = \boldsymbol{0}$より$\boldsymbol{v} \in (\text{Im } \boldsymbol{X})^{\perp}$だから$\boldsymbol{v} = \boldsymbol{0}$である。 ■

2.3. 帰無仮説に対する平方和

推定可能関数の値、すなわち$\tilde{\mathcal{E}}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{l}'\widehat{\boldsymbol{\beta}}$の分布を求め、そこから平方和を定義する。平方和についての知識は、GLM プロシジャのアウトプットを見る際に役立つ。

性質 3　$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$のとき、$\tilde{\mathcal{E}}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{l}'\widehat{\boldsymbol{\beta}}$は、平均$\boldsymbol{l}'\boldsymbol{\beta}$、分散$\sigma^2\boldsymbol{l}'\boldsymbol{c}$の正規分布に従う。$\boldsymbol{c}$は$\boldsymbol{l}' = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}$のひとつの解である。

証明　式2.2.2、正規方程式（2.1.2）及び確率モデルの式（2.1.1）より

$$\boldsymbol{l}'\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{c}'\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{l}'\boldsymbol{\beta} + \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{\varepsilon}$$

である。よって$\boldsymbol{l}'\widehat{\boldsymbol{\beta}}$は

$$E\left[\boldsymbol{l}'\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{l}'\boldsymbol{\beta},$$
$$V\left[\boldsymbol{l}'\widehat{\boldsymbol{\beta}}\right] = \sigma^2\boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c} = \sigma^2\boldsymbol{l}'\boldsymbol{c}$$

の1次元正規分布である。　■

$\boldsymbol{l}'\boldsymbol{\beta} = 0$の帰無仮説の下では$\sigma^{-2}\left(\boldsymbol{l}'\widehat{\boldsymbol{\beta}}\right)^2(\boldsymbol{l}'\boldsymbol{c})^{-1}$が自由度1のカイ二乗分布に従う。そこで平方和を次のように定義する。

定義 3a　帰無仮説$\boldsymbol{l}'\boldsymbol{\beta} = 0$に対する平方和

帰無仮説$\boldsymbol{l}'\boldsymbol{\beta} = 0$に対する平方和を$SS(H_0: \boldsymbol{l}'\boldsymbol{\beta} = 0)$と表し

$$SS(H_0: \boldsymbol{l}'\boldsymbol{\beta} = 0) = \frac{\left(\boldsymbol{l}'\widehat{\boldsymbol{\beta}}\right)^2}{\boldsymbol{l}'\boldsymbol{c}}$$

と定義する。

$SS(H_0: \boldsymbol{l}'\boldsymbol{\beta} = 0)$はデータの空間で次のような意味を持つ。$\boldsymbol{v} = \boldsymbol{X}\boldsymbol{c}$の大きさを$|\boldsymbol{v}|$と書けば

$$\boldsymbol{l}'\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{v}\cdot\boldsymbol{X}\widehat{\boldsymbol{\beta}},$$
$$\boldsymbol{l}'\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c} = |\boldsymbol{v}|^2$$

より

$$SS(H_0: \boldsymbol{l}'\boldsymbol{\beta} = 0) = \left[(\boldsymbol{v}/|\boldsymbol{v}|)\cdot\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\right]^2$$

である。これは$\boldsymbol{v}$の方向に$\boldsymbol{X}\widehat{\boldsymbol{\beta}}$を射影して測った長さの2乗だから、確かに「平方和」である。

3 章では使わないが、SAS のユーザーズガイド冒頭の式に触れておく [4]。一次独立なk個の推定可能関数の組を$\boldsymbol{L} = \begin{pmatrix}\boldsymbol{l}_1' \\ \vdots \\ \boldsymbol{l}_k'\end{pmatrix}$、それに対応する$\boldsymbol{c}'$の組を$\boldsymbol{C} = \begin{pmatrix}\boldsymbol{c}_1' \\ \vdots \\ \boldsymbol{c}_k'\end{pmatrix}$と表せば、式2.2.2は$\boldsymbol{L} = \boldsymbol{C}\boldsymbol{X}'\boldsymbol{X}$とまとめられる。この式の両辺の階数を考えると、$\boldsymbol{X}\boldsymbol{C}'$は階数が$k$で列ベクトルが一次独立であることが分かる。よって、$\boldsymbol{L}\boldsymbol{C}' = (\boldsymbol{X}\boldsymbol{C}')'\boldsymbol{X}\boldsymbol{C}'$は逆行列を持つ。$\boldsymbol{L}\widehat{\boldsymbol{\beta}}$の分布は、$E\left[\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{L}\boldsymbol{\beta}$、$V\left[\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right] = \sigma^2\boldsymbol{L}\boldsymbol{C}'$の正規分布である。$\boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$の帰無仮説の下では$\sigma^{-2}\left(\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right)'(\boldsymbol{L}\boldsymbol{C}')^{-1}\boldsymbol{L}\widehat{\boldsymbol{\beta}}$が自由度$k$のカイ二乗分布に従う。平方和は次のように定義する。

定義 3b　帰無仮説$\boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$に対する平方和

帰無仮説$\boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$に対する平方和を$SS(H_0: \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0})$と表し

$$SS(H_0: \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}) = \left(\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right)'(\boldsymbol{L}\boldsymbol{C}')^{-1}\boldsymbol{L}\widehat{\boldsymbol{\beta}}$$

と定義する。一般化逆行列を用いて$\boldsymbol{C}'$を$(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{L}'$に置き換えれば、右辺は$\left(\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right)'[\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{L}']^{-1}\boldsymbol{L}\widehat{\boldsymbol{\beta}}$である。

$\widetilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{C}'$と置き、Im $\widetilde{\boldsymbol{X}}$への射影行列を$\widetilde{\boldsymbol{P}} = \widetilde{\boldsymbol{X}}\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'$と書くと

$$SS(H_0: \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}) = \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\widetilde{\boldsymbol{X}}\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \left|\widetilde{\boldsymbol{P}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right|^2$$

となる。これは確かに「平方和」である。

3. GLM プロシジャのアウトプット

本章では、GLM プロシジャのアウトプットを調べる。SAS システムのバージョンは、Base SAS 9.4M7 及び SAS/STAT 15.2 を用いた。

3.1. 一元配置分散分析（1 要因 2 水準）— 基本と ESTIMATE ステートメント

要因A、データ$\boldsymbol{y}$、計画行列$\boldsymbol{X}$、パラメータ$\boldsymbol{\beta}$が次の場合を例にする。

要因A	y
$A1$	1
$A1$	3
$A2$	2
$A2$	4

$$\boldsymbol{X}=\begin{pmatrix}1 & 1 & 0\\ 1 & 1 & 0\\ 1 & 0 & 1\\ 1 & 0 & 1\end{pmatrix},\qquad \boldsymbol{\beta}=\begin{pmatrix}\mu\\ A1\\ A2\end{pmatrix}$$

```
data z;
   input A y;
   cards;
1 1
1 3
2 2
2 4
;
run;
```

```
proc glm data = z;
   class A;
   model y = A / e;
   estimate 'A1-A2' A 1 -1 / e;
   contrast 'A' A 1 -1 / e;
quit;
```

3.1a

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	1	1.00000000	1.00000000	0.50	0.5528
Error	2	4.00000000	2.00000000		
Corrected Total	3	5.00000000			

General Form of Estimable Functions	
効果	係数
Intercept	L1
A 1	L2
A 2	L1-L2

3.1b

推定 A1 VS A2 の係数	
	行 1
Intercept	0
A 1	1
A 2	-1

パラメータ	推定値	標準誤差	t 値	Pr > |t|
A1-A2	-1.00000000	1.41421356	-0.71	0.5528

3.1c

対比 A の係数	
	行 1
Intercept	0
A 1	1
A 2	-1

対比	自由度	対比平方和	平均平方	F 値	Pr > F
A	1	1.00000000	1.00000000	0.50	0.5528

$\boldsymbol{X}$の一次独立な行ベクトルは$(1\quad 1\quad 0)$及び$(1\quad 0\quad 1)$だから、式2.2.1より推定可能関数はこれらの線型結合である。GLM プロシジャが$\boldsymbol{X}$の行空間の基底として$(1\quad 0\quad 1)$及び$(0\quad 1\quad -1)=(1\quad 1\quad 0)-(1\quad 0\quad 1)$を取ったため、MODEL ステートメントの E オプションで得られた推定可能関数の一般的な表現は

$$\boldsymbol{l}'=L_1(1\quad 0\quad 1)+L_2(0\quad 1\quad -1)=(L_1\quad L_2\quad L_1-L_2)$$

である（3.1a 右）。ESTIMATE ステートメントで指定した$\boldsymbol{l}'=(0\quad 1\quad -1)$は、$L_1=0$及び$L_2=1$の場合で、それを E オプションで出力した（3.1b 左）。推定値$\boldsymbol{l}'\hat{\boldsymbol{\beta}}$は次のように計算する。式2.2.2は

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c}=\begin{pmatrix}4&2&2\\2&2&0\\2&0&2\end{pmatrix}\boldsymbol{c}=\begin{pmatrix}0\\1\\-1\end{pmatrix}$$

で、$\boldsymbol{c}=\begin{pmatrix}-0.5\\1\\0\end{pmatrix}$が解のひとつである。$\boldsymbol{X}'\boldsymbol{X}$はMODELステートメントのXPXオプションで出力できる。

$\boldsymbol{l}'\widehat{\boldsymbol{\beta}}=\boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\begin{pmatrix}0.5\\0.5\\-0.5\\-0.5\end{pmatrix}\cdot\boldsymbol{X}\widehat{\boldsymbol{\beta}}$だから、$\boldsymbol{X}\widehat{\boldsymbol{\beta}}$から$\boldsymbol{l}'\widehat{\boldsymbol{\beta}}$が得られる。$\boldsymbol{X}\widehat{\boldsymbol{\beta}}$はIm $\boldsymbol{X}$への射影行列$\boldsymbol{P}$を用いて計算する（式2.1.3）。$\boldsymbol{X}$の列ベクトルから一次独立な2本を選び、$\widetilde{\boldsymbol{X}}=\begin{pmatrix}1&1\\1&1\\1&0\\1&0\end{pmatrix}$と置く。

$$\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{P}\boldsymbol{y}=\widetilde{\boldsymbol{X}}\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{y}=\begin{pmatrix}0.5&0.5&0&0\\0.5&0.5&0&0\\0&0&0.5&0.5\\0&0&0.5&0.5\end{pmatrix}\begin{pmatrix}1\\3\\2\\4\end{pmatrix}=\begin{pmatrix}2\\2\\3\\3\end{pmatrix}$$

より、$\boldsymbol{l}'\widehat{\boldsymbol{\beta}}=-1$である（3.1b右）。

ESTIMATEステートメントで出力したt値は次のように計算できる。$\boldsymbol{l}'\boldsymbol{\beta}=0$の帰無仮説の下では、定義3aの$SS(H_0:\boldsymbol{l}'\boldsymbol{\beta}=0)$をErrorの平均平方（$\overline{SS_e}$）で割った値が自由度$(1,2)$の$F$分布に従う。換言すると、標準誤差を$\sqrt{\boldsymbol{l}'\boldsymbol{c}\overline{SS_e}}$と定義して$\boldsymbol{l}'\widehat{\boldsymbol{\beta}}/\sqrt{\boldsymbol{l}'\boldsymbol{c}\overline{SS_e}}$が自由度2の$t$分布に従う。Errorの平方和は$\left|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right|^2=4$で、$\overline{SS_e}=2$である（3.1a左）。$SS(H_0:\boldsymbol{l}'\boldsymbol{\beta}=0)=1$から$-0.7071$の$t$値が得られる（3.1b右、$-0.71$に丸められた）。$SS(H_0:\boldsymbol{l}'\boldsymbol{\beta}=0)$はCONTRASTステートメントで対比平方和として直接得られる（3.1c右）。推定可能関数はやはりEオプションで出力できる（3.1c左）。この例では、$SS(H_0:\boldsymbol{l}'\boldsymbol{\beta}=0)$がModelの平方和と一致し、$t$値の2乗がModelの$F$値になる（3.1c右）。

3.2. 一元配置分散分析（1要因3水準）— CONTRASTステートメント

次に要因Aが3水準の場合を例にして、CONTRASTステートメントについて考える。ESTIMATEステートメントと異なり、CONTRASTステートメントではIm $\boldsymbol{X}$の2次元以上の部分空間へ$\boldsymbol{X}\widehat{\boldsymbol{\beta}}$を射影できる。その成分の2乗が対比平方和である。

要因A	y
$A1$	1
$A1$	3
$A2$	2
$A2$	4
$A3$	4
$A3$	6

$$\boldsymbol{X}=\begin{pmatrix}1&1&0&0\\1&1&0&0\\1&0&1&0\\1&0&1&0\\1&0&0&1\\1&0&0&1\end{pmatrix},\qquad\boldsymbol{\beta}=\begin{pmatrix}\mu\\A1\\A2\\A3\end{pmatrix}$$

```
data z;
   input A y;
   cards;
1 1
1 3
2 2
2 4
3 4
3 6
;
run;
```

```
proc glm data = z;
   class A;
   model y = A / e;
   contrast 'A' A 1 -1 0, A 0 1 -1 / e;
   contrast 'A1-A2' A 1 -1 0;
   contrast '0.5*A1+0.5*A2-A3' A 0.5 0.5 -1;
quit;
```

3.2a

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	2	9.33333333	4.66666667	2.33	0.2448
Error	3	6.00000000	2.00000000		
Corrected Total	5	15.33333333			

3.2b

対比 A の係数		
	行 1	行 2
Intercept	0	0
A 1	1	0
A 2	-1	1
A 3	0	-1

対比	自由度	対比平方和	平均平方	F 値	Pr > F
A	2	9.33333333	4.66666667	2.33	0.2448
A1-A2	1	1.00000000	1.00000000	0.50	0.5305
0.5*A1+0.5*A2-A3	1	8.33333333	8.33333333	4.17	0.1339

最初の CONTRAST ステートメントで指定した部分空間は

$$\boldsymbol{l}_1' = (0 \quad 1 \quad -1 \quad 0), \quad \boldsymbol{l}_2' = (0 \quad 0 \quad 1 \quad -1), \quad \boldsymbol{l}_1' = \boldsymbol{c}_1'\boldsymbol{X}'\boldsymbol{X}, \quad \boldsymbol{l}_2' = \boldsymbol{c}_2'\boldsymbol{X}'\boldsymbol{X}$$

で得た$\boldsymbol{X}\boldsymbol{c}_1$及び$\boldsymbol{X}\boldsymbol{c}_2$が張る部分空間で（3.2b 左）、対比$A$の対比平方和は Model の平方和に一致する（3.2a、3.2b 右）。対比Aの平均平方を Error の平均平方で割った値が対比AのF値で、自由度$(2, 3)$のF分布に従う（3.2b 右）。今回の例では$\boldsymbol{X}'\boldsymbol{X}$の構造が簡単なので、$\boldsymbol{l}_1$に直交する$\boldsymbol{l}_3$とそれに対応する$\boldsymbol{c}_3$を

$$\boldsymbol{l}_3' = \boldsymbol{l}_2' + 0.5\boldsymbol{l}_1' = (0 \quad 0.5 \quad 0.5 \quad -1), \qquad \boldsymbol{l}_3' = \boldsymbol{c}_3'\boldsymbol{X}'\boldsymbol{X}$$

とすると$\boldsymbol{X}\boldsymbol{c}_1$と$\boldsymbol{X}\boldsymbol{c}_3$が直交する。$\boldsymbol{c}_1 = \boldsymbol{l}_1/2$と選べて$(\boldsymbol{X}\boldsymbol{c}_1) \cdot (\boldsymbol{X}\boldsymbol{c}_3) = (\boldsymbol{l}_1 \cdot \boldsymbol{l}_3)/2 = 0$となるからである。$\boldsymbol{X}\boldsymbol{c}_1$と$\boldsymbol{X}\boldsymbol{c}_3$の方向への平方和を 2 番目と 3 番目の CONTRAST ステートメントで計算した。$\boldsymbol{X}\boldsymbol{c}_1$及び$\boldsymbol{X}\boldsymbol{c}_3$が張る部分空間と$\boldsymbol{X}\boldsymbol{c}_1$及び$\boldsymbol{X}\boldsymbol{c}_2$が張る部分空間は同じだから$SS(H_0: \boldsymbol{l}_1'\boldsymbol{\beta} = 0) + SS(H_0: \boldsymbol{l}_3'\boldsymbol{\beta} = 0) = 1 + 8.3333 = 9.3333$で対比$A$の対比平方和が再び得られる（3.2b 右）。

3.3. 最小二乗平均 — LSMEANS ステートメント

共分散分析を例にして LSMEANS ステートメントで最小二乗平均を求める。最小二乗平均は ESTIMATE ステートメントでも計算できる。

要因A	x	y
$A1$	1.1	2
$A1$	1.2	4
$A2$	1.2	3
$A2$	1.4	5
$A2$	1.6	6

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 1.1 \\ 1 & 1 & 0 & 1.2 \\ 1 & 0 & 1 & 1.2 \\ 1 & 0 & 1 & 1.4 \\ 1 & 0 & 1 & 1.6 \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ A1 \\ A2 \\ a \end{pmatrix}$$

```
data z;
   input A x y;
   cards;
1 1.1 2
1 1.2 4
2 1.2 3
2 1.4 5
2 1.6 6
;
run;
```

```
proc glm data = z;
   class A;
   model y = A x;
   lsmeans A / diff e;
   estimate 'A1' intercept 1 A 1 0 x 1.3;
   estimate 'A2' intercept 1 A 0 1 x 1.3;
   estimate 'A1-A2' A 1 -1;
quit;
```

3.3a

最小 2 乗平均 A 係数		
	A 水準	
効果	1	2
Intercept	1	1
A 1	1	0
A 2	0	1
x	1.3	1.3

		H0:LSMean1=LSMean2
A	y の最小 2 乗平均	Pr > |t|
1	4.23529412	0.6869
2	3.84313725	

3.3b

パラメータ	推定値	標準誤差	t 値	Pr > |t|
A1	4.23529412	0.58725408	7.21	0.0187
A2	3.84313725	0.45098039	8.52	0.0135
A1-A2	0.39215686	0.84108277	0.47	0.6869

LSMEANS ステートメントの E オプションで、最小二乗平均の計算で用いた推定可能関数を出力した（3.3a 左）。要因Aの水準$A1$及び$A2$に対して、推定可能関数はそれぞれ

$$\boldsymbol{l}'_{A1} = (1 \quad 1 \quad 0 \quad 1.3),$$
$$\boldsymbol{l}'_{A2} = (1 \quad 0 \quad 1 \quad 1.3)$$

で、1.3は$\boldsymbol{x}$の平均値である。$\boldsymbol{l}'_{A1}\widehat{\boldsymbol{\beta}}$及び$\boldsymbol{l}'_{A2}\widehat{\boldsymbol{\beta}}$の推定値は ESTIMATE ステートメントでも計算でき、最小二乗平均に一致する（3.3a 右、3.3b）。ESTIMATE ステートメントに$\boldsymbol{l}'_{A1} - \boldsymbol{l}'_{A2} = (0 \quad 1 \quad -1 \quad 0)$を指定すると、推定値（0.3921）は最小二乗平均の差と一致し、p 値（0.6869）も LSMEANS ステートメントの DIFF オプションで出力した値と一致する（3.3a 右、3.3b）。

極端な状況ではあるが、次の例のように最小二乗平均が推定可能にならない場合がある。理由は明白で$(1 \quad 1 \quad 0 \quad 1.3)$と$(1 \quad 0 \quad 1 \quad 1.3)$を$\boldsymbol{X}$の行ベクトルの線型結合で作れないからである（3.3c 左右）。

要因A	x	y
$A1$	1.1	2
$A2$	1.4	5
$A2$	1.4	5

$$X = \begin{pmatrix} 1 & 1 & 0 & 1.1 \\ 1 & 0 & 1 & 1.4 \\ 1 & 0 & 1 & 1.4 \end{pmatrix}, \qquad \beta = \begin{pmatrix} \mu \\ A1 \\ A2 \\ a \end{pmatrix}$$

```
data z;
   input A x y;
   cards;
1 1.1 2
2 1.4 5
2 1.4 5
;
run;
```

```
proc glm data = z;
   class A;
   model y = A x;
   lsmeans A / e;
quit;
```

3.3c

最小 2 乗平均 A 係数		
	A 水準	
効果	1	2
Intercept	1	1
A 1	1	0
A 2	0	1
x	1.3	1.3

A	y の最小 2 乗平均
1	Non-est
2	Non-est

3.4. 平方和のタイプ — 交互作用無し（Type I 及び Type II）

要因の効果を推定する際、どのタイプの平方和を選択するかは難しい問題である。本稿ではこの問題には直接立ち入らず、3.4、3.5 及び 3.6 章で平方和のタイプごとの計算方法について筆者の理解の範囲で説明する。本節では、交互作用を考えない場合の Type I 及び Type II 平方和を考察する。次の例は、繰り返しの数が$A2, B2$のみ2回で、Type I 平方和と Type II 平方和が異なる。

要因A	要因B	y
$A1$	$B1$	2
$A1$	$B2$	4
$A2$	$B1$	3
$A2$	$B2$	5
$A2$	$B2$	6

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ A1 \\ A2 \\ B1 \\ B2 \end{pmatrix}$$

```
data z;
   input A B y;
   cards;
1 1 2
1 2 4
2 1 3
2 2 5
2 2 6
;
run;
```

```
proc glm data = z;
   class A B;
   model y = A B / e1 e2;
   contrast 'SS1 A' A 1 -1 B 0.1667 -0.1667;
   contrast 'SS1 B' B 1 -1;
   contrast 'SS2 A' A 1 -1;
   contrast 'SS2 B' B 1 -1;
quit;
```

3.4a

Type I Estimable Functions		
	係数	
効果	A	B
Intercept	0	0
A 1	L2	0
A 2	-L2	0
B 1	0.1667*L2	L4
B 2	-0.1667*L2	-L4

Type II Estimable Functions		
	係数	
効果	A	B
Intercept	0	0
A 1	L2	0
A 2	-L2	0
B 1	0	L4
B 2	0	-L4

3.4b

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	2	9.42857143	4.71428571	16.50	0.0571
Error	2	0.57142857	0.28571429		
Corrected Total	4	10.00000000			

3.4c

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
A	1	3.33333333	3.33333333	11.67	0.0761
B	1	6.09523810	6.09523810	21.33	0.0438

3.4d

要因	自由度	Type II 平方和	平均平方	F 値	Pr > F
A	1	1.92857143	1.92857143	6.75	0.1217
B	1	6.09523810	6.09523810	21.33	0.0438

3.4e

対比	自由度	対比平方和	平均平方	F 値	Pr > F
SS1 A	1	3.33363810	3.33363810	11.67	0.0761
SS1 B	1	6.09523810	6.09523810	21.33	0.0438
SS2 A	1	1.92857143	1.92857143	6.75	0.1217
SS2 B	1	6.09523810	6.09523810	21.33	0.0438

MODEL ステートメントの E1 及び E2 オプションで Type I 及び Type II 平方和とそれらの推定可能関数を出力した（3.4a 左右、3.4c、3.4d）。推定可能関数を CONTRAST ステートメントに指定することで得られる対比平方和は、それぞれ Type I 及び Type II 平方和に一致する（3.4c、3.4d、3.4e）。

この例では Type II 平方和の方が分かりやすい。推定可能関数を$\boldsymbol{X}$の行ベクトルから

$$\boldsymbol{l}_A' = (1 \quad 1 \quad 0 \quad 1 \quad 0) - (1 \quad 0 \quad 1 \quad 1 \quad 0) = (0 \quad 1 \quad -1 \quad 0 \quad 0),$$
$$\boldsymbol{l}_B' = (1 \quad 1 \quad 0 \quad 1 \quad 0) - (1 \quad 1 \quad 0 \quad 0 \quad 1) = (0 \quad 0 \quad 0 \quad 1 \quad -1)$$

と選んだのである（3.4a 右）。データの空間でこれらに対応するベクトルを$\boldsymbol{X}\boldsymbol{c}_A$及び$\boldsymbol{X}\boldsymbol{c}_B$とする。式2.2.2を内積と考えると、$\boldsymbol{X}\boldsymbol{c}_A$と$\mu$、$B1$及び$B2$に対応する$\boldsymbol{X}$の列ベクトルが直交することが分かる。同様に、$\boldsymbol{X}\boldsymbol{c}_B$と$\mu$、$A1$及び$A2$に対応する列ベクトルが直交する。Type II 平方和はこの直交性を優先するため、$\boldsymbol{X}\boldsymbol{c}_A$と$\boldsymbol{X}\boldsymbol{c}_B$が一般には直交しない。これが要因$A$及び$B$の Type II 平方和を足しても Model の平方和と一致しない理由である（3.4b、3.4d）。

Type I 平方和では$\boldsymbol{X}\boldsymbol{c}_A$と$\boldsymbol{X}\boldsymbol{c}_B$の直交性を重視する。$X$の列ベクトルから

$$\boldsymbol{X}\boldsymbol{c}_A = \begin{pmatrix}1\\1\\0\\0\\0\end{pmatrix} - \frac{2}{5}\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} = \frac{1}{5}\begin{pmatrix}3\\3\\-2\\-2\\-2\end{pmatrix}, \qquad \boldsymbol{X}\boldsymbol{c}_B = \begin{pmatrix}1\\0\\1\\0\\0\end{pmatrix} - \frac{2}{5}\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} - \frac{1}{6}\boldsymbol{X}\boldsymbol{c}_A = \frac{1}{5}\begin{pmatrix}3\\-2\\3\\-2\\-2\end{pmatrix} - \frac{1}{30}\begin{pmatrix}3\\3\\-2\\-2\\-2\end{pmatrix}$$

と$\boldsymbol{X}\boldsymbol{c}_A$及び$\boldsymbol{X}\boldsymbol{c}_B$を作ると、アウトプットに出力した推定可能関数が

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c}_A = \frac{1}{5}\begin{pmatrix}0\\6\\-6\\1\\-1\end{pmatrix} \propto \begin{pmatrix}0\\1\\-1\\0.1666\\-0.1666\end{pmatrix}, \qquad \boldsymbol{X}'\boldsymbol{X}\boldsymbol{c}_B = \frac{7}{6}\begin{pmatrix}0\\0\\0\\1\\-1\end{pmatrix} \propto \begin{pmatrix}0\\0\\0\\1\\-1\end{pmatrix}$$

で得られる（3.4a 左）。$\boldsymbol{X}\boldsymbol{c}_A$と$\boldsymbol{X}\boldsymbol{c}_B$の直交性より、要因$A$及び$B$の Type I 平方和を足すとモデルの平方和に一致する（3.4b、3.4c）。$\boldsymbol{X}\boldsymbol{c}_A$と$\boldsymbol{X}\boldsymbol{c}_B$を作る際、明らかに要因$A$と要因$B$を対称に扱っていない。つまり Type I 平方和は要因の「投入順序」に依存する。さらに、要因Bの Type I 平方和と Type II 平方和が一致するのだが（3.4c、3.4d）、この点は、Im $\boldsymbol{X}$が3次元なので絵を描けば納得できると思う。

分散分析の「釣り合い型」、「非釣り合い型」といった用語は若干注意が必要である。繰り返しの数が異なっても Type I 平方和と Type II 平方和が一致する例を示しておく。

要因A	要因B	y
$A1$	$B1$	2
$A1$	$B2$	4
$A2$	$B1$	3
$A2$	$B1$	3
$A2$	$B2$	5
$A2$	$B2$	6

$$\boldsymbol{X} = \begin{pmatrix}1&1&0&1&0\\1&1&0&0&1\\1&0&1&1&0\\1&0&1&1&0\\1&0&1&0&1\\1&0&1&0&1\end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix}\mu\\A1\\A2\\B1\\B2\end{pmatrix}$$

```
data z;
   input A B y;
   cards;
1 1 2
1 2 4
2 1 3
2 1 3
2 2 5
2 2 6
;
run;
```

```
proc glm data = z;
   class A B;
   model y = A B / e1 e2;
quit;
```

3.4f

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
A	1	2.08333333	2.08333333	10.71	0.0467
B	1	8.16666667	8.16666667	42.00	0.0075

3.4g

要因	自由度	Type II 平方和	平均平方	F 値	Pr > F
A	1	2.08333333	2.08333333	10.71	0.0467
B	1	8.16666667	8.16666667	42.00	0.0075

要因A及びBに対応する$\boldsymbol{X}$の列ベクトルから平均方向を除いたベクトルは

$$\begin{pmatrix}1\\1\\0\\0\\0\end{pmatrix}-\frac{1}{3}\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}=\frac{1}{3}\begin{pmatrix}2\\2\\-1\\-1\\-1\end{pmatrix},\qquad \begin{pmatrix}1\\0\\1\\1\\0\end{pmatrix}-\frac{1}{2}\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1\\-1\\1\\1\\-1\end{pmatrix}$$

である。これらが直交するので、この例の Type I 平方和と Type II 平方和は一致する（3.4f、3.4g）。

3.5. 平方和のタイプ — 交互作用有り（Type II 及び Type III）

Type II 平方和と Type III 平方和は、交互作用をモデルに含めない場合は一致する。交互作用を含めたモデルでは一般に Type II 平方和と Type III 平方和が異なり、どちらを選ぶかが論点になる [5]。GLM プロシジャは特にオプションを指定しなければ Type I 及び Type III 平方和を出力するので、とりあえず Type III 平方和を選ぶといった方針はあり得る。しかしながら、解析対象に応じて検討することも必要だと思う。

要因A	要因B	y
$A1$	$B1$	2
$A1$	$B2$	4
$A2$	$B1$	3
$A2$	$B2$	5
$A2$	$B2$	6

$$\boldsymbol{X}=\begin{pmatrix}1&1&0&1&0&1&0&0&0\\1&1&0&0&1&0&1&0&0\\1&0&1&1&0&0&0&1&0\\1&0&1&0&1&0&0&0&1\\1&0&1&0&1&0&0&0&1\end{pmatrix},\qquad \boldsymbol{\beta}=\begin{pmatrix}\mu\\A1\\A2\\B1\\B2\\AB11\\AB12\\AB21\\AB22\end{pmatrix}$$

```
data z;
   input A B y;
   cards;
1 1 2
1 2 4
2 1 3
2 2 5
2 2 6
;
run;
```

```
proc glm data = z;
   class A B;
   model y = A B A*B / e2 e3;
quit;
```

3.5a

Type II Estimable Functions			
	係数		
効果	A	B	A*B
Intercept	0	0	0
A 1	L2	0	0
A 2	-L2	0	0
B 1	0	L4	0
B 2	0	-L4	0
A*B 1 1	0.4286*L2	0.4286*L4	L6
A*B 1 2	0.5714*L2	-0.4286*L4	-L6
A*B 2 1	-0.4286*L2	0.5714*L4	-L6
A*B 2 2	-0.5714*L2	-0.5714*L4	L6

Type III Estimable Functions			
	係数		
効果	A	B	A*B
Intercept	0	0	0
A 1	L2	0	0
A 2	-L2	0	0
B 1	0	L4	0
B 2	0	-L4	0
A*B 1 1	0.5*L2	0.5*L4	L6
A*B 1 2	0.5*L2	-0.5*L4	-L6
A*B 2 1	-0.5*L2	0.5*L4	-L6
A*B 2 2	-0.5*L2	-0.5*L4	L6

3.5b

要因	自由度	Type II 平方和	平均平方	F 値	Pr > F
A	1	1.92857143	1.92857143	3.86	0.2998
B	1	6.09523810	6.09523810	12.19	0.1776
A*B	1	0.07142857	0.07142857	0.14	0.7699

3.5c

要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
A	1	1.78571429	1.78571429	3.57	0.3098
B	1	5.78571429	5.78571429	11.57	0.1820
A*B	1	0.07142857	0.07142857	0.14	0.7699

$\boldsymbol{X}$から交互作用を除いた計画行列$\boldsymbol{X}_0$と交互作用を表す$\boldsymbol{\Gamma}$を

$$\boldsymbol{X}_0 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\Gamma} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

と置く。要因Aの Type II の推定可能関数は

$$\boldsymbol{X}\boldsymbol{c}_A = \boldsymbol{X}_0\boldsymbol{c}_0, \qquad \boldsymbol{c}_0'\boldsymbol{X}_0'\boldsymbol{X}_0 = \boldsymbol{l}_0' = (0 \quad 1 \quad -1 \quad 0 \quad 0)$$

を充たすと$\boldsymbol{c}_A$を見つけて、$\boldsymbol{l}_A' = \boldsymbol{c}_A'\boldsymbol{X}'\boldsymbol{X}$で求める。$(\boldsymbol{X}\boldsymbol{c}_A)\cdot\boldsymbol{y} = (\boldsymbol{X}_0\boldsymbol{c}_0)\cdot\boldsymbol{y}$となるように推定可能関数を定義するため、Type II 平方和は交互作用の有無で影響を受けない（3.4d、3.5b）。$\boldsymbol{c}_A = \begin{pmatrix}\boldsymbol{c}_0 \\ \boldsymbol{0}\end{pmatrix}$とすれば

$$\boldsymbol{l}_A = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{c}_A = \begin{pmatrix}\boldsymbol{X}_0' \\ \boldsymbol{\Gamma}'\end{pmatrix}\boldsymbol{X}_0\boldsymbol{c}_0 = \begin{pmatrix}\boldsymbol{l}_0 \\ \boldsymbol{\Gamma}'\boldsymbol{X}_0\boldsymbol{c}_0\end{pmatrix}$$

なので、$\boldsymbol{X}_0\boldsymbol{c}_0$を求めれば良い。計算すると

$$\boldsymbol{X}_0\boldsymbol{c}_0 = \frac{1}{7}\begin{pmatrix} 3 \\ 4 \\ -3 \\ -2 \\ -2 \end{pmatrix}, \qquad \boldsymbol{l}_A' = \begin{pmatrix} \boldsymbol{l}_0' & \frac{3}{7} & \frac{4}{7} & -\frac{3}{7} & -\frac{4}{7} \end{pmatrix}$$

と要因Aの Type II の推定可能関数が得られる（3.5a 左）。$\boldsymbol{X}_0\boldsymbol{c}_0$を求める際は、$\boldsymbol{X}_0$の一次独立な列ベクトルで$\tilde{\boldsymbol{X}}_0$を作り、$\boldsymbol{X}_0\boldsymbol{c}_0 = \tilde{\boldsymbol{X}}_0\tilde{\boldsymbol{c}}_0$と置き換えて行列のサイズを減らすと計算が簡単になる。

要因の繰り返しの数に影響を受けないように Type II 平方和を修正したのが Type III 平方和である（3.5c）。Xの行ベクトルから重複を除いて

$$\boldsymbol{X}_{III} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

としてから Type II 平方和と同じ手順を踏む。計算すると

$$\boldsymbol{X}_{III0} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\Gamma}_{III} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{c}'_{III0}\boldsymbol{X}'_{III0}\boldsymbol{X}_{III0} = \boldsymbol{l}'_0 = (0 \quad 1 \quad -1 \quad 0 \quad 0),$$

$$\boldsymbol{l}'_{IIIA} = \begin{pmatrix} \boldsymbol{l}_0 \\ \boldsymbol{\Gamma}'_{III}\boldsymbol{X}_{III0}\boldsymbol{c}_{III0} \end{pmatrix}' = \begin{pmatrix} \boldsymbol{l}'_0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

と Type III の推定可能関数が得られる（3.5a 右）。$\boldsymbol{X}_{III}$で重複を除いたために Type III 平方和は要因の繰り返しの数に影響を受けない。一方で、データの空間との関連が Type II 平方和に比べて薄くなる。

3.6. 平方和のタイプ — 交互作用有り（Type IV）

Type IV 平方和は、要因の水準の組み合わせがカテゴリごと欠落している場合に Type III 平方和と値が異なる。どの水準を基準にするかによっても値が違ってくる。要因Aが3水準、要因Bが2水準で、$A3, B2$の測定が行われなかった場合を例にする。REF=オプションで要因Aの水準3を基準にした結果から考える。

要因A	要因B	y
$A1$	$B1$	2
$A1$	$B2$	4
$A2$	$B1$	3
$A2$	$B2$	5
$A3$	$B1$	6

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ A1 \\ A2 \\ A3 \\ B1 \\ B2 \\ AB11 \\ AB12 \\ AB21 \\ AB22 \\ AB31 \end{pmatrix}$$

```
data z;
   input A B y;
   cards;
1 1 2
1 2 4
2 1 3
2 2 5
3 1 6
;
run;
```

```
proc glm data = z;
   class A(ref = '3') B;
   model y = A B A*B / e4;
quit;
```

3.6a（左：ref＝'3'、右：ref＝'1'）

Type IV Estimable Functions			
	係数		
効果	A	B	A*B
Intercept	0	0	0
A 1	L2	0	0
A 2	L3	0	0
A 3	-L2-L3	0	0
B 1	0	L5	0
B 2	0	-L5	0
A*B 1 1	L2	0.5*L5	L7
A*B 1 2	0	-0.5*L5	-L7
A*B 2 1	L3	0.5*L5	-L7
A*B 2 2	0	-0.5*L5	L7
A*B 3 1	-L2-L3	0	0

Type IV Estimable Functions			
	係数		
効果	A	B	A*B
Intercept	0	0	0
A 2	L2	0	0
A 3	L3	0	0
A 1	-L2-L3	0	0
B 1	0	L5	0
B 2	0	-L5	0
A*B 2 1	0.5*L2	0.5*L5	L7
A*B 2 2	0.5*L2	-0.5*L5	-L7
A*B 3 1	L3	0	0
A*B 1 1	-0.5*L2-L3	0.5*L5	-L7
A*B 1 2	-0.5*L2	-0.5*L5	L7

$A3$を基準にした$A1$の推定可能関数は、$A3$と$A1$に関連した$\boldsymbol{X}$の1、2、5行目から求める。交互作用の部分を分離して

$$\boldsymbol{X}_0=\begin{pmatrix}1&1&0&0&1&0\\1&1&0&0&0&1\\1&0&0&1&1&0\end{pmatrix},\qquad \boldsymbol{\Gamma}=\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&0&0&1\end{pmatrix}$$

と置く。なお、要因の繰り返しがある場合は Type III と同様に重複を除く。

$$\boldsymbol{c}_0'\boldsymbol{X}_0'\boldsymbol{X}_0=\boldsymbol{l}_0'=(0\quad 1\quad 0\quad -1\quad 0\quad 0)$$

を解くと

$$\boldsymbol{X}_0\boldsymbol{c}_0=\begin{pmatrix}1\\0\\-1\end{pmatrix},\qquad \boldsymbol{l}_{A1}'=\begin{pmatrix}\boldsymbol{l}_0\\\boldsymbol{\Gamma}'\boldsymbol{X}_0\boldsymbol{c}_0\end{pmatrix}'=(\boldsymbol{l}_0'\quad 1\quad 0\quad 0\quad 0\quad -1)$$

が得られる。$A3$を基準にした$A2$の推定可能関数は、Xの3、4、5行目から

$$\boldsymbol{X}_0=\begin{pmatrix}1&0&1&0&1&0\\1&0&1&0&0&1\\1&0&0&1&1&0\end{pmatrix},\qquad \boldsymbol{\Gamma}=\begin{pmatrix}0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix},\qquad \boldsymbol{l}_0'=(0\quad 0\quad 1\quad -1\quad 0\quad 0),$$

$$\boldsymbol{l}_{A2}'=(\boldsymbol{l}_0'\quad 0\quad 0\quad 1\quad 0\quad -1)$$

と得られる。要因Aの効果は、$\boldsymbol{l}_{A1}'$と$\boldsymbol{l}_{A2}'$の線型結合で定まる推定可能関数で推定できる。係数をL_2、L_3とした

$$L_2\boldsymbol{l}_{A1}'+L_3\boldsymbol{l}_{A2}'=(0\quad L_2\quad L_3\quad -L_2-L_3\quad 0\quad 0\quad L_2\quad 0\quad L_3\quad 0\quad -L_2-L_3)$$

が GLM プロシジャで得られた出力である（3.6a 左）。

REF=オプションを「ref＝'1'」にして要因Aの水準1を基準にすると、Type IV の推定可能関数は変わる。$A1$を基準にした$A2$の推定可能関数は

$$\boldsymbol{X}_0=\begin{pmatrix}1&1&0&0&1&0\\1&1&0&0&0&1\\1&0&1&0&1&0\\1&0&1&0&0&1\end{pmatrix},\qquad \boldsymbol{\Gamma}=\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\end{pmatrix},\qquad \boldsymbol{l}_0'=(0\quad -1\quad 1\quad 0\quad 0\quad 0),$$

$$\boldsymbol{l}_{A2}'=(\boldsymbol{l}_0'\quad -0.5\quad -0.5\quad 0.5\quad 0.5\quad 0)$$

である。$A1$を基準にした$A3$の推定可能関数は

$$\boldsymbol{X}_0=\begin{pmatrix}1&1&0&0&1&0\\1&1&0&0&0&1\\1&0&0&1&1&0\end{pmatrix},\qquad \boldsymbol{\Gamma}=\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&0&0&1\end{pmatrix},\qquad \boldsymbol{l}_0'=(0\quad -1\quad 0\quad 1\quad 0\quad 0),$$

$$\boldsymbol{l}_{A2}'=(\boldsymbol{l}_0'\quad -1\quad 0\quad 0\quad 0\quad 1)$$

である。これらの線型結合の

$$L_2\boldsymbol{l}'_{A2} + L_3\boldsymbol{l}'_{A3} = (0 \quad -L_2 - L_3 \quad L_2 \quad L_3 \quad 0 \quad 0 \quad -0.5L_2 - L_3 \quad -0.5L_2 \quad 0.5L_2 \quad 0.5L_2 \quad L_3)$$

が GLM プロシジャで得られた出力である（3.6a 右）。

3.7. 平方和のタイプ — Type III と Type IV の関係

要因が欠落していない場合は Type III 平方和と Type IV 平方和が一致することを説明する。これは、Type IV 平方和の特殊な場合として Type III 平方和が生じることを意味する。次の計画行列を例にする。

要因A	要因B
$A1$	$B1$
$A1$	$B2$
$A2$	$B1$
$A2$	$B2$
$A3$	$B1$
$A3$	$B2$

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\boldsymbol{X}_0 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$と置く。$\boldsymbol{X}_0$の列を組み替えて$\tilde{\boldsymbol{X}}_0 = \begin{pmatrix} 1 & 1 & -1/2 & 1 \\ 1 & 1 & -1/2 & -1 \\ 1 & -1/2 & 1 & 1 \\ 1 & -1/2 & 1 & -1 \\ 1 & -1/2 & -1/2 & 1 \\ 1 & -1/2 & -1/2 & -1 \end{pmatrix}$と置くと、Im $\boldsymbol{X}_0$ への射影行列$\boldsymbol{P}$が$\boldsymbol{P} = \tilde{\boldsymbol{X}}_0\left(\tilde{\boldsymbol{X}}'_0\tilde{\boldsymbol{X}}_0\right)^{-1}\tilde{\boldsymbol{X}}'_0$であることに注意する。要因$A$の$A1$と$A2$についての Type III の推定可能関数は、$\boldsymbol{X}'_0\boldsymbol{X}_0\boldsymbol{c}_0 = \boldsymbol{l} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$の解を用いて、$\begin{pmatrix} \boldsymbol{l} \\ \boldsymbol{X}_0\boldsymbol{c}_0 \end{pmatrix}'$である。$\boldsymbol{X}'_0\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \boldsymbol{l}$より

$$\boldsymbol{X}_0\boldsymbol{c}_0 = \boldsymbol{P}\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \tilde{\boldsymbol{X}}_0\begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 3 & -3/2 & 0 \\ 0 & -3/2 & 3 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 3/2 \\ -3/2 \\ 0 \end{pmatrix} \propto \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$$

だから、$\boldsymbol{X}_0\boldsymbol{c}_0$の5、6行目が0になる。$\tilde{\boldsymbol{X}}'_0\tilde{\boldsymbol{X}}_0$がブロック対角型で、各ブロックの対角成分と非対角成分がそれぞれ特定の値のみをとることが重要である。この場合、$\left(\tilde{\boldsymbol{X}}'_0\tilde{\boldsymbol{X}}_0\right)^{-1}$についても同じことがいえ、結局、着目している要因に関係しない$\boldsymbol{X}_0\boldsymbol{c}_0$の成分が 0 になる。3次以上の逆行列でも余因子を考えれば分かる。Type IV の推定可能関数を求めるために、$\boldsymbol{X}_0$の$A1$と$A2$に関連する行を抜き出して

$$\boldsymbol{X}_0\boldsymbol{c}_0 = \begin{pmatrix} \boldsymbol{v}_0 \\ 0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{X}_0 = \begin{pmatrix} \boldsymbol{X}_{IV} \\ \boldsymbol{X}_+ \end{pmatrix},$$

$$\boldsymbol{X}_{IV} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{X}_+ = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\Gamma}_{IV} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

と置く。$\boldsymbol{X}_{IV}\boldsymbol{c}_0 = \boldsymbol{v}_0$と$\boldsymbol{X}_+\boldsymbol{c}_0 = \boldsymbol{0}$はすぐに分かる。$\boldsymbol{l} = \boldsymbol{X}'_0\boldsymbol{X}_0\boldsymbol{c}_0 = \boldsymbol{X}'_{IV}\boldsymbol{X}_{IV}\boldsymbol{c}_0$より、$\boldsymbol{X}_{IV}\boldsymbol{c}_0$から Type IV の推定可能関数が得られる。その交互作用の部分は$\boldsymbol{\Gamma}'_{IV}\boldsymbol{X}_{IV}\boldsymbol{c}_0 = \begin{pmatrix} \boldsymbol{v}_0 \\ 0 \\ 0 \end{pmatrix} = \boldsymbol{X}_0\boldsymbol{c}_0$で Type III と同一であるから、Type IV と Type III の推定可能関数は一致し、平方和も一致する。

4. まとめ

一般線型モデルの基礎と、GLM プロシジャがその基礎に調和していることを書きたかった。データの多次元空間で観測値のベクトルから知識を引き出すには、適切なベクトルを見つけて内積を取るのが簡単で、これが推定可能関数である。一般線型モデルと GLM プロシジャについては、この観点から理解できることを示したつもりである。

その一方で、3.3 章の最小二乗平均では推定可能関数に平均値を入れる必然性が説明できなかった。Type III 平方和については、3.5 章では Type II 平方和のバリエーションとし、3.7 章では Type IV 平方和のバリエーションとしたが、もっと上手い方針があるかもしれない。Type IV 平方和については [5]の 5 章を筆者なりに再構成したが、もし間違いがあれば筆者の誤解が原因である。平方和の記述が思った以上に冗長になったひとつの原因は、計画行列の定義を SAS に合わせたからだと思う。これらの点については、より正確な知識を得ることに努めたい。

引用文献

[1] J. H. Goodnight, “GENERAL LINEAR MODELS PROCEDURE,” 1976. [オンライン]. Available: https://support.sas.com/resources/papers/proceedings-archive/SUGI76/Sugi-76-02%20Goodnight.pdf.

[2] R. C. Littell, “The Evolution of Linear Models in SAS: A Personal Perspective,” 2011. [オンライン]. Available: https://support.sas.com/resources/papers/proceedings11/325-2011.pdf.

[3] 魚住龍史, “LS-Means 再考−GLM と PLM によるモデル推定後のプロセス−,” 2014. [オンライン]. Available: https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups14-a-02.pdf.

[4] SAS Institute Inc., “SAS/STAT® 13.1 User’s Guide The Four Types of Estimable Functions,” 2013. [オンライン]. Available: https://support.sas.com/documentation/onlinedoc/stat/131/introglmest.pdf.

[5] 三輪哲久, “推定可能関数による TypeII, TypeIII, TypeIV の違いの説明,” 2015. [オンライン]. Available: https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2015-a-04.pdf.

SASの中でRを動かす

○関根　暁史
（藤本製薬株式会社）

R is usable in SAS interface.

Satoshi Sekine
Fujimoto Pharmaceutical Corp.

要旨

proc FCMP の中で Python を動かすことはできる。その Python の中で R を利用できれば、SAS インターフェース上で R を直接実行できることになる。

キーワード： proc FCMP, R, Python, PC-SAS

1. はじめに

proc FCMP の中で Python を実行できることは既に示している [1]。更に Python の rpy2 モジュールは R を動かすことができるので、proc FCMP は直接 R を実行できるということになる。この PC-SAS から R を動かすという方法での、一覧表の作成、レーダーチャートの描画、ワードクラウドの作成、標準的な臨床薬理解析、R-shiny を用いたアプリの作成法等を紹介する。

但し SAS と R を接続させるためには以下の Windows 環境変数の初期設定だけは必要となる。SAS と Python を接続させる設定も追記した。以下 3 文を X ステートメントにて各 1 回ずつ実行いただくのみで、あとは半永久的に R に繋がることとなる。

```
x 'setx R_HOME "C:/Program Files/R/R-4.5.0"';

x 'setx MAS_M2PATH "C:¥Program Files¥SASHome¥SASFoundation¥9.4¥tkmas¥sasmisc¥mas2py.py"';
x 'setx MAS_PYPATH
"C:¥Users¥[user_name]¥AppData¥Local¥Programs¥Python¥Python313¥python.exe"';
```

sashelp.class データを外部ファイルに吐き出したとする。パスは C:/Users/[user_name]を想定している。

```
libname out "C:/Users/[user_name]";
proc copy in=sashelp out=out;
  select class/memtype=data;
run;
```

次にその永久 SAS データセットを R で読み込んで、R で XPT ファイルに変換することを考える。初期設定は済ませているので、以下のように proc FCMP の中に R のプログラム(赤枠部分)をそのまま貼り付けて実行できるようになっている。X ステートメント等のコマンドラインに頼ることなく、SAS から R を直接実行できる方法である。また XPT ファイルの符号化方式は UTF-8 になるので、これは PC-SAS 日本語版のみではできない動作である。

```
proc fcmp;
declare object py(python);
submit into py;
def PyProduct():
    """Output:"""
    import rpy2.robjects as ro
    ro.r('library(haven)')
    ro.r('class <- read_sas("./class.sas7bdat")')
    ro.r('write_xpt(class,path="./class.xpt",version=5)')
endsubmit;
rc=py.publish();
rc=py.call('PyProduct');
run;
```

R の構文をそのまま貼ることができる。

SAS 側から R へ引数を渡したい場合は、以下のような構文に書き換えてしまえばよい。proc FCMP を関数化しておいて、関数を実行する際に SAS マクロ変数(ここではデータセット名)を R に渡すことができる。proc FCMP 自体をマクロ化することはできないが、関数実行部分は SAS マクロ化できるということになる。

```
proc delete data=_all_; run;
proc fcmp outlib=work.fcmp.pyfuncs;
function MyFunc(arg1 $) $100;
length Result $100;
declare object py(python);
submit into py;
def PyProduct(name):
    """Output: MyKey"""
    import rpy2.robjects as ro
    ro.r('library(haven)')
    ro.r('' + name +' <- read_sas("./' + name +'.sas7bdat")')
    ro.r('write_xpt(' + name +',path="./' + name +'.xpt",version=5)')
    return name
endsubmit;
rc=py.publish();
rc=py.call('PyProduct',arg1);
```

```
Result=py.results['MyKey'];
return(Result);
endfunc;
run;
options cmplib=work.fcmp;
%macro Xport(_name=);
data _null_;
  x=MyFunc(&_name.);
  put x=;
run;
%mend;
%Xport(_name="class");
```

2. R で一覧表を作成する

sashelp.cars データを外部ファイルに SAS で吐き出したとする。

```
libname out "C:/Users/[user_name]";
proc copy in=sashelp out=out;
  select cars/memtype=data;
run;
```

以下 R で sashelp.cars の一覧表を作成するが、プログラムは proc FCMP 部分を除いた表示(第 1 章での赤枠部分内の表示のみ)とさせていただく。自動車のタイプはトラックに絞るものとし、Origin, Make, Model, DriveTrain, EngineSize を表示する。データフレームを rlistings で一覧表化し、PDF ファイル(export_as_pdf)として出力する。

```
library(haven); library(dplyr); library(rlistings);
cars <- read_sas("./cars.sas7bdat")
df <- cars |> subset(Type=="Truck") |> select(Origin,Make,Model,DriveTrain,EngineSize)
export_as_pdf(as_listing(df),"./Truck.pdf",page_type="a4")
```

1 列目の Origin だけはグループ化されてしまうようだ(図 1)。2 列目以降をグループ化させるためにはデータフレーム(df)での加工が必要となりそうだ。変数名は変数ラベルが表示されてくれる。

Origin	Make	Model	DriveTrain	Engine Size (L)
Asia	Mazda	B2300 SX Regular Cab	Rear	2.3
	Mazda	B4000 SE Cab Plus	All	4
	Nissan	Frontier King Cab XE V6	All	3.3
	Nissan	Titan King Cab XE	All	5.6
	Subaru	Baja	All	2.5
	Toyota	Tacoma	Rear	2.4
	Toyota	Tundra Regular Cab V6	Rear	3.4
	Toyota	Tundra Access Cab V6 SR5	All	3.4
USA	Cadillac	Escalade EXT	All	6
	Chevrolet	Avalanche 1500	All	5.3
	Chevrolet	Colorado Z85	All	2.8
	Chevrolet	Silverado 1500 Regular Cab	Rear	4.3
	Chevrolet	Silverado SS	All	6
	Chevrolet	SSR	Rear	5.3

図 1. トラックの一覧表

3. R でレーダーチャートを描く

SOC 別の有害事象発現割合をレーダーチャートで表現することを考える。先ずは解析するデータセットを SAS で準備する(図 2)。

```
libname out "C:/Users/[user_name]";
data out.temp;
  length SOC $100 AE ADR 8;
  SOC="感染症および寄生虫症"; AE=82.6; ADR=43.3; output;
  SOC="代謝および栄養障害"; AE=86.0; ADR=32.2; output;
  SOC="眼障害"; AE=70.0; ADR=65.0; output;
  SOC="心臓障害"; AE=35.0; ADR=35.0; output;
  SOC="胃腸障害"; AE=84.3; ADR=62.7; output;
  SOC="皮膚および皮下組織障害"; AE=55.0; ADR=43.1; output;
run;
```

VIEWTABLE: Out.Temp

	SOC	AE	ADR
1	感染症および寄生虫症	82.6	43.3
2	代謝および栄養障害	86	32.2
3	眼障害	70	65
4	心臓障害	35	35
5	胃腸障害	84.3	62.7
6	皮膚および皮下組織障害	55	43.1

図 2. レーダーチャート描画のためのデータセット

レーダーチャートを作成するために R の fmsb パッケージを使用するが、図 2 のデータセットを fmsb がそのまま活用できる訳ではなく、データセットを転置したり軸の最小値・最大値情報を追加したりする必要がある。図 2 の項目軸の数は 6 個であったが、R のプログラムは、項目軸の数が増えても自動的に対応可能としている。

```
library(fmsb); library(sas7bdat); library(dplyr); library(reshape2); library(r2rtf);
temp <- read.sas7bdat("./temp.sas7bdat") |> mutate(seq=row_number()) |>
mutate(seq2=(ifelse(seq==1,99,seq))) |> mutate(id=1,min=0,max=100)
soc <- temp |> arrange(desc(seq2)) |> mutate(seq3=row_number())
max <- dcast(soc,id~seq3,value.var="max")
min <- dcast(soc,id~seq3,value.var="min")
ae <- dcast(soc,id~seq3,value.var="AE")
adr <- dcast(soc,id~seq3,value.var="ADR")
x <- bind_rows(max,min,ae,adr)
x$id <- NULL
png("./Radar.png",width=700,height=500,res=100)
par(mar=c(0.5,0.5,0.5,0.5))
radarchart(x,axistype=1,cglcol=gray(0.6),cglwd=0.5,seg=5,pty=c(16,17),plty=c(1,2),plwd=2.
0,pcol="black",axislabcol=gray(0.3),calcex=1.0,vlcex=1.0,vlabels=soc$SOC,centerzero=TRUE,
title="",maxmin=TRUE)
legend("bottomright",legend=c("有害事象","副作用"),col=1,pch=c(16,17),lty=c(1,2),cex=1.2)
dev.off()
filename <- "./Radar.png";filename |> rtf_read_png() |> rtf_figure() |>
rtf_encode(doc_type="figure") |> write_rtf(file="./Radar.rtf");
```

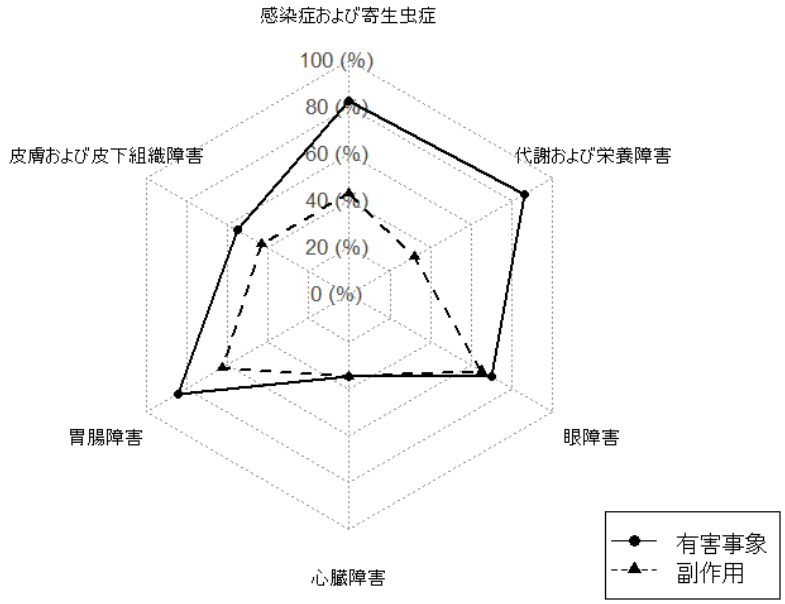


図 3.　SOC 別の発現割合を表したレーダーチャート描画

4. R でワードクラウドを作成する

PT の発現割合の大きさをワードクラウドで表現する。先ずは解析するデータセットを SAS で準備する(図 4)。

```
libname a "C:/Users/[user_name]";
data a.temp;
  length TERM $100 PCT 8;
  TERM="血中クレアチンホスホキナーゼ増加"; PCT=8.7; output;
  TERM="アレルギー性結膜炎"; PCT=21.6; output;
  TERM="便秘"; PCT=6.9; output;
  TERM="下痢"; PCT=33.0; output;
  TERM="季節性アレルギー"; PCT=17.4; output;
  TERM="挫傷"; PCT=2.7; output;
  TERM="コロナウイルス感染"; PCT=12.3; output;
  TERM="ＣＯＶＩＤ－１９"; PCT=5.4; output;
run;
```

VIEWTABLE: A.Temp

	TERM	PCT
1	血中クレアチンホスホキナーゼ増加	8.7
2	アレルギー性結膜炎	21.6
3	便秘	6.9
4	下痢	33
5	季節性アレルギー	17.4
6	挫傷	2.7
7	コロナウイルス感染	12.3
8	COVID－19	5.4

図 4. ワードクラウド作成のためのデータセット

データフレームを wordcloud パッケージにかければ、そのままワードクラウドが得られるが、割合に小数点が含まれていると不具合が生じるため、割合は四捨五入して整数値としておく。

```
library(haven); library(dplyr); library(wordcloud); library(janitor);
a<-read_sas(".temp.sas7bdat") |> mutate(pc=round_half_up(PCT,0))
b<-select(a,TERM,pc)
png("./active.png")
wordcloud(b$TERM,b$pc,scale=c(2,0.5))
dev.off()
```

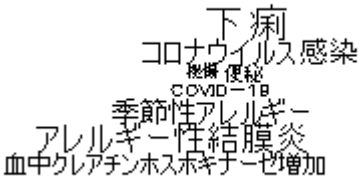


図 5. PT の発現割合の大きさを示したワードクラウド

ワードクラウドを利用して、実薬群とプラセボ群の群間比較を表現してみてはいかがだろうか(図 6)。ダミーデータではあるが、実薬群で下痢の発現割合が、プラセボ群で便秘の発現割合が高いことが見て取れる図となっている。

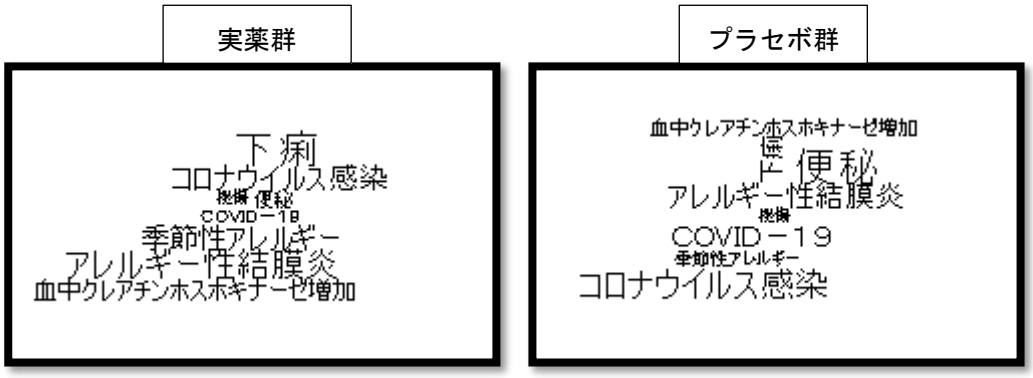


図 6. 実薬群とプラセボ群の比較(イメージ図)

5. R で標準的な臨床薬理解析を行う

ncar パッケージを使用することにより、WinNonlin と遜色のないノンコンパートメント(NCA)解析ができそうだ[2]。先ずは SAS 側で 1 症例分の血漿中薬物濃度の経時データを作ることとする(図 7)。

```
libname doc xport "C:/Users/[user_name]/adpc.xpt";
data adpc;
  SUBJID="01"; ARELTM=0.00; AVAL=0.00; output;
  SUBJID="01"; ARELTM=0.25; AVAL=3.05; output;
  SUBJID="01"; ARELTM=0.52; AVAL=3.05; output;
  SUBJID="01"; ARELTM=0.98; AVAL=7.31; output;
  SUBJID="01"; ARELTM=2.02; AVAL=7.56; output;
  SUBJID="01"; ARELTM=3.53; AVAL=6.59; output;
  SUBJID="01"; ARELTM=5.05; AVAL=5.88; output;
  SUBJID="01"; ARELTM=7.15; AVAL=4.73; output;
  SUBJID="01"; ARELTM=9.07; AVAL=4.57; output;
  SUBJID="01"; ARELTM=12.10; AVAL=3.00; output;
```

```
  SUBJID="01"; ARELTM=24.12; AVAL=1.25; output;
run;
proc copy in=work out=doc;
  select adpc/memtype=data;
run;
```

VIEWTABLE: Work.Adpc

	SUBJID	ARELTM	AVAL
1	01	0	0
2	01	0.25	3.05
3	01	0.52	3.05
4	01	0.98	7.31
5	01	2.02	7.56
6	01	3.53	6.59
7	01	5.05	5.88
8	01	7.15	4.73
9	01	9.07	4.57
10	01	12.1	3
11	01	24.12	1.25

図 7. 1 例分の血漿中薬物濃度プロファイル

24 時間のプロファイルであるが、用量は 320mg であったとして、血漿中薬物濃度の単位は mg/L であったとする。解析結果は RTF ファイルに出力する。

```
library(foreign); library(ncar);
conc <- read.xport("./adpc.xpt")
rtfNCA(fileName="./NCA-
R.rtf",conc,key="SUBJID",colTime="ARELTM",colConc="AVAL",dose=320,doseUnit="mg",timeUnit=
"h",concUnit="mg/L",down="Linear")
```

主な薬物動態パラメータについて ncar と WinNonlin の結果を比較してみたが、丸め誤差はあれど不一致はみられなかった(表 1)。

表 1. 主な薬物動態パラメータの WinNonlin との比較

Parameter	ncar	WinNonlin
CMAX (Cmax)	7.5600	7.56
TMAX (tmax)	2.0200	2.02
LAMZHL (t1/2)	8.5100	8.510038
LAMZ (kel)	0.0815	0.081451
AUCALL (AUCt)	88.5600	88.55995
AUCIFO (AUC∞)	103.9067	103.9067
VZFO (Vd)	37.8105	37.81051
CLFO (Cltot)	3.0797	3.079686
MRTEVLST (MRT)	8.3507	8.350666

6. R-Shiny による SAS ログのエラーチェッカー

SAS の実行ログを外部ファイル(UTF-8 のテキストファイルベース)として吐き出していたとする。以下の実行ログ(adae.log)には、ERROR が 1 個、エラーが 2 個、WARNIG が 1 個含まれている。

```
filename out "C:/Users/[user_name]/adae.log" encoding="UTF-8";
proc printto log=out new; run;
proc sql;
  create table AE(AESTDTC Char(10) not null);
quit;
data ADD;
  length AESTDTC $10;
  AESTDTC="";
run;
proc append base=AE data=ADD;
run;
proc sql;
  create table DM(USUBJID Char(10) unique);
  insert into DM(USUBJID) values ("01");
  insert into DM(USUBJID) values ("01");
quit;
proc printto; run;
```

```
42    proc append base=AE data=ADD;
43    run;

NOTE: WORK.ADDをWORK.AEに追加します。
WARNING: データセットWORK.AEへの追加/更新に失敗しました。
NOTE: データセットWORK.ADDから1オブザベーションを読み込み
NOTE: 0オブザベーションが追加されました。
```

図 8.　adae.log のイメージ

Shiny があれば Web アプリを簡単に作れる [3)4)]。本章では SAS 実行ログを読み込んで、エラーメッセージの個数を返すアプリを開発する。Shiny は、画面レイアウトを司っている ui、裏で計算を行っている server、アプリを起動する shinyApp から成り立っている。更にここでは session というパーツを組み込み、Web ブラウザが閉じられると R の動作が強制終了される仕組みとしている。この仕組みを採っておかないと、ブラウザが閉じられても延々と proc FCMP が回転したままという状態になってしまうからである。更に runApp(launch.browser=TRUE)とすることで、お使いの Edge が立ち上がるようにしている。

```
library(shiny); library(readr); library(stringr);
ui<-fluidPage(
  hr(),
```

```
  fileInput("file",label="ログファイル(UTF8)を選択して下さい"),
  hr(),
  actionButton("submit","エラーチェックを返します"),
  textOutput("textOut01")
)
server<-function(input,output,session){
  output$textOut01<-renderText({
    text<-read_file(input$file$datapath);
    c1=paste0("ERROR:",str_count(text,"ERROR"),"、");
    c2=paste0("エラー:",str_count(text,"エラー"),"、");
    c3=paste0("WARNING:",str_count(text,"WARNING"),"、");
    c4=paste0("欠損:",str_count(text,"欠損"),"、");
    c5=paste0("無効:",str_count(text,"無効"),"、");
    c6=paste0("欠落:",str_count(text,"欠落"));
    char=paste0(c1,c2,c3,c4,c5,c6)
  });
  session$onSessionEnded(function(){
     stopApp();
  })
}
runApp(shinyApp(ui=ui,server=server),launch.browser=TRUE)
```



図 9. R-Shiny によるログエラーチェッカー

想定通り、ERROR が 1 個、エラーが 2 個、WARNIG が 1 個返されている。こちらのチェッカーは UTF-8 版のログにしか対応していない(Shift-JIS のログでは不具合が生じる)。

7. まとめ

PC-SAS であれば、コマンドラインを使わずに R に命令することが可能であった。また SAS 側の引数を R に渡すことも可能であった。これにより、SAS が余り得意としていない、レーダーチャートやワードクラウドの作成、Web アプリ作成の機能を R で補填することができた。R の構文をコーテーションで囲まねばならない等の手間はかかるが、SAS のインターフェースの中で、SAS と R を連続的に使用することが可能となった。

参考文献

1) 関根 暁史 (2023). Python を操る FCMP プロシジャ ～SAS と Python の融合～, SAS ユーザー総会 2023 論文集
2) Hyungsub Kim (2018). Development of R packages: 'NonCompart' and 'ncar' for noncompartmental analysis (NCA), Transl Clin Pharamacol 2018;26(1):10-15
3) 水間 浩太郎 (2025). R shiny による簡単アプリ作成入門, 第 5 回ナニワデータサイエンス研究会
4) 舟尾 暢男 (2025). R shiny 超入門, 第 5 回ナニワデータサイエンス研究会

世界初の科学技術で重要なLSDの判別理論

―統計や生成AIでできない組み合わせ最適化LDF―

○新村秀一[1]、新村秀樹[2]、市川均[3]

([1]成蹊大学名誉教授、[2]イノベストメント(株)、[3]有限会社マイス(LINDO Japan 代表))

The First Discriminant Theory of Linearly Separable Data important for Science and Technology

- Combinatorial optimization LDF impossible by Statistics and ChatGPT -

○Shuichi Shinmura[1], Hideki Shinmura[2], Hitosi Ichikawa[3]

(1: Seikei University Emeritus Professor, 2: Innovestment Corp., 3: LINDO Japan)

要旨

組み合わせ最適化のRIP(Revised Optima-LDF by Integer Programming)で、以下の種々の問題を解決した。

・多重共線性のある児頭骨盤不均衡（CPD）データ（240＊19）で、RIPを2^{240}組みの240元19変数の連立不等式制約を定義域とし、データに固有の2例の誤分類をもつ一つの凸体をProgram1で特定した。2例を省けば全判別データが線形分離可能（LSD)になる（ケース選択法)。Program4で2^{19}乗の判別モデルから簡単に変数選択（FS）や検定無しで14変数の多重共線性のない最小変数のLSD（BGS: Basic Gene Set）が見つかる。BGSを含む$2^{5(=19-14)}$乗の32モデルがLSDで、524,256モデルが棄却できる革新的な判別理論である。

・Fisherが LDFの評価に用いたIris（150＊4）は3種のLSDで、14体の4変数から1変数迄のLSDでできたMatryoshka人形になる。このLSDの汎用的な構造を初めて見つけた。スイス銀行紙幣の真札/偽札（200＊6）が（X4,X6)のBGSを含む16モデルがLSDで、47モデルがLSDでない事を発見しLSD研究を開始した。

・合否判定を大問2問で50点を合格とすれば自明なLDF=T1+T2-49.5があり、自明なLSDになる。このような身近にある学生の人生に重要な合否判定がLSDである事に誰も気づかなかった。更に最小の設問のBGSで合否判定ができ、残りの設問が合否判定に関係しない。即ち設問の質の重要な研究を誰も行っていないのは教育上の重大な瑕疵である。各大学が入試問題や官庁が資格試験等の評価を行うべきである。

・優良企業と倒産企業の判別得点がAltmanスコアとして企業評価に用いられている。新しい変数を含めてRIPで見直すべきである。合否判定はこのような評価システムの一種と考えるべきだ。日経新聞が広告を取り日本の優良企業100社を掲載していたが、SASが稼働するIBM機に既存システムをリプレースする際に筆者がコンサルした。また32の製薬企業にSAS/VAXを販売し、第Ⅰ層からⅣ層臨床試験迄のソフトをSCSKで販売した。今回是非RIPで製薬企業の研究者が癌の遺伝子データ解析の理論2で癌診断の研究をしてほしい。必要な技術提供をしたい。SASを含む企業がPrecision Medicine(精密医療)の製品として遺伝子データ解析関連の製品を提供しているが、LSDでない結果は癌研究に役に立たない商品である[6]。

・小型車は座席数が4席で排気量が2000cc以下の2つのBGSで、普通車とLSDになる。恐らく多くの産業製品の分類も1変数のBGSがあると考える。統計分析の前にMatryoshka人形の構造を検討すべきだ。

・169種の遺伝子データ(約100症例＊1万遺伝子)を判別し、100変数程度の排他的なLSD(SM:Small Matryoshka)に分割でき、その中に5個以下のBGS迄の多変量の発癌遺伝子の候補を多数見つけた。

以上の実例は、統計や生成AIでは組み合わせ最適化LDFに対応できない。生成AIは画像診断や多くの専門家を代替する創薬等には役立つ。厳密解が必要な医学診断等に生成AIを用いる事は間違いである[6]。

キーワード：LSD の判別理論、生成 AI で出来ない組み合わせ最適化 LDF、癌の遺伝子データ解析（理論 2)、各種格づけ、試験の合否判定、スイス銀行紙幣の真札と偽札、小型車と普通車等の産業製品の分類、Iris データに見る動植物の種の分類、2 例の誤分類のある医学診断、間違った IP ソフトを検証する 4 データ、要員や商品等のスケジューリング、金融 Portfolio システムと SAS/IML、DEA による多入力多出力の組織や機械等の評価、真の精密医療製品、SAS の数理計画法ソフトの機能の拡大策、Sweep 演算子、多重共線性

1. 初めに

1971 年に京大の数学科の大学院に落ちて、卒業者名簿を初めて見た。数学者、高校の先生に交じって近藤次郎（文化勲章、学術会議会長）さんという先輩が東大の航空学科に入り直し、YS11 の設計から公害問題迄幅広く活躍され日本のダビンチのような先輩がおられた。NHK 出版からオペレーションズ・リサーチ（OR）の啓蒙書も出しておられる。そこで 10 月頃に SCS(現 SCSK)と NEC の役員面接で採用され SCS に入社した。しかし入社後新人教育を受けずに NEC から破格の月 40 万の派遣費で、大阪府立成人病センターで心電図診断システムの診断論理を NEC の統計ソフトの LDF と QDF で研究した。しかし野村循環器医長の開発した枝別れ論理に敵わなかった。これが医学診断に役立つ判別理論開発の動機で、52 年かけて Springer/Nature から 3 冊の判別理論の研究書[1-3]を出して LSD の判別理論を完成した。恐らく啓蒙書の販売数だけは先生を超えた。

学卒の研究者が、統計と OR で多くの実証研究し専門書や啓蒙書を出した。それは SAS と JMP を「データで表される問題解決学」の、数理計画法(Mathematical Programming)ソフトの LINDO 製品を「数式で表される問題解決学」の先生として、実証研究してきた事が成功の理由である。大学院に入らなくても、学卒でも世界の誰も行わなかった革新的な研究をしたが、企業研究者の参考になれたらと考える。

1970 年頃は第Ⅰ次統計ブームで医学データを統計分析すれば計量診断学 と言われた。筆者は、統計と MP ソフトによる 2 つの問題解決学を大学教育に取り入れる事を両学会で提案したが無視された。オバマ大統領がデータ・サイエンティストの養成を取り上げ、第Ⅱ次統計ブームで多くの大学がデータサイエンス学部を設立した。最近では統計に加えて第Ⅲ次 AI ブームの生成 AI や情報技術の問題解決が教育に取り入れられているが MP は取り入れていない。1995 年に Microarray という検査機器で、動物の**遺伝子の出す蛋白量（発現量）**が測定できるようになった。癌研究は不正防止のために高価な研究データを無償公開した。それを用いて統計、機械学習（ML)、AI 等の研究者が多くの論文を出したが LSD と指摘した研究がなく全て間違いだ。これは組み合わせ最適化を行う IP の知識が欠けていたからである。注）Word は生成 AI を使い漢字変換を行っているが、笑うような誤変換が余りにも多い。漢数字ⅠⅡⅢは内容が分かるのでそのまま使う。その他の誤変換は無駄な時間をかけて修正した。

2. Fisher の仮説で定式化した LDF の深刻な問題を２つの事実で解決

（1）多くの医学研究で企業人Ⅰ年目から研究できた幸運と判別分析の深刻な 5 問題

心電図診断や大阪府の癌の疫学研究、日本医大の CPD(児頭骨盤不均衡の 180 例の自然分娩群と 60 例の帝王切開群の 19 変数）と 32 万症例の SSM 認可のための研究、厚生省の研究者へ介護保険システムを決定木で開発する助言等の医学研究を通し、判別分析の深刻な 5 問題を見つけ Minimum Number of Misclassifications を見つける組み合わせ最適化 LDF の RIP と、使用目的の異なる RIP の 4Program をシカゴ大学発の LINGO[5]で開発した。Excel 上のデータのセル範囲を指定するだけで、RIP、ハードマージン最大化 SVM（H-SVM）と任意のペナルティｃのソフトマージン最大化 SVM（S-SVM）等が判別できる。そして多彩な判別データで驚く結果を得た。筆者は日本初の医学診断のデータ・サイエンティストの一人であろう。

問題 1 は、各種判別手法の誤分類数 NM が異なり信頼性がない。**事実 1（LDF と NM の関係）**から MNM を見つ

ける RIP がこれを解決した。ｎ件ｐ変数の 2 群データ $\mathbf{x_i}$ は、NM が 0 から(n-1)のｎ個の整数値しか取らないので組み合わせ最適化の RIP を考えた。また Fisher が LDF の評価に用いた 3 種の Iris の 4 変数が、辛うじて Fisher の仮説を満たす唯一のデータであり、5 変数以上の仮説を満たすデータがない。癌と正常のⅠ万個以上の遺伝子が正規分布と考える研究は非常識で全く役に立たない。統計や数学研究者は特異値分解で高次元に対応できる PCA と LDF を研究した。JMP が先陣を切って 2015 年の JMP ユーザー会で発表し、1 月借りて 6Array を評価したが ER は 0%でなく 1%から 17%で役に立たないことを報告した［1]。

問題 2 は、科学技術に重要な LSD の重要性が分かっているのに誰も研究しなかった。Iris に代わって判別関数の評価に用いられるスイス銀行紙幣の真札/偽札の 6 変数データの 63（$=2^6-1$）モデルを RIP で判別し 2 変数(X4,X6)が BGS であり、BGS を含む 2^4 の 16 モデルだけが LSD であることを**事実 1（LDF と NM の関係）**と**事実 2（MNM の単調減少性：$MNM_k \geq MNM_{(k+1)}$）**から見つけた。$MNM_{(X4,X6)}=0$ であり 16 モデルの MNM が 0 になる。

問題 3 は、分散共分散行列の瑕疵である。センター試験の数学IIB の合否判定で Error Rate が 30%を超え、2 次判別関数の QDF は合格群がある設問を全員正解し不合格群がバラつくと、一般化逆行列の瑕疵で合格者全員が不合格になる。筆者の指摘で JMP は合格群の分散共分散行列を全体の行列を用いることで対応した。

問題 4 は、Fisher は LDF の係数と ER の標準誤差を定義していないので推測統計手法でない。そこで判別データをｋ回コピーし擬似的な母集団の検証標本とし、乱数で並べ替えてｋ分割したｋ組の標本の学習データを考える。判別関数でｋ組の学習データを判別し、得られた LDF で検証標本を判別して係数と ER の SE が計算できる。多くのユーザーは判別係数の SE で変数選択を考えるが間違いである。判別結果を表すｋ組の学習標本の ER の平均を M1 とすれば、MNM と同じく単調減少する。検証標本の平均 ER を M2 とする。**M2 最小の Best モデル**を選ぶと Occam の剃刀を満足するモデルが得られる。発現データは全て MNM=0 で LSD であり、その中に M2=0 になる 5 個以下の遺伝子を含む BGS が「多変量の発癌遺伝子の候補」である事を指摘した。以前は Plug-in 規則で**外的基準の 2 群**を表す y_i=1/-1 という識別子をあたかも重回帰分析の目的変数として FS や検定統計量で最善と思われる FS を行ったが間違いである。外的基準は計測値より上位概念の質的変数であり重回帰分析を利用してはいけない。

問題 5 は、2025 年に多くの統計や MP ソフトが LSD を正しく判別できない事が分かり、それを確認する 2 変数の 4 データを準備した。利用者は使用ソフトが正しい結果を出すか検証してほしい。RIP と H-SVM が理論的に LSD を、最尤推定法を使った JMP のロジスティック回帰が実証研究で LSD を判別できることを確認した。他の LDF が NM=0 であれば LSD であるが、全 LSD で NM が 0 にならない（問題 1）。QDF や非線形判別手法が NM=0 であっても LSD と限らない。枝別れや決定木は分岐を増やしていけば重なりがないデータは 2 群に分けれるが LSD を意味しない。しかし多群の現実問題の解決に有用である。筆者は開発に失敗していた介護保険システムの開発に決定木を使うよう、筆者の高校の 1 年先輩の東京都の病院部門 NO.2 の D 医師にアドバイスし、彼が 3 ヶ月程でシステム化した。国会の決めた実施時期に遅れていて分析に用いていないデータでの検証を行わないで全国実施し、当初順位の逆転等の問題が指摘された。この問題を学会等で指摘した。

(2) SAS の普及に貢献した 4 名の研究者

介護保険は東大医学部の伝説の 3 秀才の開原名誉教授が厚生省の医療顧問として企画したことを 2010 年の SAS ユーザー会の懇親会で打ち明けられた。先生と私は 1977 年のトロントで開催の MEDINFO77 で **ROC** を医学診断に用いた Lusted 医師が座長のセッションで発表し、早い時期に SAS の紹介論文を共著で出した。大橋先生が開原先生の助教授時代に、筆者の 32 万例の SSM の論文[7]を医学情報学誌の編集長として採択してくれた。厚生省が SSM を認可するかもしれないと言うことで TV が医療情報学会の筆者の発表を取材予定であっ

たが、学会前日に厚生省が治験薬の続行を決め実現しなかった。恐らく開原先生の判断であろう。以上から開原先生が医学研究に SAS の普及が重要と判断されたと考える。また奥野先生と芳賀先生が日科技連の統計ソフトで品質管理のセミナーを企業研究者に行っていた。奥野先生の代理で、大橋先生が同セミナーで 1 時間の SAS の講演依頼に来社された。この 4 先生が日本製の統計ソフトから SAS へ大きく舵を切ったことが SAS 普及に貢献したと考える。蛇足になるが私が SAS の最初の代理店を SAS 社に申し込んだが日商エレに決まった。そこで世界初の SAS の計算サービスを始めた。このため塩野義から九大教授になられた浅野先生らが開発していた NISAN の普及を阻んだが、九大、岡大、北大の関係者が設立した**日本計算機統計学会（JSCS）**の発起人の末席に加えてもらった。企業の統計研究者に取って JSCS が一番適した統計学会であり勧めたい。

（3）組み合わせ最適化 LDF の RIP による 3 判別理論と目的別の 4 プログラム

RIP を LINGO の IP アルゴリズムの一つの **B&B（分枝限定）法**を用いて式（1）で定式化した。

RIP: MIN=$\sum e_i$ ； subject to $y_i*(\mathbf{b}*x_i+1) >= 1-1000*e_i$ ； ・・・（1）

n 件 p 変数の $\mathbf{x}_i$ を LDF= $\mathbf{b}*x_i+1$ で判別する。$\mathbf{x}_i$ を代入した判別得点の LDF が正(LDF>0)であれば群 1 に、負(LDF<0)であれば群 2 に判別され、逆であれば誤判別される。そして 2 群を識別する y_i =1/-1 を判別得点に掛けて拡張した判別得点を考える事で $LDF=y_i*(\mathbf{b}*x_i+1)>0$ であれば 2 群は正しく判別され MNM=0 になる。しかし 0 判定は数値計算上望ましくないので、1 以上に変えて（-1，1）にデータがないことで LSD が定義できる。誤分類がある場合を e_i=1 とし、判別される場合を 0 とする 2 値の整数変数を 240 個導入し、判別得点が $(1-1000*e_i)$ 以上で誤分類を特定できる。これで MIN=$\sum e_i$ で MNM を求める事が初めて可能になった。

次の 2 つの事実で MNM を見つける RIP を 1999 年に開発し 2015 年に最初の出版[1]で理論 1 を、2019 年に 6 種の Microarray が LSD であることを理論 2[2]で、2024 年に 52 年間で LSD の判別理論 3[3]を完成した。

・**事実 1**：誤分類数 NM と LDF の関係を発見。NM は 0 から(n-1)の整数値しか取らない組合わせ問題になる。変数が増えるほど、2 群が平均だけ異なる正規分布という仮説を満たすデータはない。

・**事実 2**：MNM=0 が LSD を定義する統計量で単調減少($MNM_k \geq MNM_{(k+1)}$)する。これが科学技術の基本である LSD 研究に重要。BGS が $MNM_k=0$ であれば、BGS を含む $MNM_{(k+1)}$ も 0 の Matryoshka 人形になる（事実 3）。

普通のデータ(n ≥ p)で次の結果を得た。RIP で CPD を判別すると、帝王切開と自然分娩の各 1 例がデータに一意に誤判別された。この 2 例を省いた 238 例が LSD になる。この「**ケースの選択法**」は全判別データを LSD で判別できる革新的な**方法 4** で判別理論の福音である。

・**方法 1**：Program2 による、通常の LOO の延長線上の間違った k 重 CV に代る正しい k 重 CV と M2。

・**方法 2**：Program3 による Matryoshka 選択法で LSD を排他的な LSD の SM に分割。

・**方法 3**：2 段階スクリーニング法(Program2 と 3 と、4 検証法の組み合わせ)で多数の M2=0 で「5 個以下の多変量の発癌遺伝子の候補の BGS」を多数見つけた。

・**方法 4**：Program1 のケース選択法で、誤分類のある全データが CPD のように LSD になる。Program4 で最小変数の BGS が見つかる。Program2 で BGS1 が M2 最小のオッカムの剃刀を満たす Best モデルになる。

CPD の RIP の定義域は 238 元 19 変数の連立不等式であるが、238 人に 0/1 の 2 値 e_i で誤分類を特定するために 2^{238} 乗の定義域から MNM になる凸体を一つ見つける。更に 19 変数の 2^{19} の 524,288 の判別モデルから最小の LSD になる 14 変数の BGS(Basic Gene Set)を見つける。選ばれない 5 変数の 32 モデルに BGS を加えた 32 モデルだけが事実 2 で LSD になり、残りの 524256 モデルは MNM が 1 以上の不要なモデルである。即ち、FS や検定統計量無しで Occam の剃刀を満たす Best なモデルになる。このような 2 段階の組み合わせの最適化問題は MP 研究でも皆無で、B&B でしか対応できないので統計や生成 AI では不可能だ。

3. Fisher の LDF と Iris データ（150 件＊4 変数）

(1)　Iris は Fisher の仮説を満たす唯一の 4 変数 3 組の LSD

Fisher は 2 群が平均だけ異なる正規分布という仮説で、2 つの正規分布の比の対数を取れば簡単に一次式になり LDF を定式化した。その後、多くの判別関数が分散共分散を用いて開発された。

彼は Iris データで検証した。図 1 の行列散布図で各 50 例の赤のセトサの G1、青のバーシクルの G2、緑のバージニカの G3 の 4 変数で 2 群が辛うじて平均だけ異なる正規分布と言える。G1&G2、G1&G3、G1&G23 が LSD で、G2&G3 が LSD でない。X3 と X4 は 1 変量の BGS で 3 群が LSD である。BGS は遺伝子データ解析の理論 2[2]で見つけた「山中 4 因子」と同じ概念の最小の LSD である。また X1 と X2 の 2 変数も BGS で、3 つの LSD が分かる。

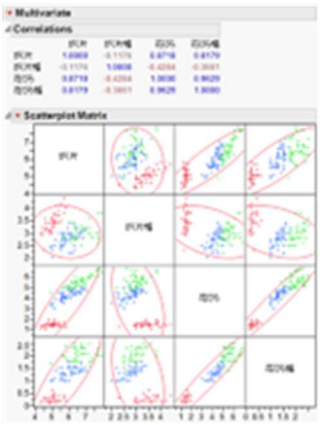


図 1　Iris の行列散布図

重回帰分析（$y_i = b*x_i+c$）は、計測値 y_i を目的変数と見なし右辺の 1 次式の計測値で予測する手法である。予測値 y_i は Gause が見つけた正規分布（Gause）になり FS や検定統計量が開発された。LDF は同じ 1 次式で量的な計測値の上位概念の外的基準を予測する記述手法である。Plug-in 規則で y_i＝1/－1 の識別子を目的変数と考えて重回帰分析すれば FS や検定統計量が利用できる。しかし Fisher の仮説を満たすデータは、この 4 変数の Iris しかなく、適用事例のないお伽話である。更に言えば、癌と正常の 1 万以上の遺伝子が仮説を満たすと考える事は異常な考えである。残念ながらこの仮説に立った精密医療の製品が多いのが現状で人類が願う正しい癌診断の発展の妨げる[6]。

(2)　Iris の LSD は 14 体の **Matryoshka 人形（事実 3）**

理論 2 で高次元の 169 の Array が LSD で、膨大な Matryoshka 人形になる LSD の重要なデータ構造(事実 3)を見つけた。表 1 の 14 体の Matryoshka 人形で説明する。4 変数が LSD で、その中に 3 変数の LSD、更に 2 変数の LSD を含み、最後に X3 か X4 の BGS を含む 12 体がある。そして X1 と X2 の BGS で終了する 2 体の計 14 体の Matryoshka 人形になる。

表 1　14 体の Matryoshka 人形

Matryoshka1: (X1, X2, X3, X4) ⊃(X2, X3, X4)⊃(X3, X4)⊃(X3),

Matryoshka2: (X1, X2, X3, X4) ⊃(X2, X3, X4)⊃(X3, X4)⊃(X4),

Matryoshka3: (X1, X2, X3, X4) ⊃(X2, X3, X4)⊃(X2, X4)⊃(X4),

Matryoshka4: (X1, X2, X3, X4) ⊃(X2, X3, X4)⊃(X2, X3)⊃(X3),

Matryoshka5: (X1, X2, X3, X4) ⊃(X1, X3, X4)⊃(X3, X4)⊃(X4),

Matryoshka6: (X1, X2, X3, X4) ⊃(X1, X3, X4)⊃(X3, X4)⊃(X3),

Matryoshka7: (X1, X2, X3, X4) ⊃(X1, X3, X4)⊃(X1, X4)⊃(X4),

Matryoshka8: (X1, X2, X3, X4) ⊃(X1, X3, X4)⊃(X1, X3)⊃(X3),

Matryoshka9　: (X1, X2, X3, X4) ⊃(X1, X2, X4)⊃(X2, X4)⊃(X4),

Matryoshka10: (X1, X2, X3, X4) ⊃(X1, X2, X4)⊃(X1, X4)⊃(X4),

Matryoshka11: (X1, X2, X3, X4) ⊃(X1, X2, X4)⊃(X1, X2),

Matryoshka12: (X1, X2, X3, X4) ⊃(X1, X2, X3)⊃(X2, X3) ⊃(X3),

Matryoshka13: (X1, X2, X3, X4) ⊃(X1, X2, X3)⊃(X1, X3) ⊃(X3),

Matryoshka14: (X1, X2, X3, X4) ⊃(X1, X2, X3)⊃(X1, X2).

普通の LSD（n >=p）は、この様に p 変数が LSD であれば p 変数から各 BGS 迄の Nest 構造になる。

一方 Array のような**高次元の LSD（n <<p）**を当初は統計、ML、AI 等の理工学研究者は高次元データは「3 つの困難な問題」があり研究が困難と考えた。単に高次元 Array は、2015 年の JMP ユーザー会で Sall 博士が SVD を使って開発した高次元の LDF と PCA の紹介で**横長データ（n＜ｐ）**と呼んだ事が正しい。理工学研究者は、自分の貴重な研究に高校数学の n 元 p 変数の連立方程式は、 p 変数から n 変数を選んで n 行 n 列の逆行列で解が得られる事実を重要な研究に生かせず間違った研究を続けている。これは理工学教育の深刻な問題である。

理論 3 の研究で、Program3 で横長の LSD は n 個以下の遺伝子の排他的な SM に分割され、残りの僅かな遺伝子の組みは LSD でない。各 SM は n 個以下から各 BGS 迄の LSD を含む膨大な Matryoshka 人形になる大きな違いがある。山中先生は 3 万個以上という遺伝子から、Rat の ES 細胞で活性化する 24 遺伝子を遺伝子 DB で特定し、3 万の変数選択に成功した。そして 24 遺伝子の組みから山中 4 遺伝子まで、1 遺伝子を省いた 23 遺伝子で LSD になれば省いていく変数減少法で山中 4 遺伝子が万能細胞を作ることを発見した。もし BGS が 1 個であれば膨大な Matryoshka 人形になり、重要な万能細胞に適さない。一方、癌は正常細胞の暴走であり、「最大 2,846 組みの 5 個以下の M2=0 である多変量の発癌遺伝子の候補」を見つけた[3]。この中に既に癌研究が見つけた 1 変量の発癌遺伝子が含まれておれば、多変量の発癌遺伝子の特徴になる（検証 4）。

4. 理論 3：Fisher の判別理論を超えた RIP と 4Program

(1)　RIP を n 元 p 変数の連立不等式で説明

理論 1[1]で RIP（MIN=$\sum e_i$ ； st　$y_i*(\mathbf{b}*x_i+1) >= 1 - 1000*e_i$ ；）の Program1 を公開した。誤分類の e_i は 1，判別できれば 0 とする e_i＝1/0 の整数値で誤分類データにも対応できる組み合わせ最適化 LDF の RIP を定式化した。RIP は n 例 p 変数の x_i を代入した判別得点に 2 群を表す y_i を掛けて拡張した判別得点が 1 以上であれば正しく判別される。そして**開区間(-1, 1)に x_i が無い事が LSD** である。しかし統計や AI はデータに一意に x_i が誤判別される事に対応できない。IP を知っておれば、1 を（1 - 1000*e_i）とすれば判別得点が－999 以上であれば誤判別例にも対応できる。しかし n 個の e_i の 2^n 組みの制約式の表す凸体の中から MNM になる一つの凸体を選ぶ必要がある。そして目的関数$\sum e_i$ の最小化で MNM が見つかる。もし MNM＝k であれば、k 例を省いた(n-k)が LSD になり、LSD の Program4 が利用できる。また定義域は n 個の判別超平面（$\mathbf{b}*x_i+1=0$）が p 次元の LDF の係数空間を判別得点が正と負の半平面に分割した 2^n 組みの制約式でできる凸体から、MNM をもつ最適凸体の SM を一つ選ぶ。そして Program4 で SM から 2^p の組み合わせで BGS を見つける。即ち、2^n と 2^p の最適組み合わせの LDF は、AI や統計ではできない精密医療等を可能にする。商用の MP ソフトでも、このような組み合わせ最適化を分枝限定法で確実にできるのは LINGO しか実績が無い。

(2)　誤分類のある CPD（児頭骨盤不均衡）データの「ケース選択法（方法 4）」

Program1 の RIP で CPD を判別すると帝王切開と自然分娩の 1 例が誤判別された。この 2 例を省くと全判別データが LSD になる。Program4 で、19 変数の 238 元 19 変数の連立不等式の 2^{238} 組みの凸体から LSD になる最適凸体の定義域を B&B 法で一つ選ぶ。定義域は 19 次元以下の凸体であり、2^{19} の 52 万個以上の凸体から最小次元の 14 変数の LSD の BGS を一つ選ぶ。この 2 つの組み合わせ最適化を同時に B&B 法で行うが、将来量子計算機が代替できる。商用の IP ソフトでも対応できないソフトがありソフトの信頼性を調べる 8 データ（5＊2）を準備した。もし SAS の MP ソフトが B&B に対応すれば、現在 IP ソフトの大きな応用分野の人や物の最適スケジューリング等にも対応でき、その後に統計処理ができてユーザーに歓迎される。

鈴村日本医科大学名誉教授が 17 変数の計測値から X12=X13-X14 と X9=X7-X8 の 2 つの等式で X12 と X9 を作り、この 2 変数の値で X 線写真の画像から出産前に出産方法を決める鈴村氏法を開発した。これを統計で証明する医学論文を 1983 年に依頼された［1］。SAS が Goodnight 博士の **Sweep 演算子**で重回帰分析の 2^{19} 個のモデルから、1 変数から 18 変数の決定係数が大きい上位 19 モデル(=19*18+1)を出力する **RSQUARE** を提供していた。筆者はその結果を踏まえて、SAS と JMP で変数増加法と減少法を行い、AIC、BIC、Cp 統計量で多重共線性のない 16 変数モデルを半年かけて選んだ。また Sall 博士の“SAS Regression Application”の多重共線性と Goodnight 博士の Sweep 演算子が分かるユニークな書籍[4]を出版した。

(3)　LSD の検証法と判別手法の 4 分類

Array の LSD の階層構造を評価する次の 4 検証法を開発した。検証 1 は、MNM=0 が LSD を確認できる唯一の統計量だ。検証 2 は RatioSV(＝SV の距離/判別得点の範囲)で線形分離度を定義するが、例数の少ない BGS では頑強な統計量でない。検証 3 は Ward 法と PCA で SM と BGS の線形分離度を 7 タイプで評価。LSD であっても、2 群が 2 クラスターに分かれない Array があるが、多分症例設計が悪いからと考える。検証 4 は M2=0 で 5 個以下の遺伝子の BGS を多変量の癌遺伝子の候補と考えた。これに遺産的な 1 変量の発癌遺伝子が含まれているかを調べ、その特徴で多変量の発癌遺伝子の候補の特徴を研究できる事を提案した。検証 5 は、医師が癌研究で行っている Genome Cohort と生存時間解析を 4 検証の後で行うべきと考える。

各種の判別手法を NM と ER の信頼性で次の 4 分類に分けた。ML 研究が教師無し手法のクラスター分析を Classifier として判別手法と同一視するのは間違いである。癌患者と正常あるいは異なった癌腫は、医学的知識で決まる外的基準で、それを発現量で判別するデータ解析の基本知識が ML 研究者に必要である。科学と工学は、質の違う手法等を ER の信頼性で厳密に分類する必要がある。

分類 1 は RIP と H-SVM の ER＝0 だけが、理論的に LSD を示す。ロジティック回帰は、全 SM の ER が 0 である事を実証分析で確認した。分類 2 は、その他の LDF は NM=0 であれば LSD であるが、全 LSD が NM=0 でない（問題 1）。**分類 3** は、QDF や非線形判別関数は NM=0 であっても LSD でない場合が多い。**分類 4** は、分類木等の枝別れ法の判別境界が複雑で、2 群のケースが重ならない場合は LSD でなくても NM=0 になる。しかし**ECG 診断や介護保険システムのような実システムの開発**に適している。結局分類 1 と分類 4 が実用に適している。他の手法は研究テーマである。癌研究者は非線形クラスターで多群を分けている。この場合重要な 2 群が分かれれば、**階層型クラスター分析で 2 群が 2 クラスターに分かれるか確認**すべきである。

(4)　**4 種の LINGO の Program と 2 段階スクリーニング法（方法 3）**

Program1 は、RIP、H-SVM、S-SVM、LP-OLDF、IPLP-OLDF 等の LDF を選択実行する。RIP は MNM を見つけ、MNM=k であればケースの選択法で(n-k)例の LSD になる。Program4 は全 LSD を判別し BGS を見つける判別理論の革命手法である。Program2 の 10 重 CV の M2 の評価で、BGS1 は最小の M2 値の Best モデルになる事が多く Occam razor を実現する。2004 年以降の第 2 世代の Array は、M2=0 になる SM や BGS だけに注目した（検証 3）。Program3 と 4 は、LSD を排他的な多数の SM と BGS に分割し理論 2 で重要である。第Ⅰ世代の 6 医学研究の Array は品質が悪く成果がなく、第 2 世代ではデータ解析的なアプローチが否定され昔ながらの 1 変量の発癌遺伝子の組み合わせで癌研究が行われている。しかし**発現量は人類が出会った最高品質の計測値**で 169 の Array が LSD で、**Program4 で多変量の発癌遺伝子を簡単に見つける**。それらに含まれている**遺産的な 1 変量の遺伝子が多変量の発癌遺伝子の候補を特徴付ける（検証 4）**。

次の方法 3 で、志のある高校生でも短時間で多変量の発癌遺伝子の候補が簡単に見つかる。将来のノーベル医学賞を狙う高校の教育の助言は喜んで行いたい。

Step1: Array を 12,625 個の遺伝子の組に分け Program3 で SM 分割すると計算時間が早くなる。10 個の SM がほぼ遺伝子数が 2000 個以下であり、Program4 で BGS 分割する。Program4 は 2000 個以上で BGS 分割すると計算時間が急に増える欠点に苦しんだ。

Step2：検証 3 と 4 を満たす Array を詳しく Step2 で分析。医学研究は 5 個以下で M2=0 の重要な BGS がない場合、コントロール群になる正常と癌が同数でない等の症例設計が悪いと考えられる。また症例数が 50 個前後の Array の結果が良いという常識に反する事が分かった。これは研究目的の違う癌症例や癌年齢に近い正常例等を集めたためと考える。NIH の肝癌は 181 例の癌と 176 例の正常例のほぼ同数のデータである。

M2=0 になる SM や BGS がなく Ward 法と PCA でも 2 群に分かれない。症例を省いていくと 50 例前後が良い結果になる。但し 30 例以下にすると Program2 で M2 が 50%の停止則が働く。これは遺伝子の出す蛋白量が癌と正常を 2 群に分ける人類が出会った最高品質の計測値のためと考えた。

表 2　16 モデルが LSD

p	Var.	RIP	SVM1	LDF	QDF	RDA
1	1−6	0	1	1	1	1
2	2−6	0	1	1	1	1
3	1, 3−6	0	1	1	1	1
4	1, 2, 4−6	0	1	1	1	1
5	1−4, 6	0	2	1	1	1
6	3−6	0	1	1	1	1
7	2, 4−6	0	1	1	1	1
8	1, 4−6	0	1	1	1	1
9	2−4, 6	0	2	1	1	1
10	1, 3, 4, 6	0	2	1	1	1
11	1, 2, 4, 6	0	2	2	1	1
12	4−6	0	1	1	1	1
13	3, 4, 6	0	2	1	1	1
14	1, 4, 6	0	2	2	2	1
15	2, 4, 6	0	2	1	1	1
16	4, 6	0	2	3	1	1

5　普通の LSD の事例

5.1　スイス銀行紙幣は判別の第 2 の評価データ

Flury と Rieduyl(1989)は真札と偽札各 100 枚の 6 変数で判別分析の本を書いた。統計ソフト等で公開され、多くのユーザーが Iris に代わる評価データとして分析。しかし散布図で LSD が判定できないので竹内啓先生も気付かなかった。6 変数の 63 モデルを RIP と H-SVM と Penalty c=1000 とした S-SVM3 とロジスティック回帰が表 2 の RIP 列のように（X4, X6）が BGS で、これを含む 16 モデルが LSD である。S-SVM1（c=1）、LDF、QDF、RDA は、NM が 1 以上で LSD でない。また MNM が単調減少する事実 2 を見つけた。事実 2 で残りの 4 変数に BGS を加えた 16 モデルが LSD で、47 モデルが LSD でないことが分かる。Program2 の 10 重 CV で、BGS が **オッカムの剃刀を満たす** M2 が**最小の Best なモデルである。即ち、変数選択と検定無しでモデル選択ができる。**しかし Iris と違い、M2 が 0 でない。これは真札／偽札を決めるインクの質やその他の偽造紙幣の特徴の外的基準を示す変数を用いず、紙幣の長さでは判定できないことを示す。Iris が 4 変数で簡単に LSD が分かったのは、種別が遺伝子の発現量に影響されるこれらの 4 計測値を用いたからと推測する。即ち動植物の種の研究は、癌研究と同じく単純な計測値で解決できると考える。

ロジスティック回帰で 6 変数を判別すると LSD である。**Firth（1993）**が指摘するようにロジスティック回帰係数の SE は大きくなり、統計的には棄却すべきモデルだが、図 2 のように JMP は ROC で 2＊2 の分割表を出力し NM=0 である事が分かる。最尤推定法を使った JMP のロジスティック回帰で 1000 個以上の LSD の SM で NM=0 を確認したので、経験的に LSD を判別できると結論した。RIP と H-SVM だけが理論的に LSD が分かる。しかし、H-SVM は LSD でないデータを判別できないし、RIP のように Shipp(77*7129)の遺伝子データを判別し、Ｉ秒で LSD であり、32 個の遺伝子の係数だけが非零の SM1 になる機能（機能を ~~Word は 2025/03/26 と誤変換~~）はない。一部の統計研究は LASSO に FS を期待したが例え数百個程度の係数を 0 にできたとしても癌診断に役立たない。

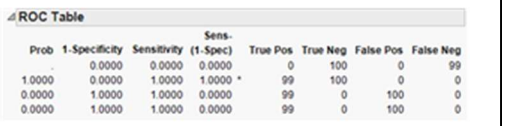


図 2　ROC による NM＝0 の確認

5.2 合否判定：身の周りにある自明な LSD は各種格づけの雛形

(1)　概略

銀行データで LSD 研究の重要性に気づき、ネットや書籍で LSD を探したが見つからない。暫く考えて合否判定が簡単に入手できる LSD である事に気づいた。大問 2 問の得点合計で 50 点を合格点とすると自明な

LDF = T1 + T2 − 49.5 が自明な LSD になる。これほど LSD 研究に適し、簡単に入手できるデータは無い。しかも多くの人生に影響するが誰も設問の質の研究をしていない。筆者が世界で最初に合否データを判別した。更に BGS に選ばれる設問だけが合否判定でき、残りの設問は合否判定に役立たない。各大学は教員の試験問題の評価は教員の同意を得られないが、入試問題の検討に BGS を利用すべきだ。JMP と LINGO ソフトの使用法を含めて大学単位や企業の部門単位で 3 時間程度の説明で簡単に実施できるので協力したい。

また筆者が 30 代の頃、NY 州立大学の Altman が、優良企業と倒産企業を LDF で判別した **Altman スコア**と呼ばれ判別得点が企業評価に広く利用されている。しかし LDF には問題があり、利用されていない変数を加えて Program4 で見直すべきであろう。日経新聞でも、一時期企業広告を伴う日本企業 100 社の優良企業ランキングを行うために SAS が使える IBM 機を導入することにし、筆者がコンサルした。日経リサーチの研究者が SAS 手法を動員し CASMA になった。この様に**合否判定は各種の格づけの雛形**として重要である。

そこで 2010 年に大学センター試験の研究員に貸し出しを依頼しセンターで研究者とセンター長に研究内容を説明し、3 年間 13 教科の本問題と予備問題の研究データ(試験当日、東京圏の旧 1 期校の現役 1 年生で実施したデータ）を貸りた。1 年間、90%、50%、10%点を合格点とした判別で次の問題点が分かった。

・合否データは例え分布が正規分布であっても、T1 + T2＝ 49.50 の判別超平面で LSD になり、Fisher の仮説を満たさない。僅か I 点差で合否が分かれるので、判別超平面で向かい合う合格と不合格者が最初のクラスターになり易く Ward 法と PCA の散布図で 2 群が 2 クラスターに分かれない点は Iris データと異なる。

・LDF は、数学IIB の 90%で ER は 32%を超える事がある。また一般化逆行列の瑕疵で、QDF は全合格生がある設問を正解し不合格群がバラつく場合、不合格群に誤判別される。JMP は対応済みでこれを解決した。

・90%の数学IIB が、偶然自明な LDF の係数になった。

・2〜3 組みの BGS が得られ、必ずしも BGS1 が Best モデルでない。

・しかし 90%点は、何れも BGS1 が M2＝0 で Best なモデルになる。一方、50%点は 1 点差で合否が入れ替わりが多いためか BGS1 が Best モデルでない試験も多い。

以上の点は、設問が発現量ほど高品質でないことを示すが BGS の検証は設問の品質向上に役立つ。

入試や資格試験の出題者等の関係者は、BGS とそれ以外の設問を検証する義務がある。BGS に選ばれた設問の組みと、BGS でない設問を出題者が検討し改善する必要がある。人の一生に影響する入試や資格試験の合否判定は、全て BGS で見直す必要がある。

某私大の問題の配布ミスで、朝日新聞が民社党の菅政権筋がセンター改組を考えているという線香花火のスクープ記事が出た。センターから、投稿済みの 3 本の論文を一次取り下げする事を依頼された。また前年に成蹊大学でセンターと応用統計学会共催で、研究用データを広く開放し合否判定の研究を促進する会議を行った[8]。この計画も頓挫してしまった。

そこで経済学部の約 130 人の 1 年生を対象とした統計入門で 10 択 100 問のマークシート試験を行い、研究を続けて経済学部の年 2 回刊行の論集に 12 の成果を発表したので参考にしてほしい。

(2) センター試験の数学IIB の見過ごせない結果

センター試験の分析結果の概略は、90%の合否判定で英語が 59 問中 83%の 49 問が 1 個の BGS になる。この対局が数学IIB で、複数の BGS が見つかり BGS1 は 129 問中 12%の僅か 16 問である。図 3 の Ward 法と PCA の散布図で学生は大きく 2 群に分かれるが、これは設問群が明らかに異なる理解力を評価していると考える。しかし出題者はこの事実を予見していないと考える。明らかに上の学生は、原点を含む多くの学生と

異なった数学の設問が得意な小数の学生である。この状況を出題者が検討する事が、高校の数学教育にとって今までと違う発見があるはずである。またセンターの研究員は 2011 年の決定を再始動すべき。

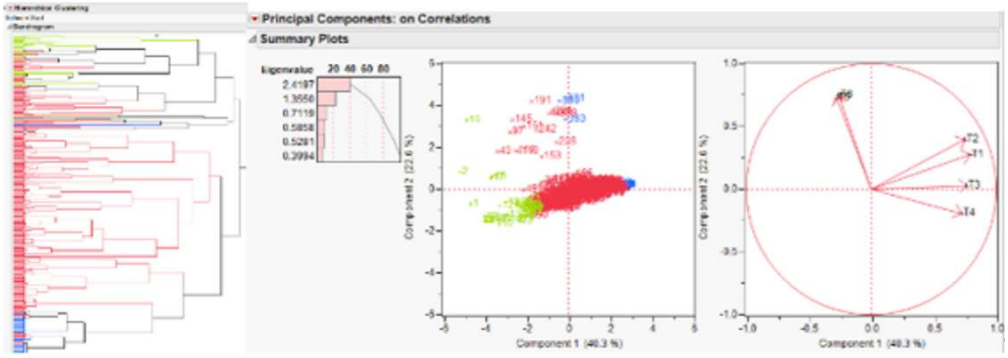


図 3　124 人の 4 大問の Ward 法と PCA の結果

5.3　普通車と小型車

小型車 15 車種と普通車 29 車種(44＊6) は大学院の 2 単位の統計演習の 50 頁以上のレポート試験で提出されたデータである。車が好きな女性で、自分で排出量（X1）、価格、座席数（X3）、CO2、燃料、売上高の 6 変数を選んで、主として売上高を残りの 5 変数で重回帰分析した。判別分析が自分の主要研究テーマであり、日科技連の『**最適判別関数（2010）**』を教えようと思ったが、本人の就職先の事を考えて学部の 1 年生に用いた統計分析の入門の『**JMP 活用　統計学とっておき勉強法（2003）**』をテキストに PC 実習した。

小型車の座席数が 4 席で 2000cc 以下の排気量という産業規格で普通車と区別される。座席数がモデルに入ると QDF の合否判定と同じく全小型車が普通車に誤判別される。即ち他の産業製品も、自動車が 2 個の 1 変量の BGS があるように、自明な BGS があると考える。また Isis の種別は遺伝子で決まるように、多くの動植物の種と同じく産業製品の種別も 1 変量の産業規格で LSD になる。理論 3 は動植物や産業製品の研究でも、iPS 研究や癌研究同様に明確な結果が期待できる。

この様に 2 個の 1 変数の BGS がある場合、表 3 のように Program2 で 63 モデルを分析するだけで BGS や SM が容易に分かる。また X3 で小型車が 4 座席と一定値であり、QDF は全車種を普通車に誤判別する。

表 3　10 重 CV で自動車の 63 モデルの評価

SN	Model	Var.	MNM	M2	x1	x2	x3	x4	x5	x6	C
1	Full	1-6	0	0.023	5.92	0	0	0	0	0	-4.893
23	SM1/BGS1	3	0	0			2				-9
48	BGS2	1	0	0	5.92						-4.893
49	bgs3	2,4-6	3	1.295		0		-46	-199	0	5342.8
60	sm3	4,5	10	2.841				3.54	10.8		-601.3
63		6	13	3.205						-0.7	6773.5

SN=1 の Var.列の 1－6 は、RIP の判別で「X1 から X6 と c」行の判別係数から X1 と C だけが選ばれることを示す。SN=48 の BGS2 は RIP=5.92＊X1-4.893 になる。SN=23 は RIP=2＊X3-9 になり、X3=4.5 が判別超平面になり 4 席と 5 席の平均（重心）である。理論的に RIP が最適凸体のどの点を選ぶか分からないが重心かもしれない。MNM 列は X1 か X3 を含む 48 モデルが LSD を示す。M2 は SN=23 と 48 の X3 と X1 の 2 つの BGS が M2=0%の Best モデルになる。SN=1 は 6 変数もあるのに X1 だけが選ばれたが、M2=0.023%で Best モデルでない。他の判別関数の係数は、6 変数とも異なった尤もらしい値になり、6 変数の判別分析の優劣が分からな

い。またこれ迄無限の判別関数の存在が自然だと間違って理解してきた。正しい認識は同一症例を誤判別する無限の LDF は同じ制約式の凸体の内点であり数学的に同値と考えるべきだ。即ちｎ例の NM は 0 から(n-1)迄のｎ個の整数値しかないが、RIP は 2^n組みの制約式から MNM に対応する最適凸体の一つを定義域に選ぶ。

6　終わりに

(1) 学卒研究者ができた世界初の革新的な成功

1971 年に大学院に落ちて先輩の近藤次郎先生が航空機の専門家に留まらず OR や環境問題の専門家であり多くの啓蒙書も執筆されていることを知り、できたばかりの SCS に入社した。そして NEC と野村循環器医長との心電図自動診断システムや鈴木先生との癌の疫学診断で学卒 I 年目から NEC の統計ソフトで研究を行う幸運に恵まれた。SCS に戻り、SAS 社が米国商務省の企画でその他の企業と代理店獲得に来日した。配属先の企画の本棚に SAS88 という薄い青のマニュアルがあり読んで驚いた。データの管理を行う**DATA ステップと分析を行う PROC ステップの革新性**に惹かれて代理店取得を申し出たが日商エレに負けた。ビルトッテン氏と共に Goodnight 社長を訪問したが契約済みで難しいことが分かり、その場で**世界初の SAS の計算サービス**を日本で行う同意を得た。そして統計ソフト等の理工系ソフトを利用し実証研究を行う方が統計全版の理解の早道と気づいた。そこで**SAS とシカゴ大学で開発された LINDO を統計と MP の個人家庭教師**として多くの論文と専門書と啓蒙書を出した。また東京都の研究者の丹後先生らがつくった統計ソフトの SAS と比べた優位性の論文に対して情報処理学会誌に反対論文を出した。OR で知り合い東工大に LINDO を導入した鈴木筑波大学副学長から、大学研究者の間で私のことを SAS 坊やと揶揄していることを指摘され驚いた。

4 年後にアシストが代理店になり、旧知の T 社長からミニコン版の SAS は IBM 版より安いので営業が販売しないので、ミニコン版の代理店の依頼を引き受けた。**SAS の医療営業の S 氏と連携し市川課員と 3 人で 32 の製薬メーカーに SAS/Vax を販売し、日本最強の企業統計家集団の誕生に貢献**したと思う。東洋信託銀行に Vax の最上位機種と SAS と LINDO、日経 Need 等の億円システム販売し、**製薬企業に加えて日本の投資分析に貢献したかった**。しかし金融担当の K 氏が**SAS/IML で投資分析**をサポートする企業と手を組みミニコン版の代理店契約が解消になり実現できなかった。三菱銀行が IML で巨額の資金で投資分析ソフトを開発し他行への販売を考えていた。しかし多くの事例で組み入れ比率の和が 100%にならず 0 になるという相談を SAS ユーザー会で知り合った A 氏から調査依頼を 150 万円で受け、LINDO で正しく分析できる事を示した。この好機を利用し LINDO で金融に投資分析を開拓する事も考えたが意欲がなく諦めた。

これ迄以下の提言を SAS と JMP にしてきたが、これを最後にしたい。市川氏 (LINDO Japan 代表) から LINDO 製品の開発に使われている C ライブラリーを購入し、SAS の MP ソフトを最先端製品にする。あるいは JMP も MP ソフトを提供する。リスク無しで、自社開発する膨大な費用と年月が削減できる。現在 IP を用いた「要員等のスケジューリング問題」は、大手流通や製造業や大学病院等の看護師要員計画問題で多数の実績があり、それが SAS の目の前にある大きな市場である。更に筆者の開発した「組み合わせ最適化の LDF」は、幅広い問題を解決できる。その中で癌研究は精密医療に AI で活路を見いだそうとしているが間違いである[6]。AI で厳密解が求まらないのに、多くの企業は遺伝子データ解析の役に立たない商品を出している。もし SAS 社が B&B 技術の組み合わせ最適化で RIP を開発すれば多くの企業で役に立つ。特に遺伝子データに限っても、医学部や製薬企業や遺伝子工学の研究機関に直ぐに SAS や JMP が役に立つ未開の分野が開拓できる。このために統計と MP ソフトが LSD を判別できるか否かを確認する必要がある。

(2) 統計と MP ソフトが LSD を判別できるか否かの簡単な検証データ

表 4　2 変数の 4 判別データ

		x1	x2	y_i	SCORE
	RIP1	4	1	-6	MNM=1
A	1	1	3	1	1
	2	2	2	1	4
	3	2	-1	1	1
	4	1	1	-1	1
	5	2	1	-1	-3
	RIP2	2	2	-7	MNM=0
B	1	1	3	1	1
	2	2	2	1	1
	3	3	1	1	1
	4	1	1	-1	3
	5	2	1	-1	1
	RIP5	-4	-1	0	MNM=0
E	1	-1	-1	1	5
	2	0	1	-1	1
	3	1	-3	-1	1
	RIP6	-0.5	0.5	1	MNM=1
F	1	-1	-1	1	1
	2	1	1	1	1
	3	0	1	-1	-1.5

表 4 は x1 と x2 が説明変数で、y_i=1/-1 が群 1/2 の外的基準を表す識別子である。データ B と E が MNM=0 で、A と F が MNM=1 である事は散布図を描けば簡単に確認できる。RIP 列は RIP の判別係数示す。そして使用する IP と統計ソフトで RIP と同じ結果が出れば、安心して世界最先端の研究に参画できる。恐らく多くの統計と IP ソフトが正しい解を出さないと考える。実際商用ソフトで、サプライ・チェーンや製造等の出荷管理の最適化で厳密な正しい最適解を出さないことが分かっている。それであれば生成 AI でアバウトな解を出すのと変わらない。このデータを JMP で判別した結果を表 5 に示す。使用する IP と統計ソフトで RIP と同じ結果が出れば、安心して世界最先端の研究に参画できる。恐らく多くの統計と IP ソフトが正しい解を出さないと考える。SAS が RIP を提供すれば、多くの既存の SAS ユーザー企業と研究者に貢献できる。また人類の癌撲滅の悲願に一歩近づける Exact Science に基づく医療システムを販売してほしい。念のため Shipp のデータで計算速度を確認した方が良い。

表 5　4 データの 3LDF の MNM と NM の結果

	A	B	E	F
RIP	1	0	0	1
LDF	2	0	1	2
Logistic	1	0	1	1

また SAS を含む多くのソフトは、非線形関数で大域最適解を求める NLP の技術に対応していない。NLP で開発した DEA[5]は多入力多出力の統計手法の代わりができる。日本の電力、空港、鉄道、金融機関、東京都の区立と市立図書館の評価を行った解説書を Amazon の Kindle で廉価で販売していた。販売が頭打ちになったので Kindle との契約を打ち切った。LINDO Japan から無料で入手できるので、各種競合組織や製品の評価等に統計のデータ解析手法の一種として利用することを勧める。

参考文献

[1]Shinmura(2016) New Theory of Discriminant Analysis After R. Fisher. Springer.

[2] Shinmura (2019) High-dimensional Microarray Data Analysis.

[3] Shinmura(2024)The First Discriminant Theory of LSD.-From Exams and Medical Diagnoses -

[4]Sall & Goodnight 著（新村秀一訳）(1986)「回帰分析の実践」朝倉書店

[5]Linus 著(新村訳)：LINGO の解説上・中巻等の多くの資料を Kindle から無償提供に変えた

[6] 2025 年度の日米癌シンポジュームの Abstract。Research Gate に UP しています。

[7]新村、飯田、丸山(1987)。SSM の癌治療における帰無仮説モデルによる評価、医療情報学,7/3. 263-276.

[8]新村（2011）合否判定データによる判別分析の問題点。学力試験データの統計解析特集号。応用統計。

[9] 新村(2017). DEA による問題の発見と解決, Kindle 版を無償公開。131 頁

注：Research Gate あるいは https://lindo.jp/から入手可能です。あるいはメールで要求してください

RTFファイルのページレイアウトを制御するマクロと コンペアマクロ

○平野　勝也

（株式会社インテリム　データサイエンスユニット）

Two SAS macros for setting page layout by TITLE, FOOTNOTE and ODS (RTF) TEXT statement and comparing with two RTF files generated by ODS RTF

Katsunari Hirano

Data Science Department, intellim Corporation

要旨

2024 年の SAS ユーザー総会において、宮辻 他(2024)「ODS RTF での出力環境の整備 : RTF スタイルと SAS マクロの運用・管理手順」[1]で、ODS RTF での出力環境と SAS マクロの管理手順を整備したことを報告した。報告後に、ページレイアウトを制御するための「タイトル・脚注の出力マクロ」の作成・ベリフィケーションを完了させ、品質管理手順を作成した。そして、2024 年末から 2025 年にかけて、整備した出力環境・SAS マクロ・品質管理手順で RTF 解析業務を実施し、大きな問題なく、解析業務を完了できた。

本報告では、解析業務を通して作成した 2 つの SAS マクロ、ページレイアウトを制御するための「タイトル・脚注の出力マクロ」、最終成果物の RTF ファイルを直接比較する「RTF ファイルのコンペアマクロ」の機能と仕組み、を紹介する。また、複数の RTF ファイルを 1 つの RTF ファイルに結合する際に発生した問題に対する 1 つのシンプルな回避策を紹介する。

キーワード： ODS RTF、ODS TEXT、SAS マクロ、命令語、タグ、制御文字列、RTF ファイル、コンペア、ページレイアウト、品質管理、SAS Macro、RTF File、Control Word、Compare、Page Layout、Quality Control

1. はじめに

日本では、2001 年にリリースされた Version 8.2 から ODS 機能を利用できるようになった。リリースから 24 年経過した今、ODS RTF を利用した SAS マクロを紹介するのは少々時代遅れと思うが、RTF 形式の解析資料の品質向上を期待して、二つの SAS マクロの機能と仕組みを紹介したい。また、解析業務の終盤に、複数の RTF ファイルを 1 つの RTF ファイルに結合した際、オリジナルの RTF ファイルのページ番号が保持されないという問題に遭遇するとともに、シンプルな回避策を見つけたため、この問題と回避策を簡単に紹介する。

2. ページレイアウトを制御する仕組み

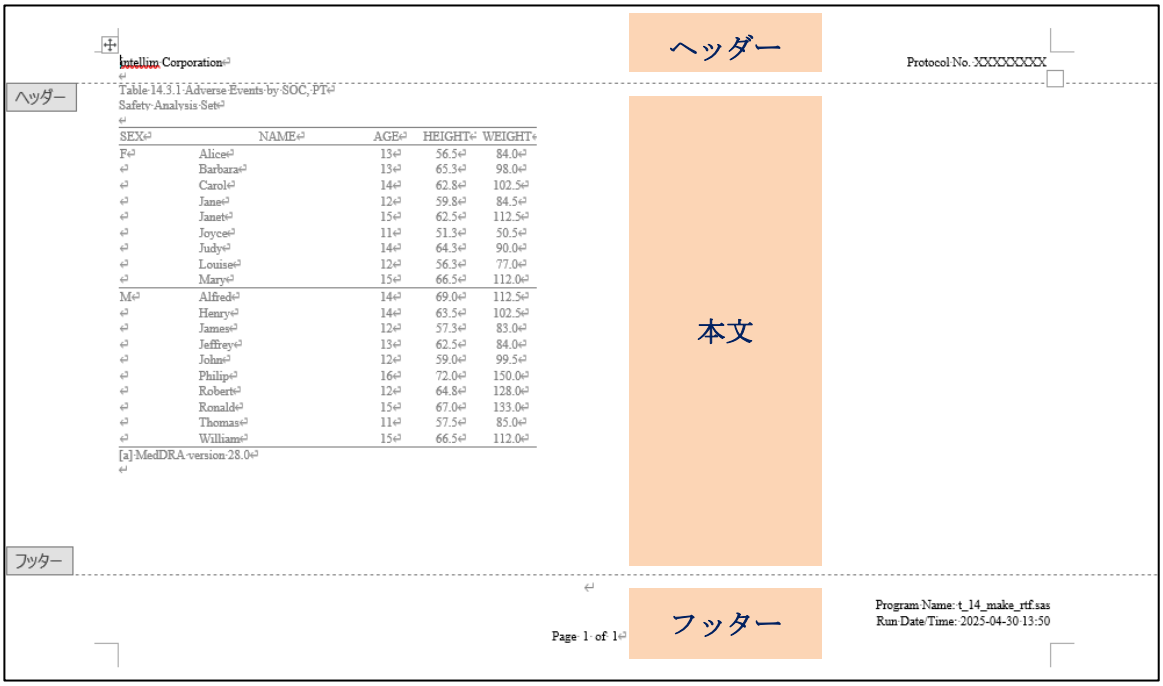
はじめに、宮辻 他(2024) [1]において Style テンプレートでページレイアウトの上下左右の余白を設定できる、と報告した。報告内容に間違いはいないが、上下左右のページ余白をプロトコル毎に変更したい場合や縦表示・横表示によりページ余白を変更したい場合には、Style テンプレートで余白設定せずに、（すべてのプログラムが、最初に読み込んで、ファイルパスや LIBNAME、オプションなどの共通設定を行うためのプログラム、例えば autoexec.sas に）OPTIONS ステートメント（TOPMARGIN、BOTTOMMARGIN、LEFTMARGIN、RIGHTMARGIN オプション）で設定したほうがよいかもしれない。Style テンプレートで余白を設定すると、OPTIONS ステートメントで指定しても、Style テンプレートの設定が優先されるためである。

本節では、結果として、RTF ファイルのページレイアウトを制御する仕組みとして、すべてのプログラムが読み込むプログラム（例えば autoexec.sas）と、SAS マクロ（%r_odstext_tf：弊社でのマクロ名）を利用する方法を紹介している。

2.1.RTF ファイルのページレイアウトとページレイアウトを制御する SAS ステートメント

RTF ファイルを MS-Word で開くと、ヘッダーとフッターと本文が存在することが分かる（図 1）。図 1 の作成プログラムでは、ODS RTF ステートメントに BODYTITLE オプションを指定していない。

図 1. RTF ファイルを MS-Word で開いた画面



ヘッダー/フッターに表示する文字列（例えば、intellim corporation、Program Name など）は、TITLE/FOOTNOTE ステートメントを利用して、JUSTIFY=オプションに left/center/right を指定することで、左/中央/右に出力する。図 1 のように、ヘッダーの左に「intellim Corporation」、右に「Protocol No. XXXXXXXX」を表示する場合には、TITLE ステートメントに JUSTIFY=left　‘intellim Corporation’を指定した後に、JUSTIFY=right　‘Protocol No. XXXXXXXX’を指定すればよい。

本文に表示する（表以外の）文字列は、ODS RTF PREPAGE=/TEXT=ステートメントもしくは ODS TEXT=ステートメントを利用して出力する（麻生（2024）[2]）。これらステートメントには JUSTIFY=オプションがないため、中央/右に文字列を出力したい場合には、文字列と共にインラインフォーマット（例えば、(*ESC*)S= {just= center}）を出力する必要がある。

2.2.タイトル・脚注の出力マクロ

タイトル・脚注を出力するための SAS マクロ（%r_odstext_tf）を作成するにあたり、①ヘッダーに出力する文字列と出力位置は TITLE ステートメントで指定する（SAS マクロの対象外）、②表タイトル/解析対象集団の出力位置（左/中央/右）、プロトコル毎にフッター部に出力する文字列と出力位置（左/中央/右）を SAS マクロの引数で指定する、③本文に出力する表タイトル/解析対象集団/脚注を別ファイルで管理する、④それぞれの帳票作成プログラムに記載するプログラム量を最小にする、という方針を定め、以下のように実装した。SAS マクロ（%r_odstext_tf）の仕組みを以下で紹介するので、各社・組織にて同様のマクロを作成する際に参考にしていただきたい。

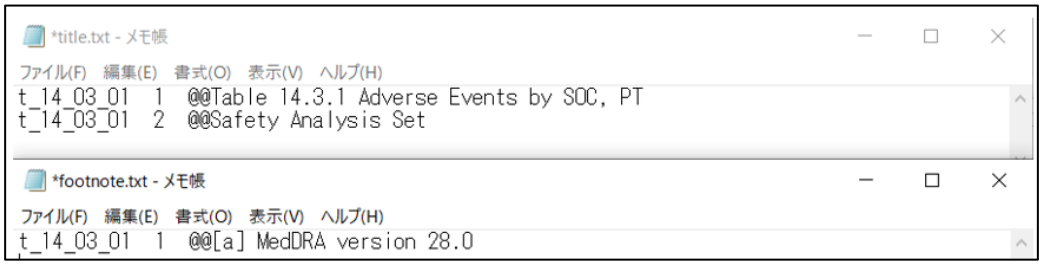
- ✓ ①②を実現するために、autoexec.sas ファイル内に TITLE ステートメントを記載するとともに、マクロ変数を定義して SAS マクロ（%r_odstext_tf）に引き渡すことにした。

Code 2. autoexec.sas に記載した SAS プログラム

```
TITLE1 JUSTIFY=left 'intellim Corporation' JUSTIFY=right "Protocol No. XXXXXXXX";
%let escapechar=~;
data _NULL_;
  length DATE TIME $10;
  DATE = put( today() , yymmdd10.);
  TIME = put( time() ,time5.);
  call symputx( "current_d" , DATE);
  call symputx( "current_dt" , catx(" " , DATE , TIME) );
run;
%let tfdata      =C:¥Title_Footer¥data;
%let title_pos   =L;
%let program_pos =R;
%let p_style     =%quote(Program Name);
%let datetime_pos=R;
%let d_style     =%quote(Run Date/Time: &current_dt.);
%let number_pos  =C;
%let n_style     =%quote(Page &escapechar.{thispage} of &escapechar.{lastpage});
```

- ✓ ③を実現するために、2 つのテキストファイル（title.txt、footnote.txt）を作成することにした（ここでは記載しないが、表タイトルと脚注の情報を Excel ファイルに入力することで、テキストファイルが自動作成される弊社自作の Excel VBA マクロを利用している）。各行の先頭に、解析図表の番号から生成した TLF_NO（t_14_03_1）が入力されており、SAS マクロ（%r_odstext_tf）の定位置マクロパラメータに TLF_NO を指定する仕様になっている。

図 3. 2 つのテキストファイル（title.txt、footnote.txt）の内容



✓ ④を実現するために、SAS マクロ（%r_odstext_tf）の定位置パラメータを 2 つ（tlfname と classification）にし、残りをキーワードパラメータとした（図 4）。その結果、作成した SAS マクロ（%r_odstext_tf）は、Code 5 のようにシンプルなプログラムで呼び出すことが可能となった（合わせて、TLF_NO の指定間違いや脚注の出力漏れを点検するために、SAS プログラムをチェックするための別の SAS プログラムを用意して目視点検タスクを低減している）。

Code 5 の(A)で TITLE/ FOOTNOTE ステートメントを実行するとともに、本文に表タイトルと解析対象集団を出力している。PROC REPORT/ ODSTABLE 実行時に TITLE/FOOTNOTE ステートメントの内容が出力され、(B)で本文に脚注を出力している。PROC REPORT /ODSTABLE を用いた帳票作成に関しては、太田（2017）[3]、山野辺（2018）[4]に詳述されている。

図 4. 作成した SAS マクロ （%r_odstext_tf） のマクロパラメータの設定

```
%macro r_odstext_tf(tlfname, classification,
        folder=&tfdata.,
        title_position=&title_pos.,
        program_position=&program_pos.,
        program_style=&p_style.,
        datetime_position=&datetime_pos.,
        datetime_style=&d_style.,
        number_position=&number_pos.,
        number_style=&n_style.
);
```

Code 5. 作成した SAS マクロ （%r_odstext_tf） の各帳票作成プログラム内での呼び出し例

```
%let TLF_NO=t_14_03_01;
%r_odstext_tf(&TLF_NO., TITLE);                    (A)
PROC REPORT/ ODSTABLE での帳票作成プログラム
%r_odstext_tf(&TLF_NO., FOOTER);                   (B)
```

作成した SAS マクロ（%r_odstext_tf）は、2 つのパート（マクロパラメータ classification が TITLE のパートと FOOTER のパート）からなる。TITLE パートでは、本文に表タイトルと解析対象集団を出力するための ODS RTF PREPAGE ステートメントと、フッターの文字列を出力するための FOOTNOTE ステートメントを記載している（FOOTNOTE ステートメントは、次の DATA ステップまたは PROC ステップの RUN ステートメントの実行時に有効になるため、FOOTER パートではなく、TITLE パートに記載している）。FOOTER パートでは、本文に脚注を出力するための ODS RTF TEXT ステートメントを記載している。

作成後に気付いたのだが、ODS TEXT ステートメントは、ODS RTF だけでなく、ODS EXCEL/ ODS PDF でも利用できるため、微妙な差異を確認した上で、ODS TEXT ステートメントでマクロ化したほうが良かった。ODS EXCEL を利用した帳票作成に関しては、松沢（2019）[5]に詳述されている。

3. RTF ファイルのコンペアマクロ

RTF ファイルは、Microsoft 社により開発された Rich Text Format と呼ばれるファイル形式で、全角・半角英数字と記号を含む文字情報に加えて、文字の装飾、画像、簡単なレイアウトを保存できる。RTF ファイルの検証手法は、山野辺（2024）[6]でも詳細に検討されている。

本節で取り上げるコンペアマクロは、ODS RTF ステートメントを利用して作成された RTF ファイルに限定しており、MS-Word/ワードパッドで作成もしくは更新された RTF ファイルを対象外としている。また、

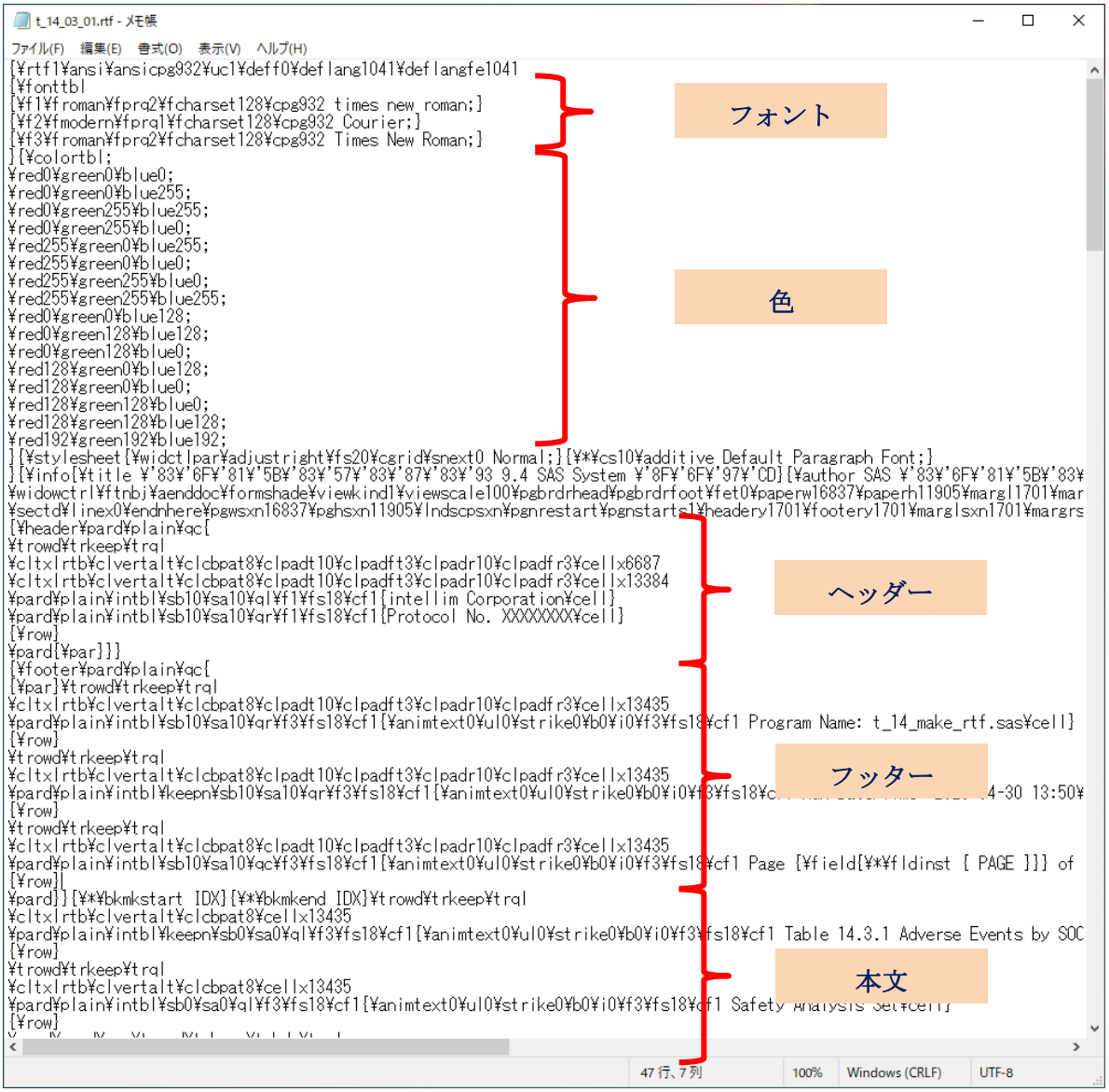
RTF ファイルに含まれるグラフデータと黒色以外の文字列もコンペア対象外としている（黒色であれば、日本語/英語、フォント種を問わずコンペア対象になる）。

3.1.RTF ファイルの中身

RTF ファイルは、Rich Text Format と呼ばれるファイル形式で、全角・半角英数字と記号を含む文字情報に加え、文字の装飾、画像、簡単なレイアウトを保存することができる。実際、図 1 に示した RTF ファイルをメモ帳で開くと、図 6 のように半角英数字と記号で記述されていることが分かる。

図 6 をじっくり観察すると、上から順にフォント定義、色定義、ヘッダー、フッター、本文の情報が、パート毎に纏まって記述されていることが分かる。プログラム化にあたり、各パートの区切り（始まりと終わり）と考えた文字列を表 7 に示す。本文のみ終わりの文字列を指定しているのは、改ページされた場合に本文パートの次に、ヘッダーパート、フッターパート、本文パートが続くためである。SAS マクロでは、これら文字列の有無を確認しながら、RETAIN ステートメントを利用して、3 変数（header/footer/body：それぞれのパートに存在する場合に 1、存在しない場合に 0）を作成している。

図 6. RTF ファイルをメモ帳で開いた画面（その 1）



ここで、色定義の 1 番目「¥red0¥green0¥blue0」が黒色であり、6 番目「¥red255¥green0¥blue0」が赤色であることをメモしておく。

表 7. 各パートの区切り文字（始まりと終わり）と考えた文字列

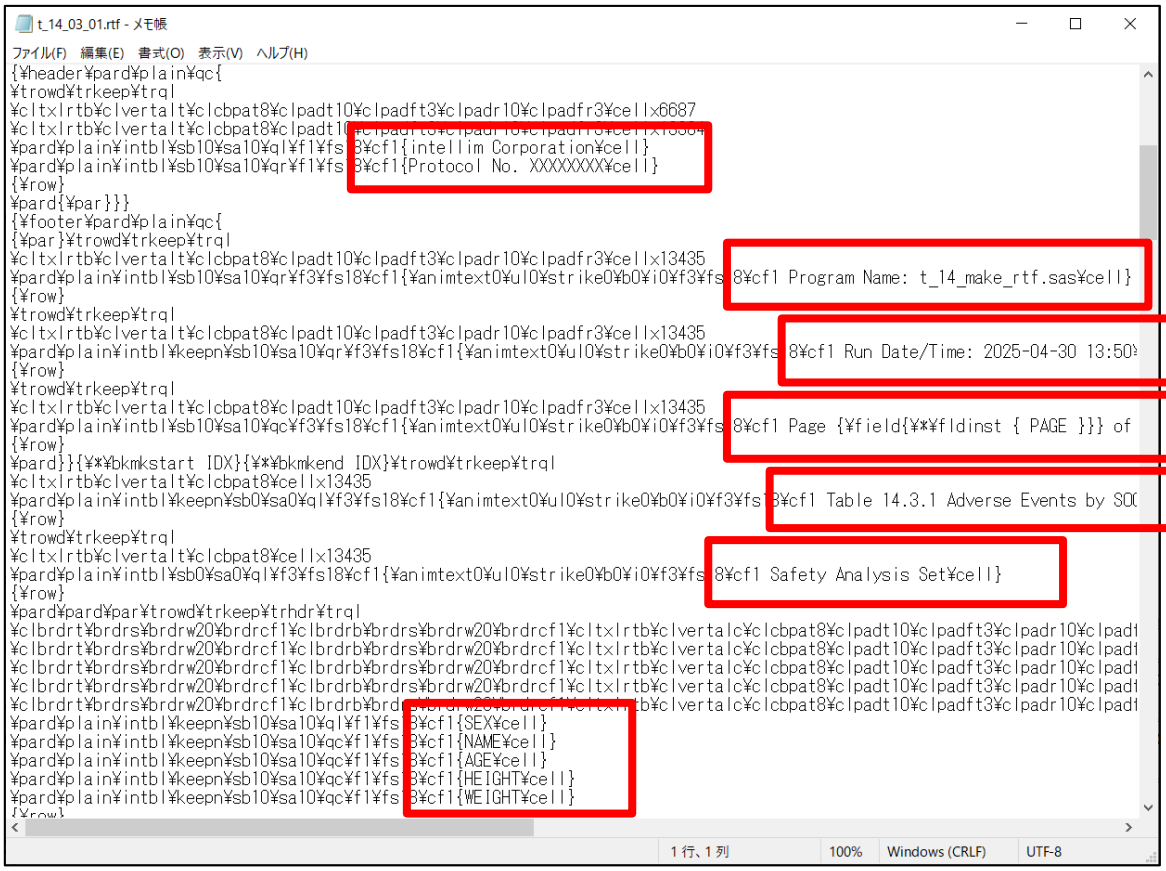
パート	始まりの文字列	終わりの文字列
フォント定義	¥fonttbl	
色定義	¥colortbl	
ヘッダー	{¥header	
フッター	{¥footer	
本文	¥bkmkend	¥pard}

3.2.RTF ファイルからの MS-Word に表示された黒色の文字列の抽出

図 6 の画面を少し下にスクロールしてヘッダーから本文の先頭までを表示した画面を図 8 に示す。赤枠部分を確認すると、「¥cf1{」または「¥cf1 」から「¥cell}」の間に、MS-Word に表示された文字列が存在している、と分かる。ここで、Rich Text Format (RTF) Specification [7]を確認すると、「¥cf」は Foreground Color、「1」は色定義の 1 番目の意味であることが分かったため、「¥cf1{」「¥cf1 」と「¥cell}」の間の文字列が黒色の文字列と解釈した。

INDEX 関数と SUBSTR 関数を利用した、MS-Word に表示された黒色の文字列を抽出する SAS プログラム例を Code 9 に示す。文字列の抽出を行う際に他の SAS 関数を利用すれば、よりスマートなプログラムにできるかもしれない。

図 8. RTF ファイルをメモ帳で開いた画面（その 2）



Code 9. RTF ファイルから MS-Word に表示された黒色の文字列を抽出する SAS プログラム例

```
filename rtf "RTF ファイル";
data RTF_DATA;
  length STRING $2000 text text1 $200;
  infile rtf dlm= '09'x lrecl = 2000;
  input STRING;

  if( index(STRING,'¥cf1 ') >0 ) then TEXT= substr(STRING, index(STRING,'¥cf1 ')+5);
  else if( index(STRING,'¥cf1{') >0 ) then TEXT= substr(STRING, index(STRING,'¥cf1{')+5);
  else if( STRING= '{¥line}' ) then TEXT= STRING;

  if( index(TEXT, '¥cell') >0 ) then do;
    if( index(TEXT, '¥cell')= 1) then TEXT1= '';
    else TEXT1= substr(TEXT,1,index(TEXT, '¥cell') -1);
  end;
run;
```

3.3.RTF ファイルから黒色の文字列を抽出する際の注意点

プログラム化する際の注意点をいくつか記載する。

- ✓ 図 1 に示したようなシンプルな帳票は少なく、セル結合されている帳票やセル内改行されている帳票が多い。そのような場合には、追加で「{¥line}」を利用することで MS-Word に表示された黒色の文字列を抽出できる。「{¥row}」を利用すれば、列数や行数もカウントできる。
- ✓ PROC REPORT で改ページされた RTF ファイルでも、文字列は抽出できる。ただし、RTF ファイル内のセクション数が改ページの数＋1 になり、ヘッダー・フッター・本文が改ページの数＋1 回 繰り返して格納されている。
- ✓ 第三者が MS-Word で RTF ファイルを開いていると、プログラム実行時に読み込みエラーが発生する。PIPE エンジンを利用した以下の filename ステートメントでエラーを回避できる。

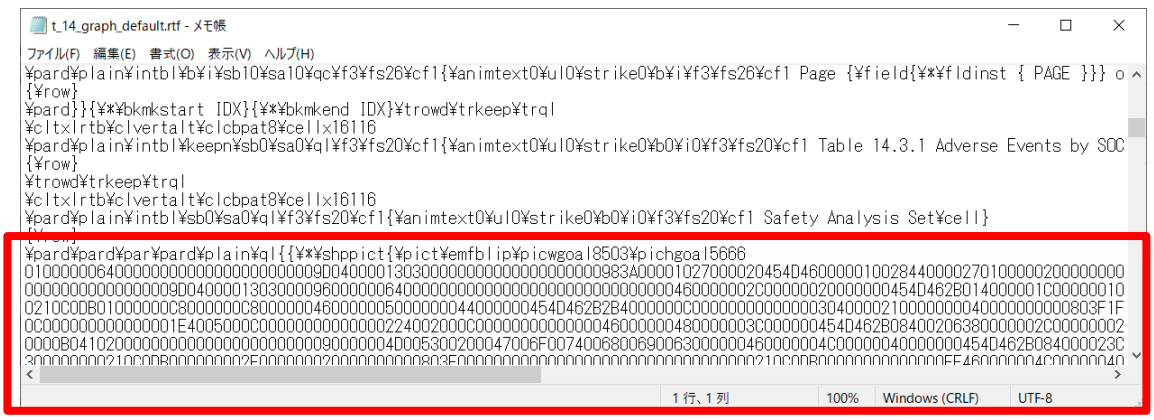
```
filename rtf pipe "type RTF ファイル";
```

- ✓ RTF ファイルには、2 バイト文字が UNICODE（例えば、"¥'83¥'6F¥'81¥'5B¥'83¥'57¥'83¥'87¥'83¥'93"、バージョン）で格納されている。読み取り可能な状態にするためには、以下の処理が必要である。

```
input( compress("¥' 83¥' 6F¥' 81¥' 5B¥' 83¥' 57¥' 83¥' 87¥' 83¥' 93", "¥' "), $hex200.);
```

- ✓ RTF ファイルにグラフが含まれる場合、グラフデータはテキスト化され、図 9 の赤枠のような形式で格納されている。グラフを含んでいても、グラフ以外の文字列を抽出できる。

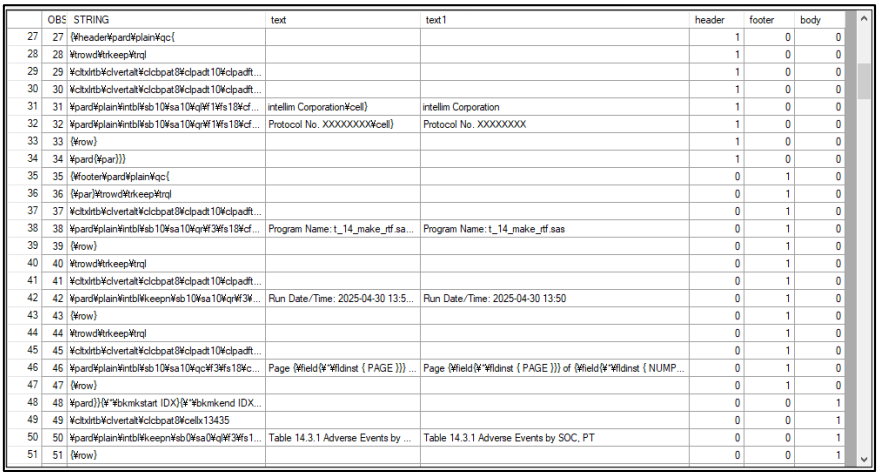
図 9. RTF ファイルをメモ帳で開いた画面（その 3）



3.4.抽出したデータのコンペア

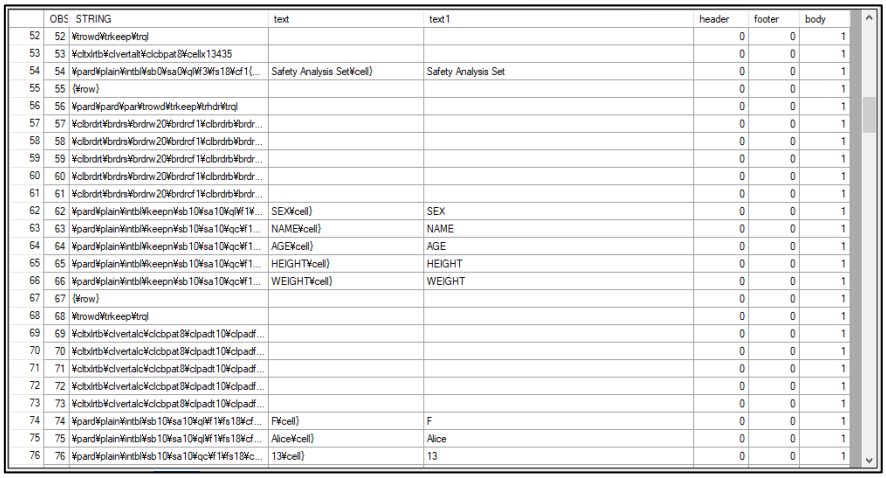
RTF ファイルから SAS マクロで作成した SAS データセットを図 10～12 に示す。ヘッダーに表示される文字列はプロトコル内で共通であること、フッターに表示されるプログラム名や実行日時は 2 系統間で異なることが想定されることから、本文中の差異を特定するために 3 変数（header/footer/body）を作成した。

図 10. SAS マクロで作成された SAS データセット（その 1）



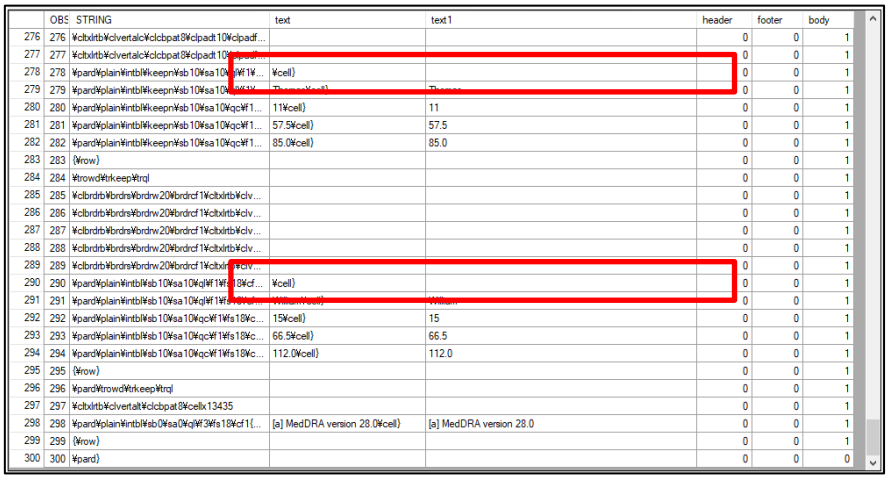
	OBS	STRING	text	text1	header	footer	body
27	27	{¥header¥pard¥plain¥qc{			1	0	0
28	28	¥trowd¥trkeep¥trql			1	0	0
29	29	¥cltxlrtb¥clvertalt¥clcbpat8¥clpadt10¥clpadft...			1	0	0
30	30	¥cltxlrtb¥clvertalt¥clcbpat8¥clpadt10¥clpadft...			1	0	0
31	31	¥pard¥plain¥intbl¥sb10¥sa10¥qj¥f1¥fs18¥cf...	intellim Corporation¥cell}	intellim Corporation	1	0	0
32	32	¥pard¥plain¥intbl¥sb10¥sa10¥qj¥f1¥fs18¥cf...	Protocol No. XXXXXXXX¥cell}	Protocol No. XXXXXXXX	1	0	0
33	33	{¥row}			1	0	0
34	34	¥pard{¥par}}}			1	0	0
35	35	{¥footer¥pard¥plain¥qc{			0	1	0
36	36	{¥par}¥trowd¥trkeep¥trql			0	1	0
37	37	¥cltxlrtb¥clvertalt¥clcbpat8¥clpadt10¥clpadft...			0	1	0
38	38	¥pard¥plain¥intbl¥sb10¥sa10¥qr¥f3¥fs18¥cf...	Program Name: t_14_make_rtf.sa...	Program Name: t_14_make_rtf.sas	0	1	0
39	39	{¥row}			0	1	0
40	40	¥trowd¥trkeep¥trql			0	1	0
41	41	¥cltxlrtb¥clvertalt¥clcbpat8¥clpadt10¥clpadft...			0	1	0
42	42	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qr¥f3¥...	Run Date/Time: 2025-04-30 13:5...	Run Date/Time: 2025-04-30 13:50	0	1	0
43	43	{¥row}			0	1	0
44	44	¥trowd¥trkeep¥trql			0	1	0
45	45	¥cltxlrtb¥clvertalt¥clcbpat8¥clpadt10¥clpadft...			0	1	0
46	46	¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f3¥fs18¥c...	Page {¥field{¥*¥fldinst { PAGE }}} ...	Page {¥field{¥*¥fldinst { PAGE }}} of {¥field{¥*¥fldinst { NUMP...	0	1	0
47	47	{¥row}			0	1	0
48	48	¥pard}}{¥*¥bkmkstart IDX}{¥*¥bkmkend IDX...			0	0	1
49	49	¥cltxlrtb¥clvertalt¥clcbpat8¥cellx13435			0	0	1
50	50	¥pard¥plain¥intbl¥keepn¥sb0¥sa0¥qj¥f3¥fs1...	Table 14.3.1 Adverse Events by ...	Table 14.3.1 Adverse Events by SOC, PT	0	0	1
51	51	{¥row}			0	0	1

図 11. SAS マクロで作成された SAS データセット（その 2）



	OBS	STRING	text	text1	header	footer	body
52	52	¥trowd¥trkeep¥trql			0	0	1
53	53	¥cltxlrtb¥clvertalt¥clcbpat8¥cellx13435			0	0	1
54	54	¥pard¥plain¥intbl¥sb0¥sa0¥qj¥f3¥fs18¥cf1{...	Safety Analysis Set¥cell}	Safety Analysis Set	0	0	1
55	55	{¥row}			0	0	1
56	56	¥pard¥pard¥par¥trowd¥trkeep¥trhdr¥trql			0	0	1
57	57	¥clbrdrt¥brdrs¥brdrw20¥brdrcf1¥clbrdrb¥brdr...			0	0	1
58	58	¥clbrdrt¥brdrs¥brdrw20¥brdrcf1¥clbrdrb¥brdr...			0	0	1
59	59	¥clbrdrt¥brdrs¥brdrw20¥brdrcf1¥clbrdrb¥brdr...			0	0	1
60	60	¥clbrdrt¥brdrs¥brdrw20¥brdrcf1¥clbrdrb¥brdr...			0	0	1
61	61	¥clbrdrt¥brdrs¥brdrw20¥brdrcf1¥clbrdrb¥brdr...			0	0	1
62	62	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qj¥f1¥...	SEX¥cell}	SEX	0	0	1
63	63	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	NAME¥cell}	NAME	0	0	1
64	64	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	AGE¥cell}	AGE	0	0	1
65	65	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	HEIGHT¥cell}	HEIGHT	0	0	1
66	66	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	WEIGHT¥cell}	WEIGHT	0	0	1
67	67	{¥row}			0	0	1
68	68	¥trowd¥trkeep¥trql			0	0	1
69	69	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
70	70	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
71	71	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
72	72	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
73	73	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
74	74	¥pard¥plain¥intbl¥sb10¥sa10¥qj¥f1¥fs18¥cf...	F¥cell}	F	0	0	1
75	75	¥pard¥plain¥intbl¥sb10¥sa10¥qj¥f1¥fs18¥cf...	Alice¥cell}	Alice	0	0	1
76	76	¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f1¥fs18¥c...	13¥cell}	13	0	0	1

図 12. SAS マクロで作成された SAS データセット（その 3）



	OBS	STRING	text	text1	header	footer	body
276	276	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
277	277	¥cltxlrtb¥clvertalc¥clcbpat8¥clpadt10¥clpadf...			0	0	1
278	278	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qj¥f1¥...	¥cell}		0	0	1
279	279	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qj¥f1¥...	Thomas¥cell}	Thomas	0	0	1
280	280	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	11¥cell}	11	0	0	1
281	281	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	57.5¥cell}	57.5	0	0	1
282	282	¥pard¥plain¥intbl¥keepn¥sb10¥sa10¥qc¥f1...	85.0¥cell}	85.0	0	0	1
283	283	{¥row}			0	0	1
284	284	¥trowd¥trkeep¥trql			0	0	1
285	285	¥clbrdrb¥brdrs¥brdrw20¥brdrcf1¥cltxlrtb¥clv...			0	0	1
286	286	¥clbrdrb¥brdrs¥brdrw20¥brdrcf1¥cltxlrtb¥clv...			0	0	1
287	287	¥clbrdrb¥brdrs¥brdrw20¥brdrcf1¥cltxlrtb¥clv...			0	0	1
288	288	¥clbrdrb¥brdrs¥brdrw20¥brdrcf1¥cltxlrtb¥clv...			0	0	1
289	289	¥clbrdrb¥brdrs¥brdrw20¥brdrcf1¥cltxlrtb¥clv...			0	0	1
290	290	¥pard¥plain¥intbl¥sb10¥sa10¥qj¥f1¥fs18¥cf...	¥cell}		0	0	1
291	291	¥pard¥plain¥intbl¥sb10¥sa10¥qj¥f1¥fs18¥cf...	William¥cell}	William	0	0	1
292	292	¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f1¥fs18¥c...	15¥cell}	15	0	0	1
293	293	¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f1¥fs18¥c...	66.5¥cell}	66.5	0	0	1
294	294	¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f1¥fs18¥c...	112.0¥cell}	112.0	0	0	1
295	295	{¥row}			0	0	1
296	296	¥pard¥trowd¥trkeep¥trql			0	0	1
297	297	¥cltxlrtb¥clvertalt¥clcbpat8¥cellx13435			0	0	1
298	298	¥pard¥plain¥intbl¥sb0¥sa0¥qj¥f3¥fs18¥cf1{...	[a] MedDRA version 28.0¥cell}	[a] MedDRA version 28.0	0	0	1
299	299	{¥row}			0	0	1
300	300	¥pard}			0	0	0

MS-Word に表示された黒色の文字列は変数 text1 に格納されることから、text1 ^='' のデータを対象にしてコンペアできる。ただし、空白セルのデータ（図 12 の赤枠のデータ）を含めてコンペアしたい場合には、

text1 ^= '' or text= '¥cell}'のデータを対象としてコンペアする必要がある。COMPARE プロシジャに関しては馴染み深いと思われるため、説明を省略する。

3.5.コンペア結果の出力

臨床試験で作成する解析帳票の数は優に 100 を超えるため、一度にすべての帳票をコンペアして、「差異なし」の帳票と「差異あり」の帳票を特定したいだろう。これを実現するために、各帳票（の本文中）の文字列をコンペアし、コンペア結果が「差異なし」の場合には AA_RESULT.csv に結果を追記し、「差異あり」の場合には AA_RESULT_NG.csv にコンペア結果を追記することにした。RTF ファイルが存在しない場合や帳票の列数が異なる場合でもプログラム実行を中断することなく、「差異あり」として AA_RESULT_NG.csv に追記する仕様にした。それぞれの出力例を図 13、14 に示す。

図 13. AA_RESULT.csv の出力例

	A	B	C	D	E	F
1	t_14_04_01_A	There is NO DIFFERENCE in the BODY.	2025/5/2	7:40	k-hirano	COMPARE RESULT
2	t_14_04_01_B	There is NO DIFFERENCE in the BODY.	2025/5/2	7:40	k-hirano	COMPARE RESULT
3	t_14_04_01_D	There is NO DIFFERENCE in the BODY.	2025/5/2	7:40	k-hirano	COMPARE RESULT
4	t_14_04_01_F	There is NO DIFFERENCE in the BODY.	2025/5/2	7:40	k-hirano	COMPARE RESULT

図 14. AA_RESULT_NG.csv の出力例

	A	B	C	D	E	F	G	H
1	t_14_04_01_C	2025/5/2	7:40	MAIN	SUB	COMPARE RESULT	RTF RESULT	
2	t_14_04_01_E	2025/5/2	7:40	MAIN	SUB	COMPARE RESULT	RTF RESULT	

図 13、14 の赤枠（COMPARE RESULT）をクリックすると、ヘッダー/フッター/本文のコンペア結果を確認できる（図 15、16）。ヘッダーに差異がないため、図 15、16 には Header=1 のレコードが存在しない。

図 15. 詳細なコンペア結果（本文中に差異がない場合）

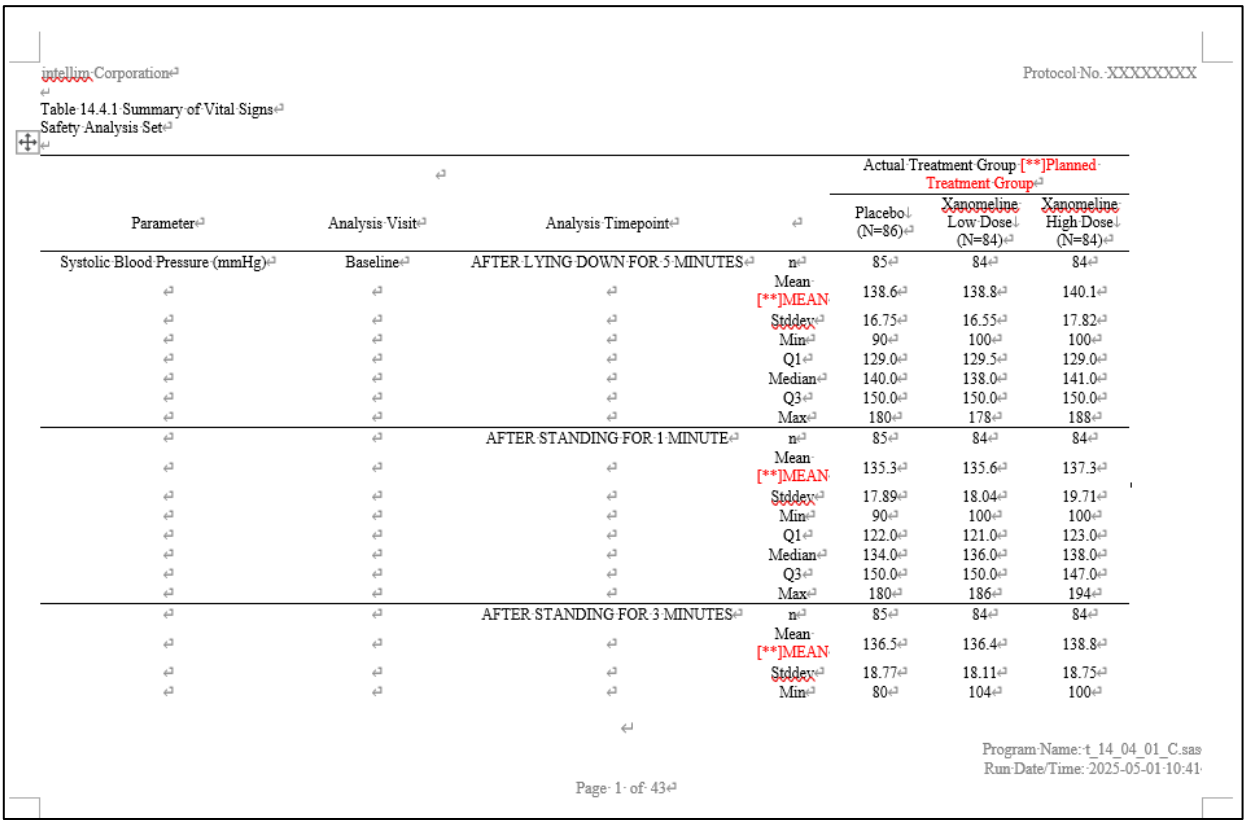
	A	B	C	D	E
1	MAIN: C:¥main¥t_14_04_01_A.rtf				
2	SUB : C:¥sub¥t_14_04_01_A.rtf				
3	MAIN(N_Column: 7, N_Row: 972)、SUB(N_Column: 7, N_Row: 972)				
4	**MAIN**	**SUB**	**Header**	**Footer**	**Body**
	|Program Name: t_14_04_01_a.sas	|Program Name: chk_t_14_04_01_a.sas	0	1	0
5					
6	|Run Date/Time: 2025-05-01 10:36	|Run Date/Time: 2025-05-01 11:53	0	1	0
	There is NO DIFFERENCE in the BODY.		0	0	1
7					
8					

図 16. 詳細なコンペア結果（本文中に差異がある場合）

	A	B	C	D	E
1	MAIN: C:¥main¥t_14_04_01_C.rtf				
2	SUB : C:¥sub¥t_14_04_01_C.rtf				
3	MAIN(N_Column: 7, N_Row: 974)、SUB(N_Column: 7, N_Row: 974)				
4	**MAIN**	**SUB**	**Header**	**Footer**	**Body**
	|Program Name: t_14_04_01_c.sas	|Program Name: chk_t_14_04_01_c.sas	0	1	0
5					
6	|Run Date/Time: 2025-05-01 10:41	|Run Date/Time: 2025-05-01 10:58	0	1	0
7	|Actual Treatment Group	|Planned Treatment Group	0	0	1
8	|Mean	|MEAN	0	0	1
9	|Mean	|MEAN	0	0	1
10	|Mean	|MEAN	0	0	1
11	|Mean	|MEAN	0	0	1

また、図 14 の青枠（RTF RESULT）をクリックすると、RTF ファイルで本文中の差異を赤字で確認できる（図 17）。

図 17. 本文中の差異を追記した RTF ファイル



intellim Corporation　　Protocol No. XXXXXXXX

Table 14.4.1 Summary of Vital Signs
Safety Analysis Set

Parameter	Analysis Visit	Analysis Timepoint		Actual Treatment Group [**]Planned Treatment Group		
				Placebo (N=86)	Xanomeline Low Dose (N=84)	Xanomeline High Dose (N=84)
Systolic Blood Pressure (mmHg)	Baseline	AFTER LYING DOWN FOR 5 MINUTES	n	85	84	84
			Mean [**]MEAN	138.6	138.8	140.1
			Stddev	16.75	16.55	17.82
			Min	90	100	100
			Q1	129.0	129.5	129.0
			Median	140.0	138.0	141.0
			Q3	150.0	150.0	150.0
			Max	180	178	188
		AFTER STANDING FOR 1 MINUTE	n	85	84	84
			Mean [**]MEAN	135.3	135.6	137.3
			Stddev	17.89	18.04	19.71
			Min	90	100	100
			Q1	122.0	121.0	123.0
			Median	134.0	136.0	138.0
			Q3	150.0	150.0	147.0
			Max	180	186	194
		AFTER STANDING FOR 3 MINUTES	n	85	84	84
			Mean [**]MEAN	136.5	136.4	138.8
			Stddev	18.77	18.11	18.75
			Min	80	104	100

Program Name: t_14_04_01_C.sas
Run Date/Time: 2025-05-01 10:41

Page 1 of 43

図 17 に示した RTF ファイルは、RTF ファイルから読み込んだテキスト（Code 9 の変数 STRING）に赤字のテキスト（命令語と差異）を追加して、外部ファイルに PUT ステートメントで出力して作成できる。色定義の **6** 番目「¥red255¥green0¥blue0」が赤色であることから、命令語を「**¥cf6** 」と指定している。

変更前；¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f1¥fs18¥cf1{Mean¥cell}

変更後：¥pard¥plain¥intbl¥sb10¥sa10¥qc¥f1¥fs18¥cf1{Mean{**¥cf6** [**]MEAN}¥cell}

3.6. RTF ファイルをコンペアする際の注意点

解析業務で RTF ファイルをコンペアする際の注意点をいくつか記載する。

- ✓ RTF ファイルから抽出した文字列をヘッダー/フッター/本文の順にコンペアしているため、改ページ位置や回数が異なる場合にコンペアで差異を検出してしまう。
  改ページ位置や回数の差異を無視するために、body 変数を利用（コンペア対象のデータセットの読み込みを where body=1 として、本文中の文字列に限定）してコンペアすることも可能だが、ヘッダー/フッターで想定外の差異を見逃す可能性があるため、注意が必要である。
- ✓ RTF ファイルのコンペアは、上から順に抽出した文字列をコンペアしているため、2 系統間でセル結合またはセル内改行のやり方が異なると、それ以降のすべてで差異を検出してしまう。
  プログラミングを開始する前に、図表計画書もしくは図表レイアウトなどの出力見本で、セル結合とセル内改行の位置やルールを確認しておくことで、プログラムの修正タスクを軽減できるだろう。
- ✓ 目視点検の代わりに RTF ファイルをコンペアするためには、1 系統の RTF ファイル作成とコンペアのタスクが増える。
  RTF ファイル作成＋コンペアのタスクと目視点検のタスク、ダブルプログラミング＋コンペアで得ら

れる品質と目視点検で得られる品質を比較して、RTF ファイルのコンペアの利用を判断いただきたい。最終実行後に目視点検する場合には、最終実行から納品までの時間も考慮する必要があるかもしれない。

✓ RTF ファイルをコンペアするなら、RTF ファイルを作成する前の SAS データセットのコンペアは不要ではないか、という意見がでるだろう。

確かに、最終成果物の RTF ファイルをコンペアして一致することが確認できれば品質的に問題ないと考えてもよいかもしれない。一方で、出力前 SAS データセットのコンペアと併用することで、目視点検の代わりに RTF ファイルをコンペアしている、と考えることもできる。目視点検を RTF ファイルのコンペアに置き換えた手順から始めて、最終成果物の品質に応じて RTF ファイルのコンペアのみの手順にシフトしていくのが現実的かもしれない。

4. RTF ファイルを結合する際に遭遇した問題と回避策

4.1.遭遇した問題

複数の RTF ファイルを結合して 1 つの RTF ファイルを作成するプログラムがいくつか公開されている。これらを利用して作成した RTF ファイルで、オリジナルの RTF ファイルのページ番号が維持されない、という問題に遭遇した。オリジナルの RTF ファイルと結合した RTF ファイルのページ番号は表 18 のようになった。この問題は、WenHsuan（2021）[8]で報告され、解決策が提示されているため、参考にされたい。また、Python を利用して RTF ファイルを PDF 化する方法が Jun（2025）[9]で紹介されている。

表 18. オリジナルの RTF ファイルでのページ番号と結合した RTF ファイルでのページ番号

RTF ファイル	オリジナルの RTF ファイル	結合した RTF ファイル
1 番目	Page 1 of 3 Page 2 of 3 Page 3 of 3	Page 1 of 6 Page 2 of 6 Page 3 of 6
2 番目	Page 1 of 2 Page 2 of 2	Page 1 of 6 Page 2 of 6
3 番目	Page 1 of 1	Page 1 of 6

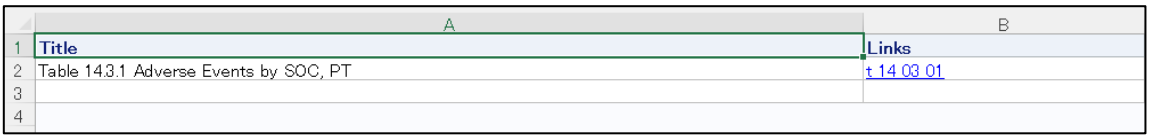
4.2.一つの回避策

CSR や CTD には RTF ファイルではなく PDF ファイルが添付されることから、「結合した RTF ファイル」の作成は必須ではない。このため「結合した PDF ファイル」を作成し、そして目次の代わりに「しおり」を自動作成する、という回避策を考えた。手順は簡単である。①作成した RTF ファイルの名称を図表タイトルに変更する（2 節で作成した title.txt を利用するプログラムを添付）。②変更したファイルを PDF 化する（いくつかのプログラムが公開されている）。③Adobe Acrobat Pro で順序を考慮しながら PDF ファイルを結合することで「しおり」が自動作成される。ヘッダーに全ページに共通のページ番号を追加して、最後に表紙を追加して PDF ファイルを完成できる。

CSR や CTD を作成する MW や薬事の担当者から、PDF ファイルとは別に、Copy&Paste できる RTF ファイルを参照しやすい形で提供してほしい、と依頼されるだろう。依頼される前に、2 節で作成した title.txt を

利用して同一フォルダ内の RTF ファイル（弊社では、RTF ファイルの名称は TLF_NO.rtf としている）にリンク付けた Excel ファイル（図 19）を ODS EXCEL で作成して提供することをお勧めする（プログラムを添付）。

図 19. 同一フォルダ内の RTF ファイルにリンク付けた Excel ファイル



	A	B
1	Title	Links
2	Table 14.3.1 Adverse Events by SOC, PT	t 14 03 01
3		
4		

5. 最後に

RTF 形式の解析資料の品質向上を期待して、ODS RTF のページレイアウトを制御するマクロとコンペアマクロを紹介した。SAS マクロを公開できないため、可能な限り丁寧に仕組みを説明したつもりである。説明が不足している点、間違っている点、プログラム化での悩みがあれば、遠慮なく、連絡いただきたい。

Excel DDE プログラミングから ODS RTF プログラミングに移行したように、近い将来、SAS から R もしくは Python に移行するかもしれないし、AI の進化で自動的にプログラム（もしくは解析資料）が生成されるようになるかもしれない。どう変わろうとも、高品質な解析資料を提供するためには、作成手順に応じた品質管理手順を構築した上で、成果物を確認するプロセスは必要であろう。そのためにも、新しい技術を継続的に習得しながら、効率的な品質管理手順を検討していきたい。

6. 参考文献

[1] 宮辻 淳、麻生 美由紀、平野 勝也. 2024. ODS RTF での出力環境の整備 ～RTF スタイルと SAS マクロの運用・管理手順～. SAS ユーザー総会
https://sas-user2024.ywstat.jp/download.html?n=35&key=paper

[2] 麻生 美由紀. 2024. RTF Output Tips -SAS 初学者も SAS 熟練者もたのしめるものを目指して-. SAS ユーザ総会
https://sas-user2024.ywstat.jp/download.html?n=30&key=paper

[3] 太田 裕二. 2017. ODS 機能と REPORT プロシジャを用いた 解析帳票の作成. SAS ユーザー総会 2017.
https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2017-b-11-02.pdf

[4] 山野辺 浩己. 2018. PROC ODSTABLE を用いた帳票作成. SAS ユーザー総会 2018.
https://www.docswell.com/s/6484025/ZN1P6L-2025-05-13-182932#p210

[5] 松沢 享. 2019. ODS EXCEL 入門. SAS ユーザー総会 2019.
https://www.sas.com/content/dam/SAS/documents/event-collateral/2019/ja/sas-users-group-meeting/Presentations/d-14-matsuzawa-presentation-rev2.pdf

[6] 山野辺 浩己. 2024. 解析結果 RTF ファイルの検証手法の検討と報告. SAS ユーザー総会 2024.
https://sas-user2024.ywstat.jp/download.html?n=70&key=dtkggzbjpctf

[7] Rich Text Format (RTF) Specification
https://www.biblioscape.com/rtf15_spec.htm

[8] WenHsuan Wu. 2021. An Application of Flexible Operation Page Number Positioning Method for RTF Outputs. PharmaSUG.
https://pharmasug.org/proceedings/2021/AD/PharmaSUG-2021-AD-112.pdf

[9] Jun Yang and Yan Moore. 2025. Packing PDF TFL with Table of Contents and Bookmarks Using Python. PharmaSUG.
https://pharmasug.org/proceedings/2025/OS/PharmaSUG-2025-OS-290.pdf

ROBUSTREGプロシジャによるロバスト回帰とその経時測定データ解析への応用

伊庭　克拓、○浅野　豊、鈴木　聡、毛利　誠、横田　大輔

（大塚製薬株式会社 新薬開発本部 バイオメトリックス部 統計解析室）

Robust Regression using ROBUSTREG Procedure and Its Application to Longitudinal Data Analysis

Katsuhiro Iba, Yutaka Asano, So Suzuki, Makoto Mouri, Daisuke Yokota
Office of Biostatistics, Department of Biometrics, Headquarters of Clinical Development, Otsuka Pharmaceutical Co., Ltd.

要旨

臨床試験では，しばしば外れ値を含んだデータが得られることがある。臨床試験での連続変数の解析では，応答変数が正規分布に従うことを仮定したanalysis of covariance（ANCOVA）やmixed model for repeated measures（MMRM）がよく用いられているが，これらの正規性を仮定した解析は外れ値の影響を受ける可能性がある。SASでは，ROBUSTREGプロシジャで外れ値に対して頑健なロバスト回帰を行うことができる。また，臨床試験では経時的に複数の時点で評価が行われ，中止等による欠測値が生じることがある。欠測値を含むデータの場合，MIプロシジャで欠測値を多重補完してからロバスト回帰を行うことが可能である。よく用いられる単調回帰法は正規性を仮定した回帰モデルから補完値を生成するのに対し，予測平均マッチング法および傾向スコア法は観測データのマッチングによって補完値を生成する。本論文では，ロバスト回帰による解析を行う方法および上記の多重補完法と組み合わせて経時測定データに適用する方法を説明すると共に，シミュレーション実験でこれらの解析方法の性能評価を行う。

キーワード：外れ値，ロバスト回帰，ROBUSTREG，多重補完法，予測平均マッチング法，傾向スコア法

1. はじめに

医薬品や医療機器の有効性および安全性を評価する臨床試験では，しばしば集団の全体的な値と比べて極端に大きい値あるいは極端に小さい値である外れ値を含んだデータが得られることがある（どのような場合に外れ値とするか厳密な定義はないが，本稿ではこのように考えることとする）。臨床試験での連続変数の解析では，応答変数に影響する共変量の調整などを目的として，応答変数が正規分布に従うことを仮定したanalysis of covariance（ANCOVA）やmixed model for repeated measures（MMRM）がよく用いられている。こ

れらの正規性を仮定した解析は，外れ値によって正規性の仮定からのずれが生じた場合，その程度により影響を受ける可能性がある。

外れ値に対して頑健な解析方法として，Wilcoxon 検定による群間比較や Hodges–Lehmann 推定量（岩崎, 2006）による群間差の推定がある。また，Wilcoxon 検定の拡張である van Elteren 検定（森川・田崎, 2009）では，層別因子を調整して群間比較を行うことができる。SAS では，NPAR1WAY プロシジャでこれらの解析を実行できるが（SAS Institute Inc., 2023），Wilcoxon 検定および Hodges–Lehmann 推定量は共変量を考慮できず，van Elteren 検定は層別因子を考慮できるものの連続変数の共変量は値の区分を設定して層別しなければならない。外れ値に対して頑健であり，かつ連続変数の共変量をそのまま考慮して群間比較や群間差の推定を行うことができる解析方法として，Huber の M 推定をはじめとしたロバスト回帰がある（蓑谷, 2016）。SAS では，ROBUSTREG プロシジャでロバスト回帰を行うことができる（SAS Institute Inc., 2023）。ROBUSTREG プロシジャはそれほど新しいプロシジャではないものの，過去の SAS ユーザー総会で ROBUSTREG プロシジャを取り上げた演題は見当たらない。本稿で紹介するロバスト回帰とは別の外れ値に対するアプローチとして，橋本（2008）では経時測定データに対して誤差の分布に Normal/independent distributions を仮定した線形混合モデルを適用する事例が取り上げられている。

臨床試験では経時的に複数の時点で評価が行われ，中止等による欠測値が生じることがある。ROBUSTREG プロシジャは欠測値を含むデータを扱うことができず，欠測値を無視して時点ごとにロバスト回帰を行うと推定精度が低下し，欠測メカニズムが missing completely at random（MCAR）ではない場合，バイアスが生じる。そのため，欠測値を含む経時測定データをロバスト回帰で解析する方法として，多重補完法で欠測値を補完してからロバスト回帰を行うことが考えられる。SAS では，MI プロシジャで多重補完法を行うことができる（SAS Institute Inc., 2023）。多重補完法での補完値の生成によく用いられる単調回帰法（阿部, 2016; 野間, 2017）は，ANCOVA や MMRM と同様に正規性を仮定した回帰モデルから補完値を生成するため，外れ値による正規性の仮定からのずれに影響を受ける可能性が考えられる。MI プロシジャでは，補完対象の変数が従う分布を仮定しない hot deck な（観測された値から補完値を生成する）方法として，予測平均マッチング法（野間, 2017）および傾向スコア法（阿部, 2016）も用いることができる。

本稿では，ROBUSTREG プロシジャでロバスト回帰による解析を行う方法およびロバスト回帰と多重補完法を組み合わせて経時測定データに適用する方法を説明すると共に，簡単なシミュレーション実験でこれらの解析方法の性能評価を行う。

2. シミュレーションデータ

本稿では，仮想の臨床試験のシミュレーションデータを生成し，解析方法の説明およびシミュレーション実験での評価に用いた。試験デザインは実薬群とプラセボ群の 2 群（各群 100 例）の無作為化比較試験で，ベースラインおよび投与後 4 時点の計 5 時点で連続変数の評価項目を測定し，規定された最終時点でのベースラインからの変化量の群間差に主な関心があるとする。

外れ値を含む経時データの生成方法について説明する。$\boldsymbol{y}_g = \left(y_{g1}, y_{g2}, y_{g3}, y_{g4}, y_{g5}\right)^T$を群$g$ (A: 実薬群, P: プラセボ群) のベースラインおよび投与後 4 時点の合計 5 時点の測定値とし，外れ値を含むデータを生成するために，以下の多変量混合正規分布からの疑似乱数を発生させた。

$$\boldsymbol{y}_g \sim 0.9MVN\left(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}\right) + 0.1MVN\left(\boldsymbol{\mu}_g, 9\boldsymbol{\Sigma}\right)$$

これは平均ベクトル$\boldsymbol{\mu}_g$，分散共分散行列$\boldsymbol{\Sigma}$の多変量正規分布と，平均ベクトルは同じで分散共分散行列 9$\boldsymbol{\Sigma}$のより裾が広い多変量正規分布を混合比 9：1 で混合した分布であり，後者の多変量正規分布が前者の多変量正規分布からは確率的にほぼ生じないような値（外れ値）を一定の確率で生じさせる。プラセボ群の平均ベクトル$\boldsymbol{\mu}_P$および実薬群の平均ベクトル$\boldsymbol{\mu}_A$は，

$$\boldsymbol{\mu}_P = (0.00, 0.00, 0.00, 0.00, 0.00)^T,$$
$$\boldsymbol{\mu}_A = (0.00, 0.20, 0.30, 0.40, 0.46)^T$$

と設定した。$\boldsymbol{\Sigma}$の共分散構造には 1 次の自己回帰（$\boldsymbol{\Sigma}$の(i,j)成分は$\sigma^2\rho^{|i-j|}$）を仮定し，$\sigma^2 = 1,\ \rho = 0.8$と設定した。最終時点でのベースラインからの変化量（$y_{g5} - y_{g1}$）の群間差の真値は 0.46 であり，正規分布の下で欠測値がない場合に両側有意水準 5%の検定に対する検出力が 90%となるように設定した。SAS でシミュレーションデータを生成するために，多変量正規分布からの乱数を発生させる SIMNORMAL プロシジャ（SAS Institute Inc., 2023）を用いて 2 つの多変量正規分布$MVN(\boldsymbol{\mu}_g, \boldsymbol{\Sigma})$および$MVN(\boldsymbol{\mu}_g, 9\boldsymbol{\Sigma})$からの乱数を含んだデータセットをそれぞれ生成し，データステップで結合した後，0~1 の一様乱数を発生させて，0.9 未満であれば$MVN(\boldsymbol{\mu}_g, \boldsymbol{\Sigma})$からの乱数を，それ以外の場合は$MVN(\boldsymbol{\mu}_g, 9\boldsymbol{\Sigma})$からの乱数を採用した。

次に上記で生成した完全データセットに対して欠測値を発生させることで，不完全データセットを生成した。ベースラインは常に観測されるとし，投与後 4 時点$(j = 2, \cdots, 5)$に対して，欠測メカニズムに missing at random（MAR）を仮定し，$y_{g(j-1)}$が観測された条件の下でy_{gj}が欠測値となる確率p_{gj}を

$$p_{gj} = \frac{\exp(-2.94 - 0.1y_{g(j-1)})}{1 + \exp(-2.94 - 0.1y_{g(j-1)})}$$

に基づいて算出する。モデル内の係数は最終時点の欠測値の割合が 20%前後となるように設定した。0~1 の一様乱数を発生させて，p_{gj}未満であればy_{gj}を欠測値とし，単調な欠測パターンとなるように，ある時点で欠測値となった場合，それ以降の時点はすべて欠測値とした。

生成したデータセットには，被験者番号 SUBJ，治療群 TRT（1=実薬群，2=プラセボ群），ベースラインおよび投与後 4 時点の合計 5 時点の測定値 Y1，Y2，Y3，Y4，Y5 の変数が含まれる。

3. ロバスト回帰

3.1 Huber の M 推定

個体$i(i = 1, \cdots, n)$の応答変数をy_i，切片項を含む説明変数を$\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip})^T$，$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$を回帰係数パラメータとしたとき，ロバスト回帰モデルは以下のように表される。

$$y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \qquad , e_i \sim L(\cdot/\sigma)$$

モデル式は通常の線形回帰モデルと同じであるが，誤差項e_iは通常の線形回帰モデルでは正規分布に従うと仮定するのに対して，ロバスト回帰ではスケールパラメータσをもつ分布Lに従うと仮定する。通常の線形回帰では，応答変数y_iとその予測値$\hat{y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$の差である残差に対する損失関数として残差平方和$\Sigma_{i=1}^n(y_i - \hat{y}_i)^2$を最小化する最小二乗法（最尤法と等価）を用いて回帰係数パラメータを推定するの対し，ロバスト回帰の代表的な方法である Huber の M 推定では，残差に対する急激に増加しない損失関数

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \hat{y}_i}{\sigma}\right)^2$$

を最小化する（$\rho(z)$が二乗関数，すなわち$\rho(z) = z^2$の場合は最小二乗法になる）。σが既知の場合，回帰係数パラメータの推定値は，$Q(\boldsymbol{\beta})$を$\boldsymbol{\beta}$の各要素で偏微分した以下の方程式の解である。

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \hat{y}_i}{\sigma}\right) x_{ij} = 0 \qquad ,j = 1, \cdots, p$$

ここで，$\psi(z) = \partial\rho/\partial z$である。重み関数を$w(z) = \psi(z)/z$と定義することによって，以下を最小化する反復重み付き最小二乗法（iteratively reweighted least squares: IRLS）で回帰係数パラメータを推定することができる。

$$\sum_{i=1}^{n} w_i \left(\frac{y_i - \hat{y}_i}{\sigma}\right)^2$$

ここで，$w_i = w\left(\frac{y_i - \hat{y}_i}{\sigma}\right)$は標準化残差の大きさに応じた重みであり，これにより外れ値の影響を弱めている。最小二乗法と異なり，上記の IRLS ではスケールパラメータの推定値が必要となる。通常の線形回帰では，スケールパラメータは残差平方和から推定するが，ロバスト回帰ではスケールパラメータの推定も外れ値に対して頑健な方法を用いる必要がある。ROBUSTREG プロシジャではいくつかのスケールパラメータの推定法が実装されているが，デフォルトでは残差の絶対値の中央値に基づく

$$\hat{\sigma} = \text{median}\left\{\frac{|y_i - \hat{y}_i|}{\Phi^{-1}(0.75)}\right\}$$

が用いられる（SAS Institute Inc., 2023）。Φは標準正規分布の累積分布関数であり，誤差項が正規分布に従う場合に一致推定量となるための定数である。最小二乗法に基づく推定値から重みの初期値を算出し，IRLS で回帰係数パラメータの推定値を求め，得られた回帰係数パラメータの推定値に基づいてスケールパラメータの推定値を求めて重みを更新する。それぞれの推定値が収束するまで，推定のプロセスを繰り返す（SAS Institute Inc., 2023）。Huber の M 推定によって得られる回帰係数パラメータの推定量は漸近正規性を持ち（蓑谷, 2016），通常の線形回帰と同様に，仮説検定を行うことや信頼区間を求めることができる。また，ROBUSTREG プロシジャでは，ロバスト回帰に対応した決定係数R^2や赤池情報量規準（Akaike's information criterion: AIC）の算出，外れ値や推定への影響が強いデータを検出する回帰診断も行うことができる。

ROBUSTREG プロシジャではいくつかの重み関数が実装されているが，デフォルトで設定されている Tukey の bisquare 重み関数

$$w(z, c) = \begin{cases} (1 - (z/c)^2)^2 & \text{if } |z| < c \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

がよく用いられており（蓑谷, 2016），本稿でも bisquare 重み関数を用いる。標準化残差zが 0 のときに最大の重み$w(0, c) = 1$となり，標準化残差zの絶対値が大きくなるほど重みが減少するような関数である。標準化残差の絶対値が調整定数c以上のときに最小の重み$w(z, c) = 0$となり，回帰係数パラメータの推定に影響しなくなる。調整定数のデフォルトは$c = 4.685$であり，誤差項が正規分布に従っている場合の漸近的有効性（回帰係数の推定量の分散の比（最小二乗法／M 推定））が 95%となるように設定されている（つまり，誤差項が正規分布に従っている場合でも最小二乗法と比べて推定効率がほとんど低下しない）。調整定数を小さくすると，より外れ値に対して敏感になる。図 1 にデフォルトの$c = 4.685$および$c = 1.548$のときの bisquare 重み関数を示す。$c = 1.548$は，頑健性を維持できる外れ値の割合として定義される breakdown value（あるいは breakdown point とも呼ばれる）が 50%となる調整定数である（breakdown value の詳細な定義は蓑谷（2016）を参照）。ROBUSTREG プロシジャでは 10 種類の重み関数が実装されているが，M 推定の性能は重み関数の形状よりも調整定数の影響を受けやすい（SAS Institute Inc., 2023）。

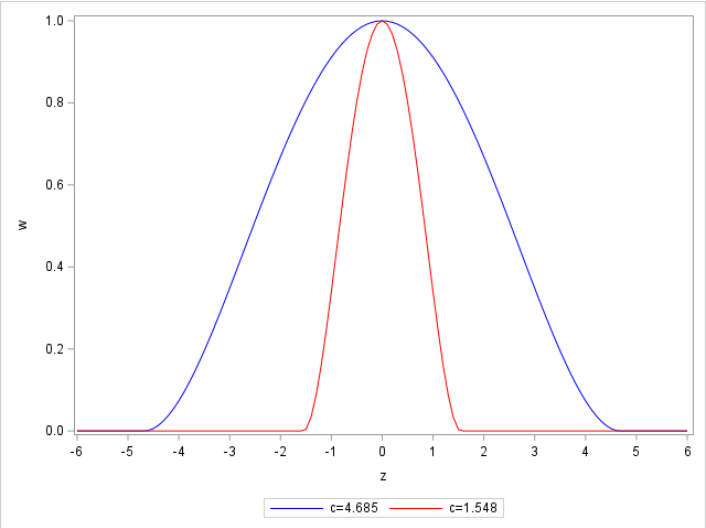


図 1　各調整定数 c に対する bisquare 重み関数

3.2 ROBUSTREG プロシジャによるロバスト回帰の実行

2 節で生成した外れ値を含むシミュレーションデータを対象に，ROBUSTREG プロシジャでロバスト回帰を行う方法を説明する。また，最小二乗法に基づく ANCOVA の結果との比較を行う。本節では欠測値を発生させる前の完全データを解析する。最終時点でのベースラインからの変化量 CHG（Y5－Y1）を応答変数，治療群 TRT（1=実薬群，2=プラセボ群）およびベースライン値 Y1 を説明変数とした ANCOVA 型のロバスト回帰を行う SAS コードを以下に示す。

```
proc robustreg data=EXAMPLE method=m;
  class TRT;
  model CHG = TRT Y1;
  ods output ParameterEstimates=OUT_REG;
run;
```

ROBUSTREG プロシジャの構文は，MIXED プロシジャや GLM プロシジャなどの回帰モデル系のプロシジャの構文と類似しており，これらのプロシジャを使用したことがあるユーザーにとっては馴染みやすいと思われる。PROC ROBUSTREG 文の METHOD オプションで推定方法を指定することができ，更にサブオプションで重み関数や調整定数といった推定に関する詳細な設定が可能である（デフォルトは M 推定であり，M 推定を行う場合は省略可能）。CLASS 文では後続の MODEL 文の右辺に含めるカテゴリー変数を指定する。なお，MIXED プロシジャや GLM プロシジャでは CLASS 文に REF オプションを指定して，基準となる水準を指定することができるが，ROBUSTREG プロシジャでは本オプションがサポートされていない。MODEL 文では応答変数とモデルに含める説明変数の変数を指定する。ODS OUTPUT 文で各解析結果をデータセット化することができる。上記 SAS コードでは ODS テーブル名 ParameterEstimates を指定し，回帰係数パラメータ推定値，標準誤差（SE），信頼区間などをデータセットに出力している。

表 1 にロバスト回帰および ANCOVA（応答変数および説明変数はロバスト回帰と同じ）の治療群の回帰係数パラメータ（最終時点の群間差）の結果を示す。点推定値は ANCOVA と比べてロバスト回帰の方が真

値の 0.46 に近い値であり，SE もロバスト回帰の方が小さかった。図 1 にベースライン値とベースラインからの変化量の散布図とロバスト回帰および ANCOVA に基づく回帰直線を示す。ロバスト回帰の回帰直線と比べて，ANCOVA の回帰直線は外れ値（実薬群は図右上，プラセボ群は図左下のデータ）の影響を受けていることが示唆される。なお，ロバスト回帰の特徴を説明するため，5 節のシミュレーション実験で生成したデータセットのうち，ロバスト回帰と ANCOVA の結果が極端に異なるケースを解析に用いた。

表 1　ロバスト回帰と ANCOVA の結果（最終時点の群間差）

			両側 95%信頼区間		
	推定値	SE	下限	上限	p 値
ロバスト回帰	0.58	0.17	0.26	0.91	0.0005
ANCOVA	1.11	0.23	0.65	1.57	<.0001

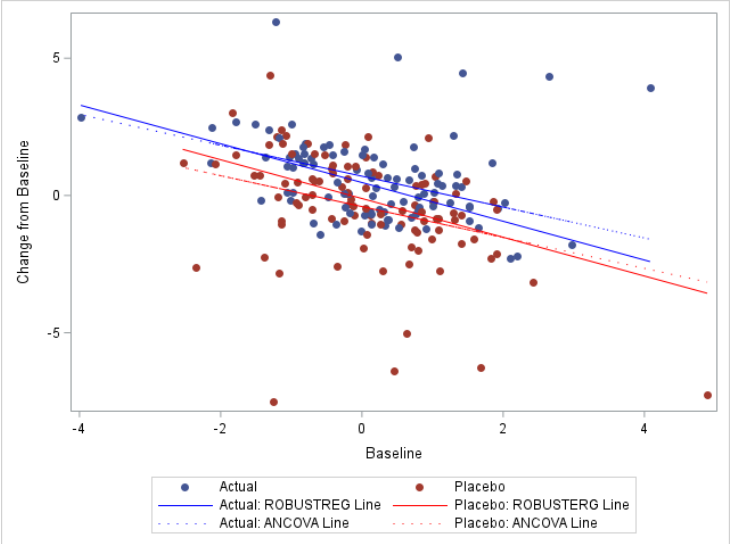


図 2　ロバスト回帰および ANCOVA に基づく回帰直線

4. 経時測定データ解析への応用

欠測値を含む経時測定データをロバスト回帰で解析するために，多重補完法で欠測値を補完したデータセットに対してロバスト回帰を行う方法を説明する。多重補完法は，欠測値を複数回補完した疑似完全データをそれぞれ解析し，得られたパラメータ推定値を Rubin のルールで統合して最終的な推定値を得ることによって，欠測値の補完に対する不確実性を考慮する方法である。多重補完法は，補完モデルがいくつかの条件を満たせば，欠測メカニズムが MAR の下で妥当な推測を行うことができる（阿部, 2016; 野間, 2017）。本稿のような単調な欠測パターンのデータでは, 解析に用いる変数を欠測値が少ない順に並べたものを$y_1,\cdots,y_J$としたとき，y_1には欠測値がないものとして，$y_j(j=2,\cdots,J)$に対して$y_1,\cdots,y_{(j-1)}$を用いて補完のためのモデルを構築し，y_jの補完値を生成する方法が用いられる。なお，経時測定データの場合，$y_1,\cdots,y_J$は測定時点の順となり，ある時点の欠測値を補完するために，1 時点前までの時点のデータを用いて補完を行う。本稿では補完値の生成方法として，単調な欠測パターンに対する回帰法（単調回帰法），予測平均マッチング法および傾

向スコア法を用いる。なお，回帰法および予測平均マッチング法は，非単調な欠測パターンの補完に用いる fully conditional specification（FCS）法でも利用可能である（SAS Institute Inc., 2023）。

4.1 回帰法

回帰法（阿部, 2016; 野間, 2017）は，ANCOVA や MMRM と同様に正規性を仮定した回帰モデルから補完値を生成する方法である。時点jの測定値y_jが欠測値の場合，時点$j-1$までの測定値$y_1,\cdots,y_{(j-1)}$を用いた以下の回帰モデルを用いてy_jの補完を行う。

$$y_j = \beta_0 + \beta_1 y_1 + \cdots + \beta_{(j-1)} y_{(j-1)}$$

観測されたデータに基づいて回帰係数パラメータの推定値$\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_{(j-1)}\right)$およびその分散共分散行列，分散パラメータの推定値$\hat{\sigma}_j^2$を得る。補完値を生成するために以下の Step を繰り返す。

Step.1　回帰モデルの推定結果に基づいて，$\boldsymbol{\beta} = \left(\beta_0, \beta_1, \cdots, \beta_{(j-1)}\right)$および$\sigma_j^2$の事後予測分布から，新たなパラメータ$\boldsymbol{\beta}_* = \left(\beta_{*0}, \beta_{*1}, \cdots, \beta_{*(j-1)}\right)$および$\sigma_{*j}^2$をランダムサンプリングする。

Step.2　y_jが欠測している個体iに対する補完値$\hat{y}_{ij}$は，以下によって得られる。

$$\hat{y}_{ij} = \beta_{*0} + \beta_{*1} y_{i1} + \cdots + \beta_{*(j-1)} y_{i(j-1)} + u_i \sigma_{*j}$$

ここで，u_iは標準正規分布からの乱数である（y_{ij}の事後予測分布からランダムサンプリングしている）。

回帰法は，このようにy_jが正規分布に従うことを仮定した回帰モデルから補完値を生成するため，外れ値による正規性の仮定からのずれに影響を受ける可能性が考えられる。

4.2 予測平均マッチング法

予測平均マッチング法（阿部, 2016; 野間, 2017）は，連続変数に対して利用可能な補完値の生成法の 1 つであり，回帰法と類似しているが，回帰モデルに基づいて欠測値の予測値に近い予測値を持つ観測値を選択し，その観測値の集合からのランダムサンプリングによって各欠測値の補完値を生成する点が異なっている。回帰法と同様に回帰モデルのパラメータを推定し，補完値を生成するために以下の Step を繰り返す。

Step.1　回帰モデルの推定結果に基づいて，$\boldsymbol{\beta} = \left(\beta_0, \beta_1, \cdots, \beta_{(j-1)}\right)$および$\sigma_j^2$の事後予測分布から，新たなパラメータ$\boldsymbol{\beta}_* = \left(\beta_{*0}, \beta_{*1}, \cdots, \beta_{*(j-1)}\right)$および$\sigma_{*j}^2$をランダムサンプリングする。

Step.2　y_jが欠測している個体iに対する予測値y_{ij*}は，以下によって得られる。

$$y_{ij*} = \beta_{*0} + \beta_{*1} y_{i1} + \cdots + \beta_{*(j-1)} y_{i(j-1)}$$

Step.3　y_jの観測値のうち，予測値y_{ij*}に近い予測値を持つk_0個の観測値を補完値の候補の集合とする。

Step.4　k_0個の補完値の候補の集合から，欠測値y_{ij}の補完値をランダムサンプリングする。

予測平均マッチング法は，y_jが従う分布を仮定せずに観測された値から補完値を生成する hot deck な方法である。ただし，k_0を小さくし過ぎると補完間の相関が高くなり，補完による不確実性を過小評価する可能性がある。一方で，k_0を大きくし過ぎると回帰モデルに基づいて補完値の候補を選択する意義が薄れることに注意が必要である。

4.3 傾向スコア法

傾向スコア法（Lavori et al, 1995; 阿部, 2016）は連続変数に対し，欠測パターンが単調であるときに利用できる。傾向スコア（Rosenbaum and Rubin, 1983）は，通常は共変量が与えられた下で処置群に割り付けられる確率と定義されるが，欠測に関する傾向スコアは，共変量が与えられた下で測定値が欠測する確率と定

義される。傾向スコア法は，ある個体の欠測値に対して，同じような傾向スコアの値を持つ別の個体の観測値を補完する方法である。補完値の生成は以下の Step に基づく。

Step.1　y_jが欠測である場合は 0，観測されている場合は 1 となる指示変数R_jを用い，y_jが欠測する確率（欠測の傾向スコア）$p_j = Pr(R_j = 0 \,|\, y_1, \cdots, y_{(j-1)})$を，ロジスティック回帰モデル

$$logit(p_j) = \beta_0 + \beta_1 y_1 + \cdots + \beta_{(j-1)} y_{(j-1)}$$

に基づいて推定する。

Step.2　傾向スコアで昇順に並べた個体を事前に決めた数の層に分類する。

Step.3　層ごとに，y_jが観測されているn_1個の測定値から復元抽出によってn_1個の補完値の候補の集合を生成する。

Step.4　層ごとに，この大きさn_1の補完値の候補の集合から，n_0個のy_jの欠測値に対して補完値を復元抽出する。

傾向スコア法は，y_jが従う分布を仮定しないノンパラメトリックな方法である。傾向スコアに基づく層の数は，予測平均マッチング法の補完値の候補の数と同様に，解析者が決定する必要がある。傾向スコア法は，補完値を生成する際に$y_1, \cdots, y_j$間の相関の情報を考慮していないため，補完された個々の変数の推測には有効であるが，回帰分析のような変数間の関係を含む解析ではバイアスを生じさせる可能性がある（SAS Institute Inc., 2023; 阿部, 2016）。

4.4 MI プロシジャと ROBUSTREG プロシジャによる実行

2 節で生成した外れ値を含むシミュレーションデータを対象に SAS での実行方法を回帰法，予測平均マッチング法および傾向スコア法の順に説明する。本節では欠測値を発生させた後の不完全データを解析する。ベースラインおよび投与後 4 時点の合計 5 時点の測定値 Y1-Y5 を用いて治療群 TRT ごとに MI プロシジャで回帰法によって多重補完（補完回数は 100 回）を行った。補完後のデータセットに対し，最終時点でのベースラインからの変化量 CHG（Y5－Y1）を算出し（最終時点以外も解析可能である），ROBUSTREG プロシジャで 3.2 項と同様の ANCOVA 型のロバスト回帰を行った。得られた治療群の回帰係数パラメータ（最終時点の群間差）の推定値を MIANALYZE プロシジャで Rubin のルールによって統合した。SAS コードを以下に示す。

```
proc mi data=EXAMPLE seed=123456 out=OUTEX nimpute=100;
  var Y1 Y2 Y3 Y4 Y5;
  monotone reg(Y2 Y3 Y4 Y5 / details);
  by TRT;
run;

data OUTEX;
  set OUTEX;
  CHG = Y5 - Y1;
 run;
```

```
proc sort data=OUTEX;
  by _Imputation_;
run;

proc robustreg data=OUTEX;
  class TRT;
  model CHG = TRT Y1;
  by _Imputation_;
  ods output ParameterEstimates=OUT_REG(where=(Parameter='TRT' and Level1='1'));
run;

proc mianalyze data= OUT_REG;
  modeleffects estimate;
  stderr stderr;
run;
```

単調な欠測パターンとなる順番に VAR ステートメントで変数を指定する。指定順が単調な欠測パターンと合っていない場合は，「A MONOTONE statement is specified for a data set with a non-monotone missing pattern.」のようなエラーとなるため注意が必要である。次に MONOTONE 文で REG を指定し，括弧内に補完対象の変数を指定する。DETAILS オプションを指定することで，補完モデルの詳細が得られる。ROBUSTREG プロシジャの指定は，BY 文で補完回数ごとに実行する以外は 3.2 項と同じである。ODS OUTPUT で回帰係数パラメータの推定値および SE を出力し，MIANALYZE プロシジャで統合する。

予測平均マッチング法を行う場合，補完方法以外は回帰法と同様のため，SAS コードのうち MI プロシジャの部分を以下に示す。予測平均マッチング法では MONOTONE 文で REGPMM を指定すれば良い。補完候補の数k_0は K オプションで設定できる。本稿では補完候補の数をデフォルトの 5 とした。

```
proc mi data=EXAMPLE seed=123456 out=OUTEX nimpute=100;
  var Y1 Y2 Y3 Y4 Y5;
  monotone regpmm(Y2 Y3 Y4 Y5 / details k=5);
  by TRT;
 run;
```

傾向スコア法も同様であり，SAS コードのうち MI プロシジャの部分を以下に示す。MONOTONE 文で PROPENSITY を指定する点が異なっている。DETAILS を指定することにより，上記 Step.1 で計算したロジスティック回帰モデルの係数を表示できる。また，NGROUPS オプションで傾向スコアによって分類する層の数を指定できる（デフォルトは NGROUPS=5）。注意が必要な点として，層の数が多すぎるとエラーとなり，例えば「ERROR: パラメータ値 25 が 1 と 20 の有効範囲内にありません。」と表示される。また，層の数とデータの数の関係によっては補完値の候補が 1 つもないということが起こりえるが，1 つでもそういっ

た層があると別のエラーとなる。一方，層が少ないと傾向スコアを使っている意味が薄れる（例えば層の数を 1 に指定することもできるが，傾向スコアを計算する意味がなくなる）。本稿では層の数をデフォルトの 5 とした。

```
proc mi data=EXAMPLE seed=123456 out=OUTEX nimpute=100;
  var Y1 Y2 Y3 Y4 Y5;
  monotone propensity(Y2 Y3 Y4 Y5 / details ngroups=5);
  by TRT;
run;
```

表 2 に多重補完法と組み合わせたロバスト回帰の治療群の回帰係数パラメータ（最終時点の群間差）の結果を示す。また，比較のために欠測値を伴う経時測定データの解析方法としてよく用いられる MMRM（五所・丸尾, 2017）による結果も示した。MMRM は MIXED プロシジャで実行し（O'Kelly and Ratitch, 2014），ベースラインからの変化量を応答変数，群，時点，ベースライン値，群と時点の交互作用，ベースライン値と時点の交互作用を固定効果とし，共分散構造は unstructured を仮定した。結果の傾向は表 1 と同様であり，点推定値は MMRM と比べて多重補完法と組み合わせたロバスト回帰の方が真値の 0.46 に近い値だった（3.2 項の解析に用いたデータセットに対して欠測値を発生させたデータセットを本項の解析に用いている）。また，いずれの補完値の生成方法を用いた場合も，同様の結果が得られた。

表 2　多重補完法と組み合わせたロバスト回帰および MMRM の結果（最終時点の群間差）

			両側 95%信頼区間		
	推定値	SE	下限	上限	p 値
回帰法＋ロバスト回帰	0.69	0.20	0.29	1.09	0.0007
予測平均マッチング法＋ロバスト回帰	0.63	0.18	0.27	0.98	0.0005
傾向スコア法＋ロバスト回帰	0.65	0.19	0.27	1.03	0.0008
MMRM	1.21	0.26	0.70	1.72	<.0001

5. シミュレーション実験

5.1 シミュレーション実験の計画

本節では，2 節の外れ値と欠測値を含み得る仮想の臨床試験のシミュレーションデータに基づいて，ロバスト回帰，多重補完法と組み合わせたロバスト回帰，正規性を仮定した ANCOVA および MMRM を比較した。表 3 に示す 4 つのシナリオに基づいて，欠測値がない場合と欠測値がある場合のシミュレーションデータを乱数で生成した。シナリオ 1 は 2 節と同じ設定であり，シナリオ 2 は群間差がないシナリオとして，実薬群の平均ベクトルを$\boldsymbol{\mu}_A = (0.00, 0.00, 0.00, 0.00, 0.00)^T$に変更した。シナリオ 3 は外れ値なしのシナリオとして，応答変数の分布を多変量正規分布に変更（混合比を 1：0 に変更）した。シナリオ 4 は外れ値がなく，群間差もないシナリオとした。

表 3 シミュレーション実験のシナリオ

シナリオ	応答変数の分布	真の群間差
1	多変量混合正規分布（外れ値あり）	あり
2	多変量混合正規分布（外れ値あり）	なし
3	多変量正規分布（外れ値なし）	あり
4	多変量正規分布（外れ値なし）	なし

欠測値なしのデータに対して，3.2 項と同様の方法で，ロバスト回帰および ANCOVA による解析を行った結果を比較した。また，欠測値ありのデータに対して，4.4 項と同様の方法で，多重補完法（回帰法，予測平均マッチング法および傾向スコア法）と組み合わせたロバスト回帰および MMRM による解析を行った結果を比較した。各シナリオおよび各手法について，それぞれ 5000 回反復計算を実施し，以下の評価指標（Morris et al., 2019）を計算した。

- バイアスの平均値：バイアス（群間差の推定値－群間差の真値）の平均値
- Root mean squared error（RMSE）：バイアスの平均平方の平方根
- 第 1 種の過誤確率：帰無仮説下（真の群間差がなしのシナリオ）で群間差が 0 という帰無仮説が両側有意水準 5%で棄却された割合
- 検出力：対立仮説下（真の群間差がありのシナリオ）で群間差が 0 という帰無仮説が両側有意水準 5%で棄却された割合
- 95%信頼区間の被覆確率：群間差の真値が群間差の両側 95%信頼区間の中に含まれた割合
- モデル SE の相対誤差：$100 \times (\text{モデル SE}/\text{経験 SE} - 1)$

経験 SE：群間差の推定値の標準偏差，モデル SE：群間差の SE の推定値の平均値（SE の 2 乗の平均値の平方根）

5.2 シミュレーション実験の結果

図 3～図 8 に各指標のシミュレーション実験の結果を示す。

バイアスの平均値は欠測値の有無（図 3 左および図 3 右），外れ値の有無，群間差の有無によらず，どの解析方法でも 0 に近い値だった。

RMSE（バイアスが 0 の場合は経験 SE と同等）は，欠測値がない場合（図 4 左），外れ値ありのシナリオでは ANCOVA の RMSE がロバスト回帰の RMSE よりも大きかった。一方で，外れ値なしのシナリオでは，ANCOVA の RMSE の方が若干小さかった。欠測値がある場合（図 4 右）でも同様の傾向であり，外れ値ありのシナリオではロバスト回帰の各方法の RMSE の方が小さく，外れ値なしのシナリオでは MMRM の RMSE が若干小さかった。各補完値の生成法と組み合わせたロバスト回帰同士の比較では, RMSE は同程度だった。

第 1 種の過誤確率は，欠測値がない場合（図 5 左）では外れ値の有無によらずロバスト回帰，ANCOVA 共に名義水準である 0.05 に近い値だった。欠測値がある場合（図 5 右），回帰法＋ロバスト回帰は，外れ値ありのシナリオでは 0.05 より低く，外れ値なしのシナリオでは 0.05 に近い値だった。予測平均マッチング法＋ロバスト回帰は，外れ値の有無によらず 0.05 よりも若干高かった。傾向スコア法＋ロバスト回帰は，外れ値の有無によらず 0.05 よりも低い値だった。MMRM は外れ値の有無によらず 0.05 に近い値だった。

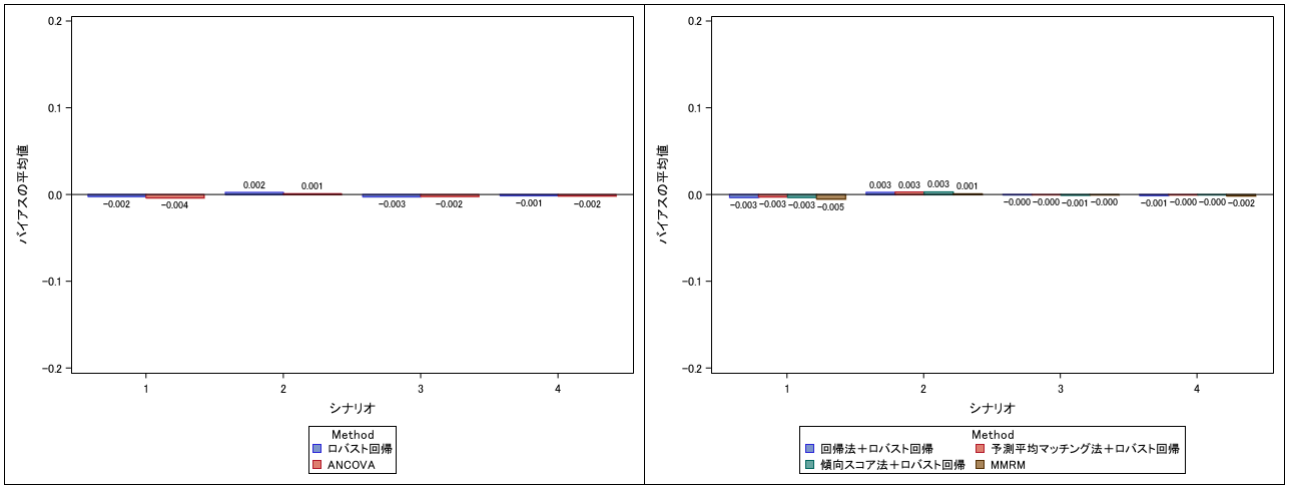


図 3 各解析法のバイアスの平均値（左：欠測値なし，右：欠測値あり）

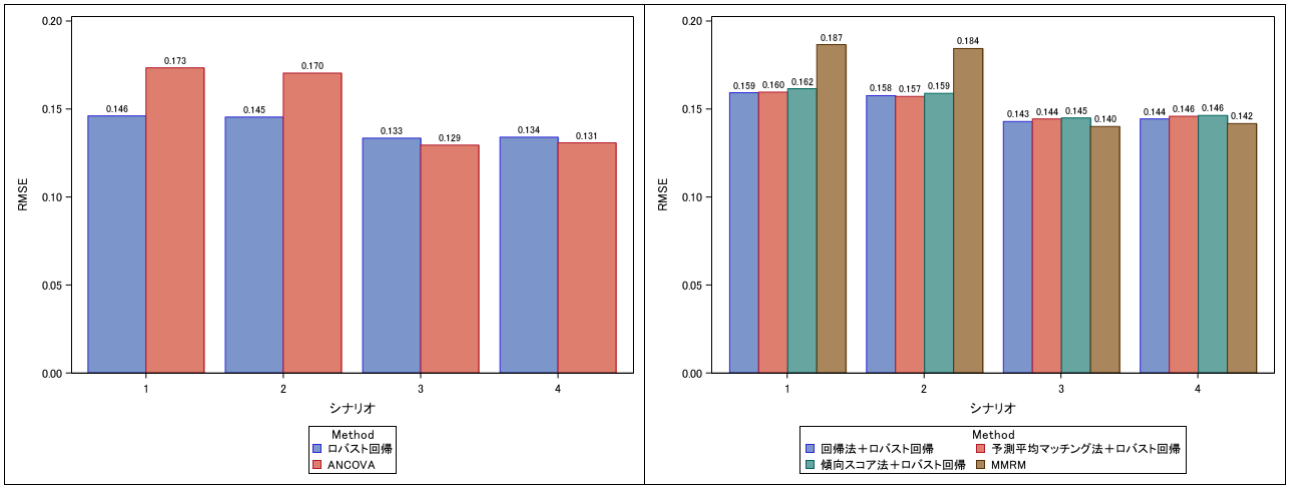


図 4 各解析法の RMSE（左：欠測値なし，右：欠測値あり）

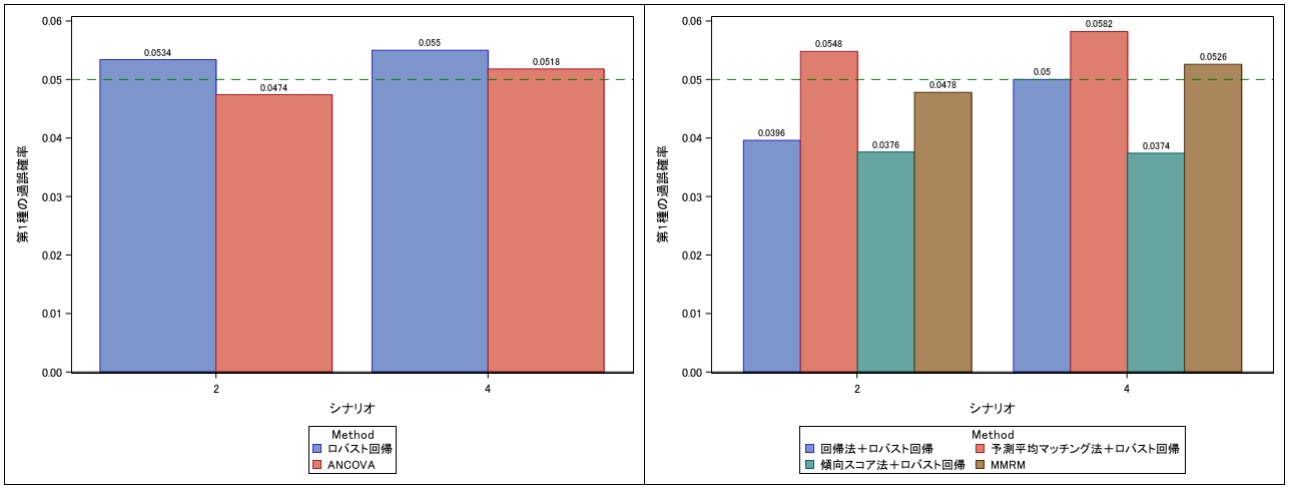


図 5 各解析法の第 1 種の過誤確率（左：欠測値なし，右：欠測値あり）

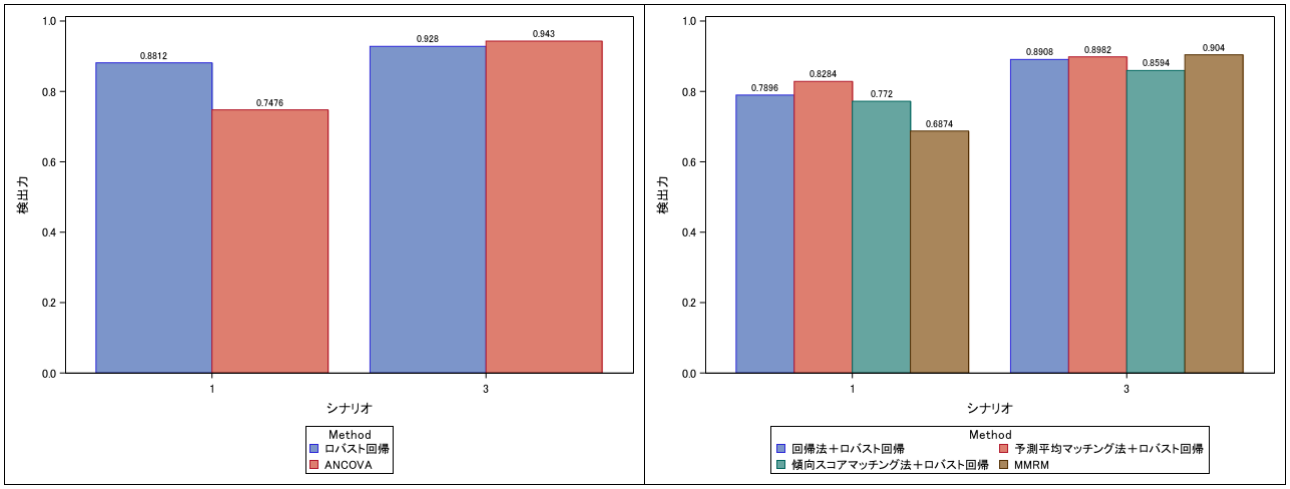


図 6 各解析法の検出力（左：欠測値なし，右：欠測値あり）

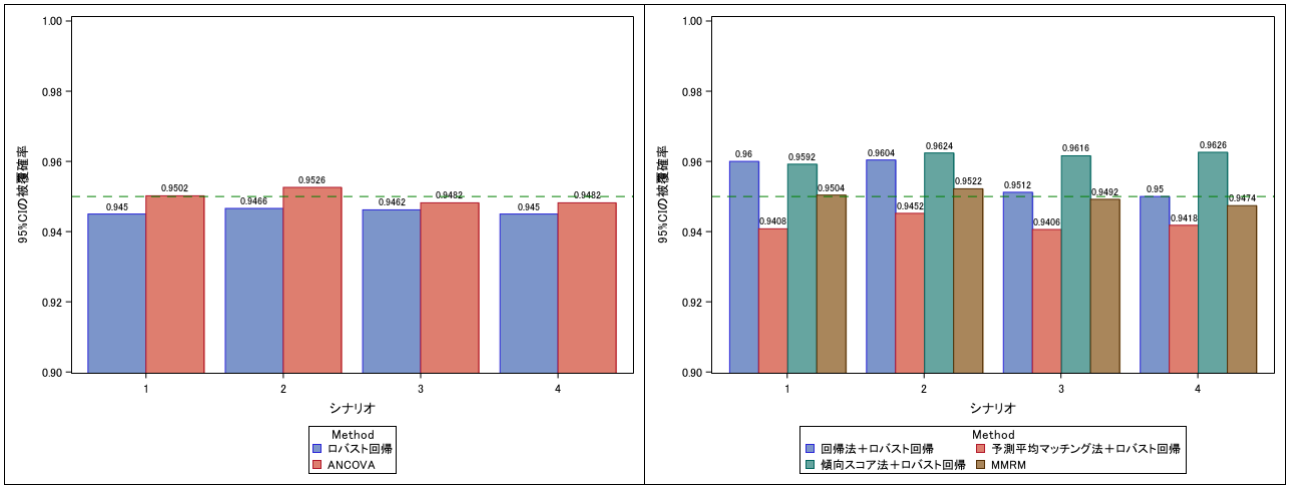


図 7 各解析法の 95%信頼区間の被覆確率（左：欠測値なし，右：欠測値あり）

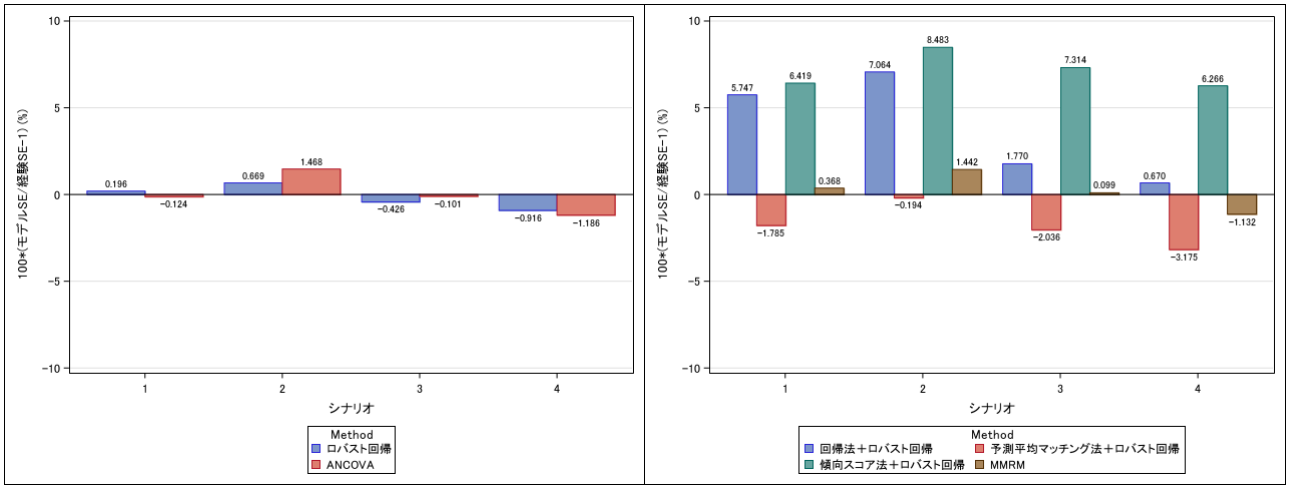


図 8 各解析法のモデル SE の相対誤差（%）（左：欠測値なし，右：欠測値あり）

検出力は，欠測値がない場合（図 6 左），外れ値ありのシナリオでは ANCOVA と比較してロバスト回帰の方が高かった。一方で，外れ値なしのシナリオでは，ANCOVA の検出力が若干高かった。欠測値がある場合（図 6 右）も同様の傾向であり，外れ値ありのシナリオではロバスト回帰の各方法の検出力が高く，外れ値なしのシナリオでは MMRM の検出力が若干高かった。各補完値の生成法と組み合わせたロバスト回帰同士の検出力の比較は，第 1 種の過誤確率と同様の傾向だった。

95%信頼区間の被覆確率は，欠測値がない場合（図 7 左）では外れ値の有無によらずロバスト回帰，ANCOVA 共に名義水準である 0.95 に近い値だった。欠測値がある場合（図 7 右），回帰法＋ロバスト回帰は，外れ値ありのシナリオでは 0.95 よりも高い値であり，外れ値なしのシナリオでは 0.95 に近い値だった。予測平均マッチング法＋ロバスト回帰は，外れ値の有無によらず 0.95 よりも低い値だった。傾向スコア法＋ロバスト回帰は，外れ値の有無によらず 0.95 よりも高い値だった。MMRM は，外れ値の有無によらず 0.95 に近い値だった。

群間差の SE（モデル SE）が群間差の推定値の本来のばらつき（経験 SE）と一致しているかを評価するモデル SE の相対誤差は，欠測値がない場合（図 8 左）では外れ値の有無によらずロバスト回帰，ANCOVA 共に 0 に近く，経験 SE とモデル SE が同程度だった。欠測値がある場合（図 8 右），回帰法＋ロバスト回帰は，外れ値ありのシナリオでは正の誤差を示し，経験 SE よりもモデル SE が大きかったが，外れ値なしのシナリオでは 0 に近い値だった。予測平均マッチング法＋ロバスト回帰は，外れ値の有無によらず負の誤差を示す傾向があり，経験 SE よりもモデル SE が小さかった。傾向スコア法＋ロバスト回帰は，外れ値の有無によらず正の誤差を示し，経験 SE よりもモデル SE の方が大きかった。各補完値の生成法と組み合わせたロバスト回帰のモデル SE の相対誤差の結果は，上述の第 1 種の過誤確率，検出力，95%信頼区間の被覆確率で認められた傾向と一致していた。MMRM は外れ値の有無によらず 0 に近い値だった。

6. 考察とまとめ

6.1 臨床試験でのロバスト回帰の実践

本稿では，ロバスト回帰および多重補完法とロバスト回帰を組みあわせて外れ値が存在する仮想データの解析を行うとともに，簡単なシミュレーション実験で解析方法の性能評価を行った。本稿のシミュレーション実験では，実薬群とプラセボ群の応答変数が同じ対称分布に従っており，その平均の群間差を評価していたことから，外れ値がある状況でもいずれの解析方法も平均的なバイアスはなかった（データや解析の内容が異なれば外れ値がバイアスを生じさせることはありうる）。一方で，シミュレーション実験の個々の結果では，3.2 項および 4.4 項の解析結果のように正規性を仮定した解析は外れ値の影響を受ける場合があるのに対して，ロバスト回帰では外れ値の影響が緩和され，その結果としてロバスト回帰の RMSE の方が小さく，正規性を仮定した ANCOVA および MMRM よりも頑健な推定値が得られることが示された。また，正規性が成立する状況でもロバスト解析の推定精度は正規性を仮定した解析方法と比べてほとんど低下しないことが示された。これらのことから，臨床試験の解析を計画する際，あらかじめ外れ値が生じることが想定される評価項目に対して，ロバスト回帰を主たる解析として計画することが考えられる（Raal et al. 2020; Robinson et al., 2015; 医薬品医療機器総合機構, 2019）。ICH E9 ガイドライン「臨床試験のための統計的原則」（厚生省医薬安全局審査管理課長, 1998）では，治験実施計画書中に外れ値に対処する方法が記載されていない場合，実際の値を用いた解析のほかに，外れ値の影響を除くか小さくする別の解析を少なくとも 1 つ行うべきであり，それらの結果の間の差異を議論すべきと述べられており，ICH E9(R1)ガイドライン「臨床

試験のための統計的原則 補遺 臨床試験における estimand と感度分析」（厚生労働省医薬局医薬品審査管理課長 ,2024）でも，主とする推定量による推測について，その仮定からのずれに対する安定性を調べるために，同じ estimand を対象とした 1 つ以上の解析として，感度分析を実施すべきと述べられている。したがって，評価項目の主たる解析方法として，ANCOVA または MMRM といった正規性を仮定した方法を用いる場合でも，外れ値が生じる可能性が否定できない場合，正規性の仮定に対する感度分析としてロバスト回帰を用いることも考えられる。ただし，多重補完法と組み合わせて経時測定データに適用する場合，いくつかの注意が必要である。

補完値の生成方法として回帰法，予測平均マッチング法および傾向スコア法とロバスト回帰を組み合わせた際の RMSE は同程度であり，回帰係数パラメータの推定値に違いはなかったが，その SE の推定値には補完値の生成方法によって違いがあり，その違いが検出力および第 1 種の過誤確率，95%信頼区間の被覆確率にも影響していると考えられた。回帰法とロバスト回帰を組み合わせた際，外れ値がある（すなわち補完モデルを誤特定している）状況では，経験 SE よりもモデル SE が大きく，推定値のばらつきを過大推定していることが示された。Mehrotra et al. (2012)でも回帰法とロバスト回帰を組み合わせた際に，非正規分布の下では Rubin のルールに基づく SE は真のばらつきよりも大きく推定される傾向があり，第 1 種の過誤確率が名義水準を下回ると報告されている。また，多重補完法での解析結果の統合方法として，Rubin のルールの代わりに Robins and Wang (2000)の方法を用いることで，よりバイアスのない SE の推定値が得られ，第 1 種の過誤確率が名義水準に近くなることをシミュレーション実験で示している。なお，Robins and Wang の方法は現時点（SAS 9.4 TS1M8）では SAS のプロシジャで実装されていない。また，予測平均マッチング法および傾向スコア法は外れ値の有無によらず同様の傾向であり，回帰法と比べて外れ値の影響を受けにくいことが示唆されたものの，予測平均マッチング法では経験 SE よりもモデル SE が小さく，傾向スコア法では経験 SE よりもモデル SE が大きい傾向があった。なお，本稿のシミュレーション実験では，これらの方法の補完値の候補数や傾向スコアの層数は SAS のデフォルトの設定のみを用いており，シミュレーション実験のシナリオも包括的ではないため，補完値の生成方法が SE の推定値に与える影響については，更なる検討が必要と考えられる。

6.2 他のロバスト回帰

本稿ではロバスト回帰の代表的な方法であり，ROBUSTREG プロシジャのデフォルトである Huber の M 推定を用いた。Huber の M 推定は応答変数の外れ値に対しては頑健であるが，説明変数の外れ値に対しては頑健ではない（SAS Institute Inc., 2023; 蓑谷, 2016）。また，Tukey の bisquare 重み関数のデフォルトの調整定数c =4.685 を用いた場合，ロバスト回帰が頑健性を維持できる外れ値の割合として定義される breakdown value は約 12%であり（蓑谷, 2016），外れ値の割合がそれ以上になると M 推定は外れ値の影響を受けてしまう。ROBUSTREG プロシジャでは，M 推定以外にも，least trimmed squares（LTS）推定，S 推定および MM 推定によるロバスト回帰を行うことができる（SAS Institute Inc., 2023）。これらの方法の概要を紹介する。

LTS 推定（SAS Institute Inc., 2023; 蓑谷, 2016）は，残差平方を小さい順に並び変えた残差平方の順序統計量を$r_{(1)}^2 \le r_{(2)}^2 \le \cdots \le r_{(n)}^2$としたとき，その$h$番目までの和$\sum_{i=1}^{h} r_{(i)}^2$を最小化することにより回帰係数を推定する。$\frac{n}{2}+1 \le h \le \frac{3n+p+1}{4}$であり，$h=\frac{n}{2}+1$のとき約半数の残差の大きいデータを無視することとなり，breakdown value は上限の 50%（50%を超えると外れ値の方が多いことになる）を達成できる。一方で，LTS 推定は残差の大きいデータを完全に無視する方法であり，誤差項が正規分布に従っている場合（以下同様）の漸近的有効性が低いという問題がある。

S 推定（SAS Institute Inc., 2023; 蓑谷, 2016）は，

$$\frac{1}{n-p}\sum_{i=1}^{n}\rho\left(\frac{y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}}{S}\right)=b \tag{2}$$

のS，つまりスケールパラメータの推定値を最小化する回帰係数を推定する方法である。ここで，$b=\int\rho(z)d\Phi(z)$である。$\rho(z)$として，ROBUSTREG プロシジャでは Tukey の bisquare 重み関数（(1)式）と対応する

$$\rho(z)=\begin{cases}3\left(\frac{z}{c}\right)^2-3\left(\frac{z}{c}\right)^4+\left(\frac{z}{c}\right)^6, & \text{if } |z|\le c\\ 1 & \text{otherwise}\end{cases}$$

がデフォルトで用いられる。S 推定のデフォルトの調整定数c =2.9366 の場合，breakdown value は 25%および漸近的有効性は 75.9%であり，M 推定のデフォルト（breakdown value：12%，漸近的有効性：95%）と比べて breakdown value は高いものの，漸近的有効性は低い。

MM 推定（SAS Institute Inc., 2023; 蓑谷, 2016）は，高い breakdown value と高い漸近的有効性を併せ持つ方法であり，以下の Step で推定される。

Step.1　LTS 推定または S 推定などの高い breakdown value を持つ推定方法で初期の回帰係数の推定値を得る。

Step.2　Step.1 で推定した初期の回帰係数の推定値を用いて，(2)式からスケールパラメータを推定する。調整定数は，breakdown value が高くなるように設定する。

Step.3　スケールパラメータを Step.2 で推定した値に固定し，以下の損失関数を最小化することによって最終的な回帰係数の推定値を得る。調整定数は，漸近的有効性が高くなるように設定する。

$$Q(\boldsymbol{\beta})=\sum_{i=1}^{n}\rho\left(\frac{y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}}{\hat{\sigma}}\right)$$

MM 推定の breakdown value は Step.1 および Step.2 で決まる。Step.1 の LTS 推定で$h=\frac{n}{2}+1$，Step.2 の調整定数をc =1.548 に設定することで，50%の breakdown value を達成できる。更に，Step.3 の調整定数を M 推定のデフォルトと同じc =4.685 に設定することで，95%の漸近的有効性を達成できる。3.2 項のロバスト回帰（M 推定）と同じ解析を MM 推定で行うための SAS コードを以下に示す。

```
proc robustreg data= EXAMPLE method=mm(inith=101 k0=1.548 eff=0.95) seed=1234;
  class TRT;
  model CHG = TRT Y1;
run;
```

PROC ROBUSTREG 文で METHOD=MM を指定することで MM 推定を実行できる。括弧内のサブオプションは，INITH が Step.1 が LTS 推定の場合のh，K0 が Step.2 の調整定数，EFF が Step.3 の調整定数（漸近的有効性の値で指定する）に関する指定である。結果は 3.2 項の M 推定とほぼ同じなので割愛する。なお，MM 推定のデフォルトの設定（Step.1 の$h=\frac{3n+p+1}{4}$，Step.2 の調整定数c =2.9366，Step.3 の調整定数c =3.440）では breakdown value は 25%，漸近的有効性は 85%となっており（SAS Institute Inc., 2023），設定を変更しないと MM 推定の性能を十分に発揮できない可能性があることに注意が必要である。しかしながら，S 推定および MM 推定では漸近的有効性の値を指定できるものの，それ以外では breakdown value およ

び漸近的有効性は調整定数などを介してしか指定できず，これらを意図した値にするためにはロバスト回帰に対する知識が必要となる。また，ROBUSTREG プロシジャで採用されている LTS 推定および S 推定の計算アルゴリズムでは，データセットからサブセットのサンプリングを行うため，ユニークな値が少ない変数（例えば，カテゴリカル変数のダミー変数）がすべて同じ値になってしまいモデルが退化するといった計算上の問題が生じる可能性がある（SAS Institute Inc., 2023）。これらのことから，MM 推定は高い breakdown value と高い漸近的有効性を併せ持つ有用な方法ではあるものの，M 推定で対処可能と考えられる状況では，M 推定の方が実用的と考えられる。また，M 推定で対処できないような外れ値の割合が高い状況では，そもそも外れ値とすべきか検討が必要となる場合も考えられる。

6.3 まとめ

本稿では，SAS で ROBUSTREG プロシジャを用いてロバスト回帰を行う方法および多重補完法と組み合わせて経時測定データ解析に適用する方法を説明し，簡単なシミュレーション実験でこれらの解析方法の性能を評価することで，臨床試験でのロバスト回帰が活用できる状況や，その際の留意事項を示した。なお，ICH E9 ガイドライン（厚生省医薬安全局審査管理課長, 1998）では，特定の値が外れ値であると明確に判断するのに最も説得力があるのは，統計的に正当であるだけではなく医学的にも正当な場合であり，医学的な背景によりしばしば適切な対応が決定される例がみられると述べられている。そのため，ロバスト回帰によって外れ値の影響を弱めて解析する前に，外れ値がどういった背景から生じており，全体の結果にどのように影響しているかを検討し，またその影響がロバスト回帰によってどのように対処されているかを確認することが重要と考えられる。

本稿が，SAS で外れ値に対処するためにロバスト回帰を用いる際の参考になれば幸いである。

参考文献

Lavori PW, Dawson R, Shera D (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine*. 14:1913-1925.

Mehrotra DV, Li X, Liu J, Lu K (2012). Analysis of longitudinal clinical trials with missing data using multiple imputation in conjunction with robust regression. *Biometrics* 68(4):1250-1259.

Morris TP, White IR, Crowther MJ (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 38(11):2074-2102.

O'Kelly M, Ratitch B (2014). Clinical Trials with Missing Data -A Guide for practitioners-. John Wiley & Sons.

Raal FJ, Rosenson RS, Reeskamp LF, Hovingh GK, Kastelein JJP, Rubba P, Ali S, Banerjee P, Chan KC, Gipe DA, Khilla N, Pordy R, Weinreich DM, Yancopoulos GD, Zhang Y, Gaudet D; ELIPSE HoFH Investigators (2020). Evinacumab for Homozygous Familial Hypercholesterolemia. *N Engl J Med*. 383(8):711-720.

Robins JM, Wang N (2000). Inference for imputation estimators. *Biometrika* 87**:**113‑124.

Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, Averna M, Stroes ES, Langslet G, Raal FJ, El Shahawy M, Koren MJ, Lepor NE, Lorenzato C, Pordy R, Chaudhari U, Kastelein JJ; ODYSSEY LONG TERM Investigators (2015). Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N Engl J Med*. 372(16):1489-1499.

Rosenbaum PR, Rubin DB (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70:41–55.

SAS Institute Inc. (2023). SAS/STAT 15.3 User's Guide. SAS Institute Inc., Cary, NC, USA.

阿部 貴行 (2016). 欠測データの統計解析. 朝倉書店.

医薬品医療機器総合機構 (2019). ロスーゼット配合錠 LD, 同配合錠 HD, 審査報告書. https://www.pmda.go.jp/drugs/2019/P20190325001/170050000_23100AMX00300_A100_1.pdf [2025/8/5 アクセス]

岩崎 学 (2006). 統計的データ解析入門 ノンパラメトリック法. 東京図書.

厚生省医薬安全局審査管理課長 (1998). 「臨床試験のための統計的原則」について. https://www.pmda.go.jp/files/000156112.pdf [2025/8/5 アクセス]

厚生労働省医薬局医薬品審査管理課長 (2024). 「臨床試験のための統計的原則」の補遺について. https://www.pmda.go.jp/files/000269154.pdf [2025/8/5 アクセス]

五所 正彦, 丸尾 和司 (2017). 経時測定データ解析における mixed-effects models for repeated measures（MMRM）の利用. 応用統計学 46(2):53-65.

橋詰 公一 (2008). 経時臨床データに対する Robust Linear Mixed Models の適用 -SAS/lML を利用した MCMC 法による推定-. SAS ユーザー総会論文集 2008.

野間 久史 (2017). 連鎖方程式による多重代入法. 応用統計学 46(2):67-86

蓑谷 千凰彦 (2016). 頑健回帰推定. 朝倉書店.

森川 馨, 田崎 武信 (監訳) (2009). 治験の統計解析 理論と SAS®による実践. 講談社サイエンティフィク.

兵庫県下の消防本部における救急搬送データのJMPを活用した分析

○川向　肇[1]・野田晴菜[2]
（[1]兵庫県立大学社会情報科学部・大学院情報科学研究科・ソーシャルデータサイエンス研究所,
[2]兵庫県立大学社会情報科学部）

A Statistical Analysis on Ambulance Service Time Data in Hyogo, utilizing JMP

Hajime Kawamukai, Haruna Noda
School of Social Information Sciences, Graduate School of Information Sciences, Research Institute for Social Data Sciences / University of Hyogo

要旨

本稿では，総務省消防庁救急企画室救急連携係から提供を受けた 2015 年から 2023 年に実際に発生した救急搬送に関するデータを利用し，兵庫県下の消防本部ごとの救急搬送関連時間に関する分析結果を報告する．

まず，救急搬送に関する入電時刻から現場到着時刻までの現場到着所要時間，および入電時刻から病院等で医師に引き渡されるまでの病院収容に関する記述統計（最小値，四分位，25%分位，中央値，75%分位，90%分位，最大値，平均値）を求め，これらに関する特性についての検討を行った．さらに消防本部単位について市民が居住している面積や緊急輸送道路長などの空間的特性値を含む各種属性値に関するデータベースを構築し，現場到着所要時間や病院収容所要時間の 10%分位範囲（10%分位点から 90%分位点の間の範囲）の対数値を従属変数とする回帰モデルによる分析を行った結果を紹介する．

キーワード：救急搬送データ，消防庁，兵庫県，救急搬送所要時間，空間的要素

---本文開始点---

１．はじめに

1.1　研究の背景

消防法第 2 条 9 項では，救急業務について以下のように定義されている．

> 救急業務とは、災害により生じた事故若しくは屋外若しくは公衆の出入する場所において生じた事故（以下この項において「災害による事故等」という。）又は政令で定める場合における災害による事故等に準ずる事故その他の事由で政令で定めるものによる傷病者のうち、医療機関その他の場所へ緊急に搬送する必要があるものを、救急隊によつて、医療機関（厚生労働省令で定める医療機関をいう。第七章の二において同じ。）その他の場所に搬送すること（傷病者が医師の管理下に置かれるまでの間において、緊急やむを得ないものとして、応急の手当を行うことを含む。）をいう。

ところで，総務省消防庁（2024）では，高齢化社会の進展に伴う救急活動所要時間の延伸を予防することの重要性に関して，以下のような記述がなされている．

> *令和６年に行った将来推計によると、高齢化の進展等により救急需要は今後増大する可能性が高いことが示されており、救急活動時間の延伸を防ぐとともに、これに伴う救命率の低下を防ぐための対策が必要である。*

この指摘を踏まえるならば，高齢化の進展状況や救急活動に影響すると考えられる道路の整備状態や各消防本部内の消防署が管轄する面積やその管轄内の人口に関する空間分布，積雪状態を中心とする気象条件といった救急活動時間の延伸との何らかの関係が生じている可能性の強い要因について，定量的に捕捉しておく必要があろう．

筆者とその共同研究者は，兵庫県北部の豪雪地帯を含む消防本部の所轄区域において，積雪による救急搬送の影響を確認する研究を企図したが，その研究に着手するにあたり，各消防本部が置かれた環境，具体的には，消防本部ごとの管轄域の面積，管轄区域における人口が分布する面積，実効人口密度，管轄区域の物流道路，消防署の設置状況や救急自動車の配備状況といった諸属性により，救急搬送に関する所要時間がどのような要因により影響を受けているのかについて，定量的に分析を行い，その実態を把握するため，救急搬送人員データの個票データをもとに記述分析統計を行い，実態の把握に努めることとした．

類似の視点に立った研究としては，山岡（2015）の消防本部単位の住民基本台帳による人口密度，面積当たりの救急病院数，面積当たりの消防署数，住民基本台帳による高齢化率，交通センサスから導出した道路混雑度や昼夜間人口比などの指標を利用した研究があるが，同論文では，人口密度，面積当たりの救急病院数，同消防署数を地理的要因として利用しようとしているものの，これらの指標間に多重共線性があることから，これらの変数により主成分分析を行い地理的要素に関する成分の変数を独自に作成し，その変数に加え，管轄外搬送率や救急病院に占める公的病院の比率，人口当たり搬送件数や人口当たり搬送件数時間帯比，高齢化率，交通センサスデータから計算した都道府県単位の混雑度や交通量昼夜比を利用した回帰分析を行っている．

実際には，救急搬送に利用できる国道や主要道の整備状況は一様ではなく，また消防本部の管轄領域内の人口分布は，一様ではない．より具体的に言うならば，山間部を管轄する消防本部の管轄区域では，全く人が居住していない山林地がある半面，河川や道路，営農可能な平坦な地区では，相対的に高密度での居住傾向が確認される．実際，山間部を管轄する消防本部の管轄区域内においては，DID 地区の面積および DID 地区内の人口も限られる．救急搬送に影響する要因としては，非可住地を含む単純な人口密度や単純な高齢化率というよりは，実効人口密度や実数としての管轄内に居住する高齢者数ではないか，と考えた．

また，鈴木・平田（2023）では，2019 年の茨城県の救急搬送人員データによる搬送時間（病院収容所要時間）に消防本部間で差異があることに着目し，疾患の種別や年齢層，重症度などを含めたモデル化を行い，搬送にかかわる出発地と到着地（病院）を推定するシミュレーションモデルを構築している．その上で，シミュレーションモデルによる結果を用いて，地域での人口減少・少子高齢化の動向が病院収容所要時間にどのような影響をあたえ得るのか，診療科目および病院の減少が病院収容所要時間にどのように影響するのか，空飛ぶクルマの導入による現場到着所要時間や病院収容所要時間への効果などについて分析しているが，あくまでも厳密な出発地と到着地ではなく，シミュレーション結果を利用した分析であるという限界がある．

本稿では，これらの問題意識に立ち，近年極めて充実してきた空間的オープンデータを可能な限り利用し，どのような空間的要素や消防本部単位での救急車両の配備状況が現場到着所要時間や病院収容所要時間にどのように影響するのかについて，日本海から太平洋側までを含み，北部の山間地や南部の平坦地，島しょ部などを含み地形的要素の点でも，空間的要素の点でも多様な環境を有する兵庫県についての救急搬送に関する諸要因についての空間的データを可能な限り収集・利用して，分析することとした．

1.2 利用した救急搬送人員データについて

筆者らは，本研究の着手にあたって，総務省消防庁救急企画室に，救急搬送人員データの利用申請を行い，2008 年から 2023 年の全国の救急搬送に関するデータの提供を受けた．2008 年から 2014 年までのデータと，2015 年以降のデータでは，提供されているデータ項目や，データの欠損値がかなり異なるため，欠損値がほぼ存在せず，データの組成が安定している 2015 年から 2023 年までのデータを利用した．

この救急搬送人員データは，救急搬送ごとに，一連番号，西暦による年次，都道府県コード，消防本部コード（6 桁），事故種別（傷病　3 桁）発生年月日（時刻情報無し），入電時刻（年月日時分），指令時刻（年月日時分），現場到着時刻（年月日時分），傷病者接触時刻（年月日時分），車内収容時刻（年月日時分），現場出発時刻（年月日時分），収容時刻（年月日時分），傷病者番号，年齢区分（新生児・乳幼児・少年・成人・高齢者の 5 区分），性別，年齢，居住地（管内・管外・外国人等その他の 3 区分），発生場所大分類（住宅・公衆出入場所・仕事場・:道路・その他の 5 区分），救急搬送先機関（告示・告示外の 2 区分），救急搬送先機関（管内・管外の 2 区分），3 次医療機関の有無，初診医による重症度区分（6 区分），救急救命士の同乗の有無，救命士の各種資格の有無，医師の搭乗の有無，口頭指導の有無，指導実施の場合の対象（家族等 7 区分），指導実施がない場合の様態（5 区分），現場出場の医師種別（3 区分），転送回数とその理由，具体的な応急措置の種別（88 種別）の項目が含まれており，1 レコードあたり 133 項目で構成された救急搬送に関する特定年の複数の都道府県のレコードが 1 つの Excel ファイルに収録される形で提供されている．なお，2020 年以前と 2021 年以降ではレコードの項目順に若干の違いが存在している．

当初は，Excel のフィルタリング機能を利用し，特定府県のデータを抽出していたが，ファイルの読み込み，抽出作業にかなりの時間がかかるため，Excel ファイルを JMP に直接読み込み，抽出した．

1.3 現地到着所要時間および病院搬送時間の分位点・要約統計量について

総務省消防庁（2025）年における第 2 章第 5 節「救急態勢」では，全国での救急自動車に対する出動回数と入電時刻（119 番通報を受けた時刻）から現場に到着した時刻（現場到着時刻）の時間である現地到着所要時間と入電時刻から病院等で医師に引き継いた時刻（収容時刻）までに要した時間である病院収容所要時間のそれぞれについて，以下のような記述がみられる．

> *救急自動車による全国の救急出動件数は*、763 *万*8,558 *件（対前年比* 40 *万*8,986 *件増*、5.7%*増）となっている。これは*1 *日平均では約*2 *万*928 *件（同約*1,121 *件増）*

であり，現場到着所要時間に関しては，

> *現場到着所要時間の平均は約*10.0 *分（前年約*10.3 *分）となっており、（中略）令和元年と比べ、約*1.3 *分延伸している*

同じく病院収容所要時間についても

病院収容所要時間の平均は約 45.6 分（前年約 47.2 分）となっており、（中略）令和元年と比べ、約 6.1 分延伸している

といずれも所用時間の増大（延伸）傾向にあることが示されている．

ところで，兵庫県全体での出動回数や現場到着所要時間，病院収容所要時間について検討してみよう．まず，出動回数であるが，兵庫県の場合，2015 年 1 年間の出動回数は 236,381 回であったが，2015 年以降，出動回数はおおむね逓増している（図 1）．ただし，2020 年には，COVID-19 の流行により社会活動そのものが若干停滞したこともあるためか対前年の出動回数と比べ若干の減少がみられたが，それ以降については出動回数自体は逓増し，2023 年の出動回数は，293,104 回を数えるに至った．

次に，兵庫県全体における 2015 年以降の現場到着所要時間と病院収容所要時間の平均値と分位点がどのように変化したのかを図 2 および図 3 に示した．平均値で見る限り，現場到着所要時間については 2022 年の値と比べ 2023 年は微減しているが，2015 年以降，10%分位点，75%分位点，および 90%分位点において微増傾向にある．病院収容所要時間に関しては，2022 年までは，すべての分位点および平均値において微増しているが，2023 年は，10%分位点においてのみ指標値の微増が確認されるものの，10%分位点以外の分位点および平均値においては対前年の各指標と比べ微減している．総務省消防庁（2025）と同様に 2019 年の現場到着所要時間および病院収容所要時間と比較した場合，2023 年の平均値については，微増しており延伸傾向にあるものの，中央値についてはより小さな値を取っている．病院収容所要時間に関しては 2019 年の各指標値の比較において，いずれの分位点および平均値ともより大きな値であり，兵庫県でも病院収容所要時間における延伸傾向が確認された．

まず，JMP で年次別の兵庫県内のみの救急搬送に関するデータとして編成し，入電時刻から現場到着時刻まで現場到着所要時間（分単位），現場到着時刻から傷病者接触時刻までの現場移動所要時間（分単位），

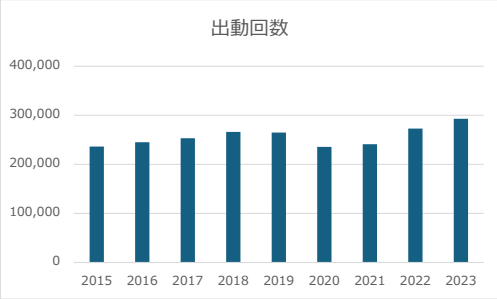
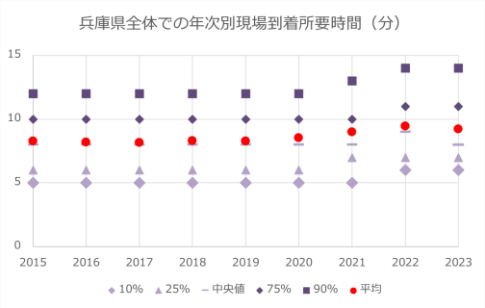


図 1　兵庫県下の救急自動車の出動回数



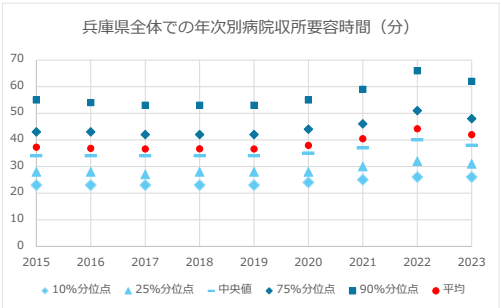


図 2 兵庫県での現場到着所要時間の年次変化　図 3 兵庫県での病院収容所要時間の年次変化

傷病者接触時刻から現地出発時刻までの現地対応所要時間（分単位），現地出発時刻から病院到着時刻までの医療機関移送所要時間（分単位），医療機関到着時刻から，医師に引き渡した時刻である病院収容時刻までの病院内移送所要時間（分単位）と分けて分析も行ったが，現場移動所要時間，病院内移送所要時間など大半の所要時間が 1 分以内のものも多数あったため，入電時刻から現地到着時刻までの分単位での現地到着時刻（以下，現場到着所要時間）と，入電時刻から収容時刻までの病院収容時刻までの分単位での病院収容所要時間（以下，病院収容所要時間）に限って分析することとした．この 2 種類の救急搬送にかかわる所要時間について，県全体と各消防本部別に記述統計量として（最小値，10%分位，25%分位，中央値，75%分位，90%分位，最大値，平均値など）を求めた．

図 4 および図 5 に示した兵庫県全体での 2023 年の 1 年間の現場到着所要時間（対数値）と病院収容所要時間（対数値）に示すように，現場到着所要時間，病院収容所要時間ともに非常に短いものから非常に長いものまで存在しており，正規分位プロットによる確認結果からは，現場到着所要時間に関しては対数正規分布とはいいがたい可能性が高いものの，病院収容所要時間に関してはほぼ対数正規分布に近いものと判断された．

次に，消防本部別に救急自動車の出動回数，現場到着所要時間と病院収容所要時間を分位点およびについて比較をしてみよう．現場到着所要時間および病院収容所要時間の平均値は消防本部ごと，また年度ごとにかなり変動がある．

まず，現場到着所要時間について検討してみた結果の一部を紹介したい．2015 年から 2021 年の期間に関しては，現場到着所要時間の平均値による比較において最短平均所要時間を示した消防本部の消防本部コードは 282065 でありその平均値は 6 分 30 秒から 7 分以下であり，2022 年度及び 2023 年度における現場到着所要時間の平均値が最小であった消防本部コードは 282189 でその平均値は 6 分 55 秒から 7 分 3 秒であっ

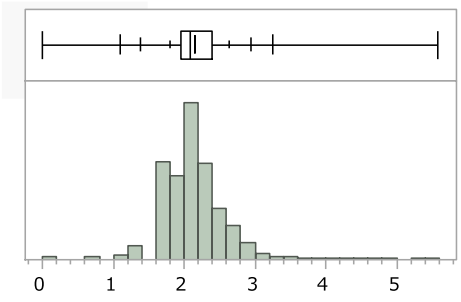


図 4　2023 年の兵庫県下の救急搬送に関する現地到着所要時間のヒストグラム

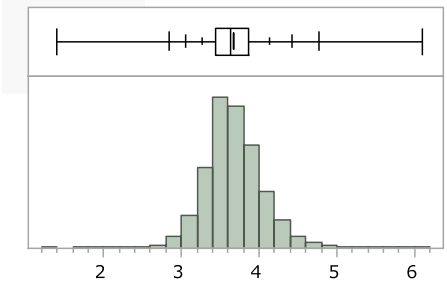


図 5　2023 年の兵庫県下の救急搬送に関する病院収容所要時間のヒストグラム

表 1　消防本部コードが 282065 での救急出動の分位点と要約統計量

分位点	最大値	31
	90%分位	10
	四分位点	8
	中央値	7
	四分位点	5
	10%分位	5
	最小値	1
要約統計量	平均	6.95
	標準偏差	2.46
	出動数	4,345
	最頻値	6

表 2　消防本部コードが 282191 での現地到着時間の分位点と要約統計量

分位点	最大値	50
	90%分位	19
	四分位点	14
	中央値	11
	四分位点	9
	10%分位	7
	最小値	0
要約統計量	平均	12.10
	標準偏差	5.66
	出動数	1,484
	最頻値	9

た．2015 年から 2023 年までの現場到着所要時間の平均値において最大であった消防本部の消防本部コードは 289281 であり，同消防本部における 2015 年以降での現場到着所要時間の平均値の最大値は 12 分 27 秒（2023 年）であった．このように，現場到着所要時間の平均値が最小・最大となる消防本部は比較的安定していることから，消防本部が置かれた環境（例えば，管轄面積や人口分布，高齢化状況）により，現場到着所要時間は影響を受けているものと考えられる．

次に病院収容所要時間の平均値についての一部の結果を紹介する．2015，2016，2018，2019，2021 の各年間の病院収容所要時間の平均値が兵庫県下で最小値を示した消防本部コードは 282162 であり，2017 年に平均値が最小値をとったのは 282090 であった．逆に 2015 年から 2023 年の各年次で，病院収容所要時間の平均値が最大値であった消防本部コードは 289281 であった．そこで，現地到着所要時間の平均値が 2015 年以降 2021 年度まで県内最小であった消防本部コード 282065 と，現地到着所要時間の平均値が 2015 年以降県内最大であった消防本部コード 282191 の 2 消防本部を取り，この 2 消防本部についての 2021 年度の現地到着所要時間のヒストグラムを図 4 および図 5 に示し，その要約統計量を表 1 および表 2 に示す．

この 2 消防本部についての結果から，消防本部ごとに現地到着所要時間および病院収容所要時間の平均値や四分位の結果からは，これらの指標値そのものに，消防本部間でかなりのばらつきがあることが想定される．この結果を踏まえ，データが利用できる期間における直近年となる 2023 年の兵庫県の各消防本部別の現地到着所要時間の分位点およびについて平均値に関してグラフとして示したものを図 6 に，同病院収容所要時間の分位点および平均値に関してグラフに示したものを図 7 として示す．図 6 からは，現場到着所要時間の 4 分位範囲や 10%分位範囲についてもかなりの差があるということが直感的に理解される．10%分位範囲および 4 分位範囲で最小値である 5 分を示した消防本部の消防本部コードは 282065 及び 282073 であり，10%分位範囲の値は 5 分であり，同じく 10%分位範囲で最大値を示した消防本部コードは 289281，10%分位範囲その値 14 分であった．2023 年における兵庫県内の現場到着所要時間の 10%分位範囲の最小値と最大値の差は 9 分であった．

次に，病院収容所要時間についての図 8 からは，現場到着所要時間よりも病院収容所要時間かなりばらつきがあることが分かる．10%分位範囲，4 分位範囲による比較では，これらの値が最小であった消防本部コードは 282065 であり．10%区間で 29 分，4 分位範囲で 14 分であった．次にこれらの値が最大値であった消防本部コードは，10%分位の最大値を取った消防本部の消防本部コードは 282219 で 50 分，4 分位範囲が最大値であった消防本部の消防コードは 289281 であり，その値は 26 分であった．

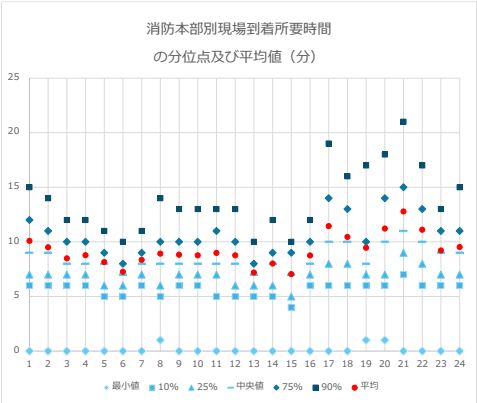


図 6 消防本部別の現場到着所要時間の分位点及び平均値(2023 年)

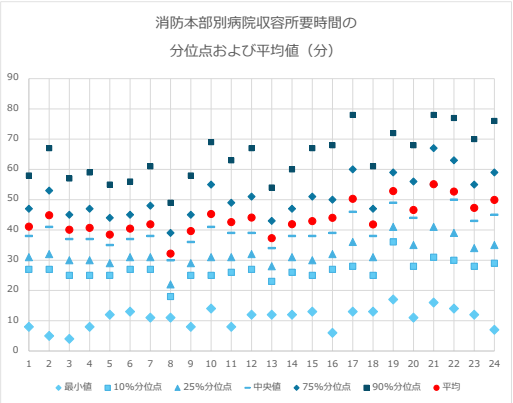


図 7 消防本部別の病院収容所要時間の分位点及び平均値(2023 年)

２救急搬送にかかわる所要時間のばらつきに影響する要因の分析

2.1 救急搬送にかかわる所要時間のばらつきに影響する可能性のある諸要因

1.3 において紹介した結果からは，現地到着所要時間や病院収容所要時間については消防本部間でかなりのばらつきが確認される.救急搬送時間にかかわる所要時間の長短も問題ではあるが，救急搬送にかかわる所要時間のばらつきも検討する必要があるのではないだろうか．この救急搬送にかかわる所要時間のばらつきは，救急自動車の配備状況や救急告示病院などの資源の地域内での偏在，救急通報が行われた現場と救急告示病院の関係，道路環境，また，降雪等の気象条件による通行困難などにより生じると考えられる．とりわけ病院収容所要時間の変動の大小は，救急活動の成否に大きく影響するものと思われる．そこで，これらの所要時間の変動が，どのような要因によって影響されているのかについて分析することにした．

緊急搬送にかかわる所要時間のばらつきに影響を与える諸要因に関する分析を実現するためには，まず，消防本部コードと消防本部の名称との対応関係を特定する必要がある．消防本部コードがどの消防本部に該当するのかを特定するためにまず，全国公共団体コード仕様及び一部事務組合等コードを参考に国土数値情報ダウンロードサイトから入手可能な消防署データの一部として提供されている消防署名の先頭部分に付与されている消防本部名を抽出した．この消防本部名を基準に，各消防本部のサイトなどで提供されている管轄域情報などを参考に消防本部コードと消防本部名との対応表を作成したうえで．消防署の管轄領域から消防本部単位の管轄領域データに再編した．

国土数値情報ダウンロードサイトで提供されている重要物流道路や医療機関情報をもとに，各消防本部別の重要物流道路長や救急告示病院ベッド数，公的病院ベッド数などを集計し，さらに総務省統計局がオープンデータとして提供している 5 次メッシュ集計による消防本部管轄内の属性別人口・世帯数および人口が存在する 5 次メッシュ面積などを面積按分法で求めた．また，各消防本部に配備されている救急自動車数については，全国消防士長会による令和 5 年版消防現勢からデータを取得した．なお，消防現勢や国土数値情報ダウンロードサイトから入手した消防署データから類推した消防本部名と消防本部コードの一致性は厳密には保証されない．このため，各消防本部が公開している統計書や各市町が公開している統計書等を確認し，90％弱（21/24 消防本部）の消防本部については出動回数が一致したため，消防本部コードと消防本部名を対応させた．

これらの独自に整備した消防本部別の指標値を独立変数，2023 年度における現地到着所要時間，病院収容所要時間を従属変数とする重回帰分析を行った．

各消防本部の管轄内における各種の属性別一般世帯数と複数の属性別の人口総数などの独立変数間に強い相関関係が確認されたため，本稿では，分析に利用する説明変数として，消防本部の非居住地を含む管轄面積（㎢）の対数値，救急自動車台数，救急自動車出動回数（回），管轄区域内の緊急輸送道路長（㎞），高齢者のみで構成される世帯数（単身高齢世帯と高齢夫婦のみ世帯の和，千世帯），消防本部管轄域内において人口の存在が確認された 5 次メッシュ面積（㎢），救急告示病院のベッド数（床）を利用し分析を行った．表３に分析に用いた指標に関する記述統計量を示す．

2.2 救急搬送にかかわる所要時間のばらつきに関するモデル化と分析結果

分析に利用した独立変数についての符号条件についてであるが，消防本部ごとの現地到着所要時間の 10％分位範囲の対数値に対する回帰分析で正のパラメータを取ると考えられる独立変数としては，消防本部の管轄区域内の非居住地を含む面積（㎢）の対数値，救急自動車出動回数（回），高齢者のみで構成される世帯数（千世帯），混雑の要因となる消防本部管轄域内において人口の存在が確認された 5 次メッシュ面積

表 3　分析に利用した各種指標に関する記述統計量

	最大値	90.00%	四分位点	中央値	四分位点	10.00%	最小値	平均	標準偏差
現場到着所要時間の10パーセンタイル区間（分）	14	12.5	9.75	7.5	6	5	5	8.13	2.58
病院収容所要時間の10パーセンタイル区間（分）	50	47	42	36.5	32.25	30.5	29	37.83	6.20
消防本部管轄面積（?）の対数値	7.16	6.74	6.41	5.28	4.11	5.28	2.91	5.25	1.25
救急自動車数（台）	34	17.5	10	5.5	4	4	3	8.29	6.84
出動回数（回）	83231	28134.50	14512.50	5699	3206.75	2012	1499	12212.67	17240.13
管轄域内の緊急物流道路長(km)	557.49	319.26	162.08	94.75	30.64	18.03	12.03	123.39	127.86
高齢者のみで構成される世帯数（千世帯）	188.86	54.64	32.80	13.06	6.83	3.88	3.39	25.93	38.22
人口存在メッシュ面積（?）	271.83	256.75	113.92	66.31	40.78	21.62	11.73	96.93	83.50
救急告示病院ベッド数（床）	18506	6054.50	3379.50	1905.50	749.25	352.50	232	2753.79	3749.32

（㎢）を取った．逆に，現地到着所要時間の 10%分位範囲の対数値に対する回帰分析で負のパラメータを取ると考えられる独立変数としては，救急自動車台数，緊急輸送道路長（㎞）を想定した．また，病院到着所要時間の 10%分位範囲の対数値に対する回帰分析において正のパラメータを取る独立変数としては，人口の存在による道路上での混雑を反映するよう，消防本部の管轄域内において人口の存在が確認された 5 次メッシュ面積（㎢），同じく負のパラメータを取る独立変数として消防本部管轄区域内の緊急輸送道路長（㎞），救急告示病院ベッド数（床）を取った．2023 年度の現地到着所要時間と病院収容所要時間の 10%分位範囲を従属変数とする対数線形回帰分析を行った結果について，現地到着所要時間に関する分析結果を表 4，病院収容所要時間に関する分析結果を表 5 に示す．

2.3 救急搬送にかかわる所要時間のばらつきに関する分析結果の解釈

表 4 に示す現地到着所要時間の 10 分位区間の対数を従属変数とした回帰分析の結果からは，現地到着所要時間に関するモデルとしての説明力は，かなり高いとはいうものの，消防本部の管轄区域内の面積（㎢）の対数値，救急自動車出動回数（回），高齢者のみで構成される世帯数（千世帯），救急車台数について符号条件は一致しているものの，一部 t 値が相当小さいパラメータがみられ，さらに，緊急輸送道路長（㎞）と消防本部管轄域内において人口の存在が確認された 5 次メッシュ面積（㎢）に関しては， t 値は小さく，符号条件は当初の想定とは逆の結果となった．VIF 値を検討してみると，変数間の多重共線性による影響がかなり発生していることが想定された（表 4 左）．そこで，分析に用いる変数を必要最小限にまで削減したモデル（表 4 右）を構築し，分析したところ，おおむね当初の想定通り，広域の管轄域を担当している消防本部，及び単身高齢世帯と高齢夫婦のみ世帯の和で計算される高齢者のみで構成される世帯数が多い消防本部では救急車の現地到着所要時間は延伸し， 救急自動車の配備台数が多い消防本部では，現地到着所要時間が短縮される傾向にあることが明らかになった．

表 5 に示した病院収容所要時間の 10%分位範囲の対数値を従属変数とした分析結果からは，現地到着時間ほどのモデルとしての説明力は高くはないものの，ある程度の説明力を持っている．しかしながら，当初予想した緊急輸送道路長のパラメータの符号条件とは異なり，パラメータは正の値をとった．VIF 値をもとに変数間の多重共線性を検討したところ，人口が多い領域に緊急輸送道路が多数整備されていることもあり，緊急輸送道路長と人口存在メッシュ面積の間に強い正の相関が確認された．そこで，緊急輸送道路長を独立変数から除き，人口存在メッシュ面積と救急告示病院のベッド数だけを独立変数とする回帰分析を実施したところ，それぞれの独立変数について当初想定したパラメータの符号条件と一致していた．つまり，消防本部管内において人口が存在するメッシュ面積が大きな消防本部では，病院収容所要時間は延伸傾向にあり，管内の救急告示病院のベッド数の多い消防本部では，逆に病院収容所要時間は短縮傾向にあることが分析結果から明らかになった．

表 4 現地到着所要時間の 10%分位範囲を従属変数とする対数線形回帰分析結果

自由度調整R2乗	0.596
誤差の標準偏差(RMSE)	0.194
現地到着所要時間の10%分位区間の対数値の平均	2.050
N	24

分散分析

要因	自由度	平方和	平均平方	F値	
モデル	6	1.497	0.250	6.66	
誤差	17	0.637	0.037	有意水準	
全体(修正済み)	23	2.134		***	

パラメータ推定値

項	推定値	標準誤差	t値	有意水準	標準β	VIF
定数項	1.190	0.366	3.260	***	0	.
消防署管轄面積（㎢）の対数値	0.209	0.079	2.640	**	0.858	6.045
出動回数	4.22.E-06	0.000	0.130		0.239	188.509
救急自動車数（台）	-0.071	0.033	-2.150	**	-1.595	31.310
高齢者のみで構成される世帯数（千世帯）	0.005	0.014	0.340		0.602	181.289
緊急輸送路長(km)	0.002	0.002	1.040		0.689	24.801
人口存在メッシュ面積（㎢）	-2.61.E-04	0.001	-0.200		-0.072	7.553

自由度調整R2乗	0.629
誤差の標準偏差(RMSE)	0.186
現地到着所要時間の10%分位区間の対数値の平均	2.050
N	24

分散分析

要因	自由度	平方和	平均平方	F値	
モデル	3	1.446	0.482	14.000	
誤差	20	0.688	0.034	有意水準	
全体(修正済み)	23	2.134		***	

パラメータ推定値

項	推定値	標準誤差	t値	有意水準	標準β	VIF
定数項	0.876	0.189	4.63	***	0	.
消防署管轄面積（㎢）の対数値	0.273	0.045	6.11	***	1.123	2.092
救急自動車数（台）	-0.054	0.018	-2.93	***	-1.215	10.689
高齢者のみで構成される世帯数（千世帯）	0.007	0.003	2.41	**	0.921	9.011

表 5 病院収容所要時間の 10%分位範囲を従属変数とする対数線形回帰分析結果

自由度調整R2乗	0.313
誤差の標準偏差(RMSE)	0.134
病院収容所要時間の10%分位区間の対数値の平均	3.620
N	24

分散分析

要因	自由度	平方和	平均平方	F値	
モデル	3	0.244	0.081	4.498	
誤差	20	0.362	0.018	有意水準	
全体(修正済み)	23	0.605		**	

パラメータ推定値

項	推定値	標準誤差	t値	有意水準	標準β	VIF
定数項	3.597	0.043	82.86	***		
人口存在メッシュ面積（㎢）	1.28.E-04	0.001	0.17		0.066	5.259
緊急輸送路長(km)	0.001	0.001	1.53		0.750	8.020
救急告示病院ベッド数	-3.88.E-05	1.E-05	-3.31	***	-0.895	2.452

自由度調整R2乗	0.269
誤差の標準偏差(RMSE)	0.139
病院収容所要時間の10%分位区間の対数値の平均	3.620
オブザベーション(または重みの合計)	24

分散分析

要因	自由度	平方和	平均平方	F値	
モデル	2	0.201	0.101	5.238	
誤差	21	0.404	0.019	有意水準	
全体(修正済み)	23	0.605		***	

パラメータ推定値

項	推定値	標準誤差	t値	有意水準	標準β	VIF
定数項	3.585	0.04	81.40	***		
人口存在メッシュ面積（㎢）	0.001	0.00	2.74	***	0.584	1.431
救急告示病院ベッド数	-2.72.E-05	0.00	-2.95	***	-0.628	1.431

2.4 救急搬送にかかわる所要時間のモデルの結果から大きく乖離している事例の検討

2023 年度の現地到着所要時間と病院収容所要時間の 10%分位点の対数を取った値に関する予測モデルによる推計値と現実の値とを比較してみたい．回帰モデルにおける予測値と実際の値とが大きく外れている消防本部のみについてラベルを付したグラフを図 9 から図 12 に示す．図 9 及び図 10 からは，消防本部の管轄域が隣接している猪名川町と三田市の両消防本部については，現地到着所要時間の 10%分位範囲の予測値と実際の値の間に大きな差がある，より具体的には，猪名川町消防本部は，モデルの予測値と比べ，実際の値はかなり大きく現場到着所要時間のばらつきが大きく，逆に三田市消防本部では，モデルの予測値と比べ，実際の値はかなり小さく，三田市消防本部における現場到着所要時間のばらつきはかなり小さいという結果となった．このような結果となった背景として，三田市消防本部が管轄する領域内に 1980 年代に開発された北摂三田ニュータウンの大規模開発に伴う高規格道路による道路ネットワークが整備されていることが指摘できよう，これに対し，猪名川町消防本部の管轄領域は，猪名川流域の南北方向に走る対面 2 車線の県道 12 号線のみが主要幹線として存在し，その沿線に住民が居住する領域が広がっているという特性があ

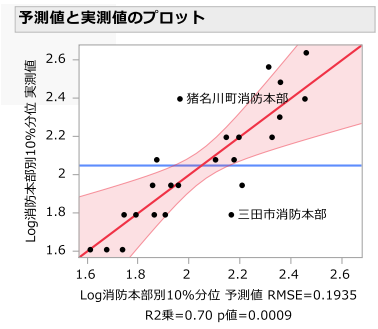


図 9 現場到着所要時間の 10%分位範囲に関する予測値と実測値（表 4 左側）

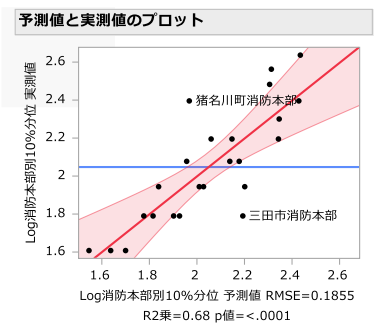


図 10 現場到着所要時間の 10%分位範囲に関する予測値と実測値（表 4 右側）

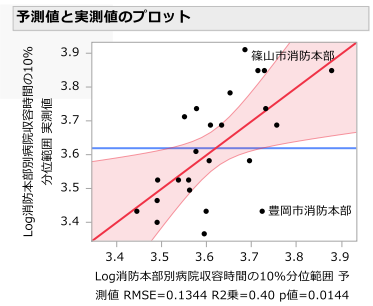


図 11　病院収容所要時間の 10%分位範囲に関する予測値と実測値（表 5 左側）

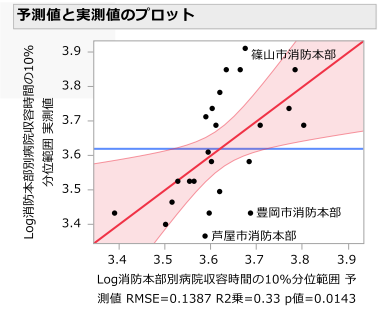


図 12　病院収容所要時間の 10%分位範囲に関する予測値と実測値（表 5 右側）

る．道路ネットワークの特性が隣接するこれらの消防本部間で大きく異なることが，モデルの予測と実際の値の差に反映されているものと考えられる．

次に，病院収容所要時間の 10%分位範囲の対数値に関する回帰モデルによる予測値と現実との値の比較をしてみたい．豊岡市消防本部の病院収容所要時間はモデルによる推計値よりも大幅に短い．兵庫県全体の病院収容所要時間との比較や，現地到着時刻と現地出発時刻の差，現地出発時刻から病院収容時刻の差で定義される時間について検討した結果からも，兵庫県のこれらの時間に関する全県の平均値や中央値と比べ，豊岡市消防本部におけるこれらの時間は，明らかに短い．現地到着所要時間については，全県平均や中央値と比べ豊岡市消防本部の同統計量は 1 分程度大きいが，現地に到着してからの所要時間が極めて短いという特徴がみられる．このような結果の背景については，空間的要素では説明できないため，何らかの別要素が影響しているものと考えられる．これに対し，篠山市消防本部（現丹波篠山市消防本部）については，現地到着所要時間の中央値で県の中央値と比べ 2 分，同平均値の比較において 2.2 分の延伸がみられた．さらに現地出発所要時間についての比較結果からは，全県の中央値，平均値と比べ，篠山市消防本部においては 9 分弱の延伸が確認されている．この背景として，篠山市消防本部管内での救急告示病院の数そのものの少なさと管轄域の広さの影響を指摘できよう．

これらの結果を踏まえ，救急告示病院へのアクセシビリティや主要幹線道路ネットワークのフラクタル構造や，そのほかの要因についても何らかの定量的な指標化を行い，今後分析を進めていくことを予定している．

3 小括と今後の課題

本稿では，救急搬送にかかわる個票データをもとに消防本部別の分位区間などを算出し，現場到着所要時間および病院収容所要時間の 10%分位範囲というばらつきを示す指標値の対数値を従属変数とする回帰分析を実施した結果，空間的要素や地域における救急搬送にかかわる資源の状況や空間分布によりこれらの指標値が影響を受けている可能性を示した．しかし，本来個票データが存在していることから，分位点回帰等の分析手法が利用可能であると考えるが，残念ながら時間的な制約等からその実施にまで至っていない．

今後は，分位点回帰や一般化回帰などの分析手法の適用を検討している．また，本研究で明らかになったように，道路構造や病院（救急告示病院とそのベッド数）の空間分布など，空間における要素に関するより現実的な定式化に基づく研究や，現在は事故（傷病）の種別，疾患の種別や年齢層，重症度，傷病の発生場所の種別，発生月や気象条件など，救急搬送に影響を与える諸条件についての検討を行っていないが，これらの諸要因を反映する形で，より詳細に分析を進めることを検討している．

なお，本稿は，CSIS Days 2025 で発表予定の川向・野田(2025)の内容に新規の分析結果と内容を加え，大幅に加筆修正したものである．

謝辞

貴重な救急搬送人員データを研究目的で提供いただいた総務省消防庁救急企画室にこの場を借りて謝意を表したい．

参考文献

総務省消防庁（2024）『令和 6 年版消防白書』第 2 章第 5 節，76-87.
川向　肇・野田　晴菜(2025)　兵庫県下における救急搬送データを利用した現地到着所要時間に関する分析．「CSIS Days 2025 研究アブストラクト集」. Forthcoming.
鈴木 万生・平田 輝満（2024），救急搬送人員データを活用した搬送時間分布の分析手法に関する研究，「土木学会論文集」, 80（20）論文 ID: 24-20062
全国消防長会（2024）『令和 5 年版消防現勢』（消防本部別），http://www.fcaj.gr.jp/info/download/%E4%BB%A4%E5%92%8C%EF%BC%95%E5%B9%B4%E7%89%88%E6%B6%88%E9%98%B2%E7%8F%BE%E5%8B%A2.pdf
山岡 淳（2015）救急搬送における地域の諸条件と搬送時間の関連性．「経済社会学会年報,」 37, 94-103.

SASによる症例数再推定を伴うアダプティブデザインの実装と統計的評価

～CPZ（Constrained Promising Zone）デザインの実践～

○内藤　寿輝，森田　祐介，太田　裕二，浜田　泉

（ノーベルファーマ株式会社　データサイエンス部）

Implementation and Statistical Evaluation of Adaptive Design
with Sample Size Re-estimation using SAS
- Practical Application of the Constrained Promising Zone Design -

Toshiki Naito，Yusuke Morita，Yuji Ohta，Izumi Hamada
Data Science Department，　Nobelpharma Co.，　Ltd.

要旨

本稿では，Constrained Promising Zone（CPZ）デザインに基づく症例数再推定を取り入れたアダプティブデザインの実装例と，その統計的評価について報告する．中間解析時の治療効果に応じて追加症例数を柔軟に調整可能とする本手法は，近年注目を集めているが，国内での具体的な適用例や解析手順の紹介は限られている．本稿では，CPZ デザインを用いた症例数再推定の考え方を整理し，実際の臨床試験を想定したシミュレーション結果を提示する．

また，CPZ デザインの仕組みを視覚的に理解しやすくするために，SAS を用いて作成したグラフの実例も併せて示す．

キーワード : アダプティブデザイン，中間解析，症例数再推定，CPZ デザイン，条件付き検出力，SGPLOT，ICH E20，rpact

1．はじめに

アダプティブデザインは，臨床試験に参加した被験者の蓄積されたデータに基づいて，試験デザインの一部について予め計画された変更を行うことができる柔軟な手法であり，試験の効率性向上や成功確率の増加，さらには被験者がより良い治療を受けられる機会の提供など，多くの利点が期待されている．この手法は，欧州医薬品庁（EMA）および米国食品医薬品局（FDA）によりすでにガイダンスが発行されており，近年では革新的試験デザイン（CID）の一例としても言及されている[1] [2] [3]．また国際的にも，アダプティブ臨床試験は ICH E20 のトピックとして採択され，ガイドライン案が作成されており（2025 年 6 月時点で Step 2），本邦でも近くパブリックコメントの募集が始まる見込みである[4].

アダプティブデザインの一種に症例数再推定がある．これは中間解析の結果に基づき最終症例数を調整する方法であり，Bothwell らは，文献と ClinicalTrials.gov のアダプティブデザイン 142 件を集計し，seamless II/III 57%，群逐次デザイン 21%，症例数再推定 8%であったと報告している[5]．さらに，FDA ガイダンス

と ICH E20（Step 2）でも，症例数再推定はアダプティブデザインの一類型として明示的に取り上げられている[2][4].

本稿では，Constrained Promising Zone（CPZ）デザインに着目する[6]．このデザインは，中間解析で条件付き検出力（Conditional Power， CP）を評価し，その結果に基づいて必要に応じて症例数を再設定するアダプティブデザインである．臨床的に意味のある最小の治療効果（δ_{min}）を仮定し，症例数の追加により一定以上の条件付き検出力の確保が可能な場合のみ症例数の調整を行う．症例数再推定を行っても，事前に設定した第一種の過誤確率（Type I error rate）を保持できるように設計されており，かつ中間結果に応じて症例数を柔軟に追加できる点が特徴である．

アダプティブデザインは柔軟で魅力的な手法である反面，統計的な課題への適切な対応が不可欠である．具体的には，第一種の過誤確率の増大，推定値のバイアス，信頼区間の被覆率の名目水準との乖離といった問題が生じ得るため，計画段階で十分なシミュレーションを行い，動作特性を事前に把握しておくことが重要である[2] [4]．この点，R のパッケージ「rpact」では CPZ デザインのシミュレーションや可視化機能が提供されているが，SAS による実装例は限られているのが現状である[7]．そこで本稿では，CPZ デザインの実装に伴う課題とその対応策を整理するとともに，SAS による CPZ デザインのシミュレーション手法および可視化手法を紹介する．

2. 想定する試験デザイン

本稿では，主要評価項目として全奏効率（ORR）を用いた，Active 群と Placebo 群を比較する無作為化優越性試験を想定する．なお，割付比は 1：1 の均等割付とする．本デザインは非盲検下の中間解析を含む 2 段階デザインとして計画される．第 1 ステージを中間解析まで，第 2 ステージを中間解析後から最終解析までと定義する（図 1）．

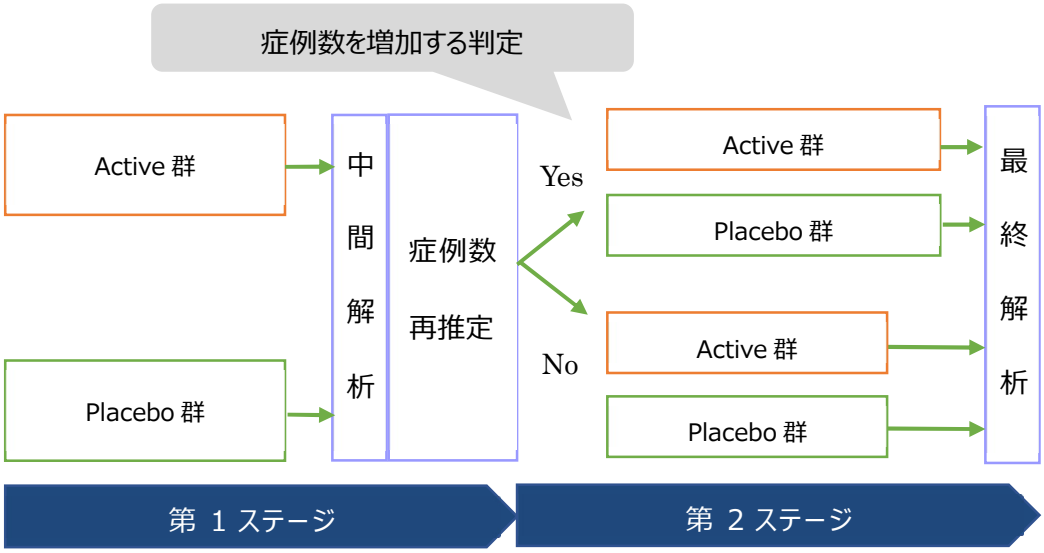


図 1．中間解析を含む，2 段階デザインの症例数再推定を伴う無作為化優越性試験

本デザインは，中間解析結果に基づき，有効性が認められない場合に無効中止（早期中止）が可能であること，中間解析で条件付き検出力が「有望領域（Promising zone）」にある場合には，症例数を追加して検

出力を高められること，といった柔軟性を備えている点で魅力的である．なお，本稿では有効中止は考慮しない．

一方で，症例数再推定を伴う 2 段階デザインでは，

(1) 中間解析と最終解析で繰り返し有効性評価を行うこと

(2) 中間解析に基づいて症例数を追加すること

の二点から，第一種の過誤確率が上昇する可能性がある[8].

したがって，3 章で述べるように，本デザインにおいては第一種の過誤確率を適切に制御可能な統計手法を適用することが必須である．

3．症例数再推定における第一種の過誤確率制御の一般的な手法

本章では，試験の途中での計画的な変更（アダプテーション）が，最終的な結論の解釈に影響を与え，第一種の過誤確率の増大を引き起こす可能性があるという課題に対処するための，一般的な統計的手法について概説する[8].

1. 統合検定による方法（Lehmacher and Wassmer （1999）[9]）

   この手法は，各ステージで得られたデータから算出される検定統計量を，以下に定義された統計量Z_{LW}で統合することにより，帰無仮説下で第一種の過誤確率を目標水準以下に制御する．

   $$Z_{LW} = \omega_1 \Phi^{-1}(1 - p_1) + \omega_2 \Phi^{-1}(1 - p_2)$$

   ここで，Φ^{-1}（・）は標準正規分布の分布関数の逆関数，ステージ i（i=1,2）の p 値と重みはp_iおよびω_iである．

   また重みω_1，ω_2は事前に規定され，ω_1^2+ω_2^2=1 である．この方法を用いた最終解析の p 値は$p_{LW} = 1 - \Phi(Z_{LW})$で得られる．

   なお，この方法以外にも，Bauer and Kohne （1994）[10]による Fisher の基準に基づく各ステージの p 値の積を用いて p 値を統合し，棄却域を導出する方法，また Chang （2006）[11]による各ステージの p 値の和を用いて検定統計量を構築し，有効・無益性中止の境界，調整済みp値を計算するための閉形式を導出する方法などがある．

2. 条件付き過誤関数（Proschan and Hunsberger （1995）[12]）

   この手法は，第 1 ステージで得られた検定統計量あるいは p 値に基づいて，最終解析の棄却境界を調整することで，試験全体の第一種の過誤確率を制御する．$A(p_1)$をp_1のもとで，第 1 ステージで試験継続が判断された場合の第 2 ステージの第一種の過誤確率の条件付き確率とすると，試験全体の第一種の過誤確率αは以下のようになる．

   $$\alpha = \alpha_1 + \int_{\alpha_1}^{\alpha_0} A(p_1) dp_1$$

   ここで，α_1は第 1 ステージのデータに基づく，有効と判断するための閾値，α_0は第 1 ステージのデータに基づく，無効と判断するための閾値である．

上記の式は，試験全体の第一種の過誤確率αに抑えられるように$A(p_1)$を設定していることになり，$A(p_1)$に基づく調整した最終解析の棄却域を用いる事で試験全体の第一種の過誤確率を制御する．

Proschan and Hunsberger （1995）は以下の関数を提案しており，

$$A(p_1) = 1 - \Phi\left(\sqrt{[\Phi^{-1}(1-\alpha_1)]^2 - [\Phi^{-1}(1-p_1)]^2}\right), \quad \alpha_1 \leqq p_1 \leqq \alpha_0$$

これを用いることによって，最終解析の棄却域は以下のようになる．

$$c_{CEF} = \frac{\sqrt{n_1}Z_1 + \sqrt{\hat{n}_2}\Phi^{-1}(1-A(p_1))}{\sqrt{n_1 + \hat{n}_2}}$$

ここで，第 1 ステージで得られた検定統計量をZ_1，症例数再推定後の第 2 ステージにおける 1 群あたりの症例数を$\hat{n}_2$，試験計画時の第 1 ステージにおける 1 群あたりの症例数をn_1とする．

4. Constrained Promising Zone（CPZ）デザイン（Hsiao et al.（2019））

本章では，CPZ デザインの概要を示す．CPZ デザインは，中間解析の情報を活用して限定的に症例数を調整し，第一種の過誤確率を制御しながら試験の実行性と効率を向上させるデザインである．

4.1. CPZ デザインの概要

アダプティブデザインの一つである症例数再推定は，中間解析の結果に応じて症例数を調整することで試験の効率性と成功確率を高める手法である．その中で CPZ デザインは，Hsiao ら（2019）により提案された方法である[6].

CPZ デザインの特徴は，中間解析時に「有望領域（Promising Zone）」を設定し，この範囲に入った場合のみ症例数を追加する点にある．具体的には，以下で定義される中間解析で得られた検定統計量（Z 値）がz_1^Lとz_1^Uの間にある場合に症例数が追加される．

- z_1^L：この値を上回る場合は症例数が追加される．一方，この値を下回る場合は症例数を最大限追加しても条件付き検出力が不十分であり症例数は増えない.
- z_1^C：この付近では症例数を最大限追加した場合に条件付き検出力が上限基準（CP_{max}）に達することを示す.
- z_1^U：この値を上回る場合は症例数を追加しなくても十分な検出力が確保できているため症例数は増えない.

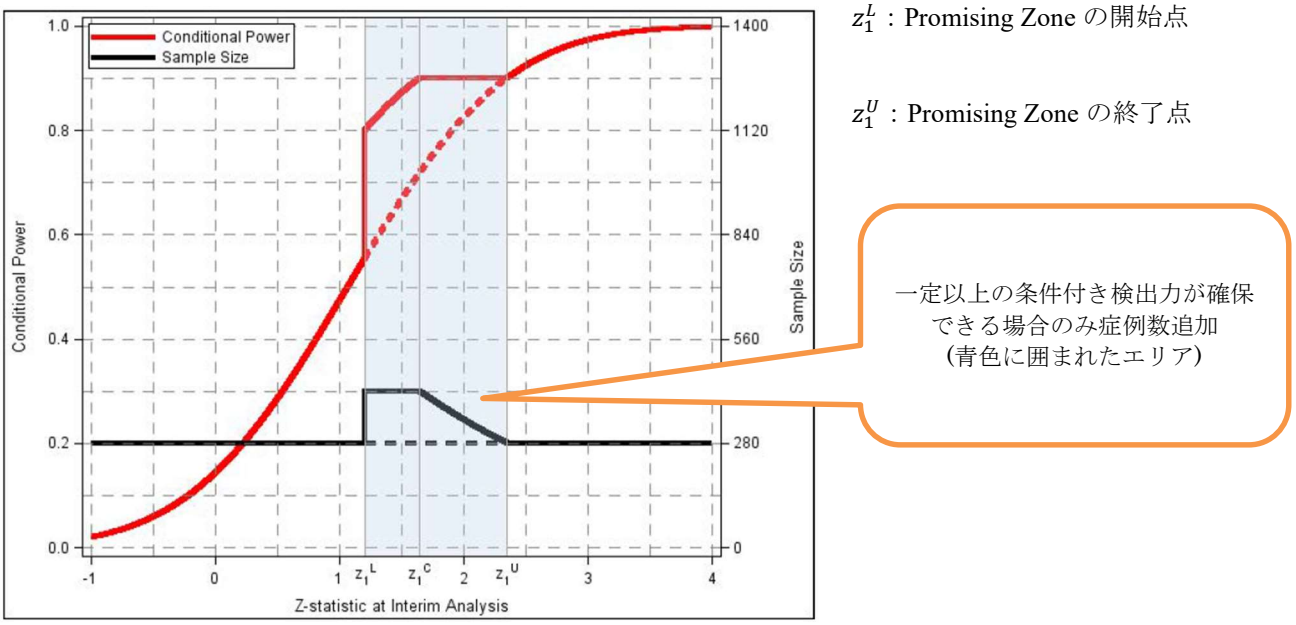


図 2．条件付き検出力，Z 統計量およびサンプルサイズの関係図

Source: Figure 2 from Zhao Yang, Shivani Nanda. Promising Zone Designs for Sample Size Re-estimation in Clinical Trials: Graphical Approaches Using SAS [7].

4.2．Five-zone アダプティブデザイン

Five-zone アダプティブデザインとは，Liu （2021）[13] によって提唱された考え方であり，非盲検下における中間解析の結果に応じて，試験の継続方針や症例数の再設定を柔軟に行う設計戦略である．この「Five-zone」という名称は，中間解析の結果に基づいて試験の方針を導く 5 つの判定領域（ゾーン）を意味している[8].

各ゾーンは，統計的有意性，条件付き検出力，および事前に設定された閾値に基づいて分類され，以下のように定義される．

1. **Efficacy zone（有効領域）**
   中間解析時点で有効性の判定基準を満たした場合，試験はその時点で有効と判断され，早期終了となる．
2. **Futility zone（無益領域）**
   効果が非常に小さい，あるいは臨床的に意味がないと判断される場合，無益性中止の基準を満たしているとみなされ，試験は早期に中止される．
3. **Favorable zone（良好領域）**
   中間解析の結果が十分に良好であり，事前に設定した目標条件付き検出力を上回っている場合には，試験は計画された症例数でそのまま継続される．
4. **Promising zone（有望領域）**
   条件付き検出力が目標値には到達していないが，結果としては有望と判断される範囲にある場合，試験の成功確率を高めるために，最大症例数の上限を設けた上で症例数を増加させる適応的な対応が行われる．

5. **Unfavorable zone（不利領域）**

条件付き検出力が目標値にも達せず，有望とも無益とも明確に判断できない場合，追加の症例数は導入せず，試験は当初計画された症例数のまま継続される．

このように，Five-zone アダプティブデザインでは，中間解析の結果に応じて，試験の早期終了，計画通りの継続，または症例数の増加による適応のいずれかの選択が合理的に行われるよう設計されている．このアプローチは，必要以上の症例数の投入を避けつつ，試験成功の可能性を最大化することを目的としている．

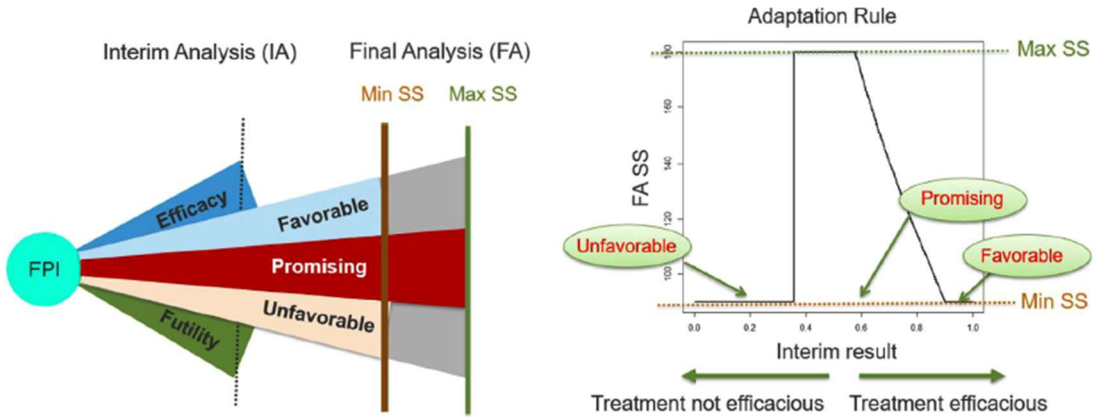


図 3．Fize zone の概念図と症例数再推定のルール図

SSR design concept (on the left) where interim result is divided into five zones: starting from first patient in (FPI), the trial stops at IA in efficacy or futility zone, minimum planned sample size (Min SS) will be used for FA in favorable or unfavorable zone, sample size will be between minimum and maximum planned sample size (Max SS) in promising zone; An example of sample size adaptation rule where FA sample size (SS) is a function of the interim result (on the right).
Source: Fig. 1. from Liu et al., Sample size re-estimation for pivotal clinical trials, Contemporary Clinical Trials 2021[13].

4.3．条件付き検出力

評価項目が連続変数の場合，中間解析の検定統計量が与えられた下で，最終解析において帰無仮説が棄却される確率，すなわち条件付き検出力（Conditional Power，　CP）を以下の式で定義される[8].

$$\mathrm{CP}\ (\delta,\ z_1,\ n_2)\ = \Phi\ \left(\frac{\delta\sqrt{n_2}}{\sqrt{2}\sigma} - \frac{z_\alpha\sqrt{n_1+n_2}-z_1\sqrt{n_1}}{\sqrt{n_2}}\right)$$

ここで，

- δは平均値の群間差の真値
- z_1は第 1 ステージのデータに基づく標準化検定統計量
- n_1，　n_2は試験計画時の第 1，第 2 ステージにおける 1 群あたりの症例数
- z_αは標準正規分布の上側α%点
- σは標準偏差の真値

である．

上記の式に示される通り，CP は真の治療効果δの関数である．よって，未知数であるために実務では 1.症例数設計の仮定を用いたδ_{pre}，2.中間解析時点での結果に基づく推定値$\hat{\delta}_{pre}$を用いる事になる．後者の場合，条件付き検出力の算出の際に中間データを用いるため，中間データが繰り返し使用される点が問題点である[8]．CPZ デザインでは，この問題点を解決するために前者のδ_{pre}を用いている．

4.4. 最適化問題として取り扱い症例数再推定を行う方法

アダプティブデザインでは，中間解析の結果に基づいて，試験の後半部分における症例数（第 2 ステージの症例数）を再評価し，調整することが可能である．この条件付き検出力と症例数の調整を「最適化問題」として捉える方法が近年注目されており，代表的なアプローチとして Jennison and Turnbull （2015）[14] と Hsiao, Liu and Mehta （2019）[6] の提案が挙げられる．

特に，Hsiao, Liu and Mehta （2019）による方法は，Jennison and Turnbull （2015）の試験デザインと Lehmacher and Wassmer （1999） [9]の統合検定を組み合わせた目的関数と制約条件からなる以下の最適化問題として提案し，このデザインを Constrained Promising zone（CPZ）デザインと呼ぶ[6] [8].

- 目的関数: $Maximize\{CP_{\delta_{min}}（Z_1，\ \hat{N}_{final}）\}$
- 制約条件 1: $N_{initial} \leq \hat{N}_{final} \leq N_{max}$
- 制約条件 2: $CP_{\delta_{min}}(Z_1，\ \hat{N}_{final}) \geq CP_{min}$
- 制約条件 3: $CP_{\delta_{min}}(Z_1，\ \hat{N}_{final}) \leq CP_{max}$
- Z_1: Stage 1 の片側 p 値に基づく標準化検定統計量
- N_{max}: 試験全体の 1 群あたりの最大症例数
- $\hat{N}_{final}$: 症例数再推定後の試験全体の 1 群あたりの症例数
- $N_{initial}$: 試験計画時の試験全体の 1 群あたりの症例数
- δ_{min}: 臨床的に意味のある最小の治療効果
- CP_{min}: Promising zone 内の条件付き検出力が満たすべき最低要件．$\delta=\delta_{min}$，$\hat{N}_{final}=N_{max}$ としたときの条件付き検出力が CP_{min} に達していない場合，その中間解析結果は Unfavorable zone にあるとみなし，症例数は変更しない．さらに条件付き検出力が 20%未満にある場合は Futility zone（promising zone の範囲外）にあるとみなし，試験を無効中止する．
- CP_{max}: Promising zone 内の条件付き検出力の最大値．$\delta=\delta_{min}$，$\hat{N}_{final} \leq N_{max}$ としたときの条件付き検出力が CP_{min}～CP_{max} になるように症例数を追加する．なお，症例数を追加しなくても条件付き検出力が CP_{max} を超えている場合は，Favorable zone（promising zone の範囲外）にあるとみなし，症例数は変更されない．

上記の CPZ デザインをもとに再推定された$\hat{n}_2$を用いた条件付き検出力は以下のように定義される．

$$CP（\delta，\ z_1，\ \hat{n}_2） = \Phi\ (\frac{\delta\sqrt{2\hat{n}_2}}{\sqrt{2}\sigma} - \frac{z_\alpha\sqrt{n_1+n_2}-z_1\sqrt{n_1}}{\sqrt{n_2}})$$

さらに評価項目が二値変数の場合，rpact の実装（f_simulation_base_rates.cpp）を基に，観測された治療効果および中間時点での条件付き棄却限界値に応じて正規化効果量 θを定義し，臨床的に意味のある最小の治療効果で設定した治療群と対照群の割合をp_1，p_0，第 1 ステージで得られた治療群と対照群の割合を$\hat{p}_1$，$\hat{p}_0$とすると以下の式で表すことができる．

$$CP（\theta，\ z_1，\ \hat{n}_2） = \Phi\ (\theta\sqrt{2\hat{n}_2} - \frac{z_\alpha\sqrt{n_1+n_2}-z_1\sqrt{n_1}}{\sqrt{n_2}})$$

ここで,

$$\theta = \frac{1}{\sqrt{2}}\left[\frac{(p_1 - p_0)}{\sqrt{p_1(1-p_1)+p_0(1-p_0)}} + sign(p_1 - p_0) * \frac{z_\alpha\sqrt{n_1+n_2} - z_1\sqrt{n_1}}{\sqrt{n_2}} * \frac{\sqrt{2}}{\sqrt{\hat{n}_2}} * \left(1 - \frac{\sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1)+p_0(1-p_0)}}\right)\right]$$

$$z_1 = \frac{(\hat{p}_1 - \hat{p}_0)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_{p_1}} + \frac{\hat{p}(1-\hat{p})}{n_{p_0}}}}$$

$$p = \frac{(p_1 + p_0)}{2}$$

$$\hat{p} = \frac{(\hat{p}_1 + \hat{p}_0)}{2}$$

であり，治療群および対照群の中間時点での観測割合，および計画時のパラメータに基づいて構築されたものであり，標準化された効果の強さを反映する[15]．また，割付比は 1：1 を想定している．

5. シミュレーションによる検討

本章では，rpact の Simulation of a Trial with a Binary Endpoint and Unblinded Sample Size Re-Calculation with rpact の臨床試験事例を用いて，シミュレーションによる性能評価とその可視化を行う[16].

また CPZ デザインの可視化をするグラフについても説明する．

5.1. 想定する試験デザイン・シナリオ

まず，シミュレーションで想定する試験デザイン，パラメータ設定，中間解析の設定条件およびシナリオ設定を下表に示す.

シミュレーションでは，主要評価項目として全奏効率（ORR）を用いた 1:1 の無作為化優越性試験を想定している．主要な要約指標は Active 群と Placebo 群の間の全奏効率の群間差とし，各群の症例数はそれぞれ 240 例，検出力 90%，片側有意水準は 2.5%と設定した．中間解析は全体の 50%の症例が集積した時点で 1 回実施し，条件付き検出力（CP）に基づいて無効中止または症例数再推定を行う．症例数再推定は Promising Zone（CP が 80%以上 90%未満）のみで実施し，最終的な最大例数は各群 392 例とする．脱落や欠測は発生しないものとし，すべての設定について 10,000 回のシミュレーションを行う．

想定する試験デザイン／シミュレーション条件

主要評価項目	全奏効の有無
要約指標	全奏効率（ORR）の群間差
割付比率	1：1
症例数（検出力）	Active 群 240 例，　Placebo 群 240 例（90%）
有意水準	片側 2.5%
繰り返し数	10,000 回
解析手法	各ステージ（第 1 ステージ及び第 2 ステージ）別に算出した Active 群と Placebo 群の全奏効率の差を，計画時の各ステージの症例数比による加重平均により算出する．

	また，奏効率および差の標準誤差は正規近似により算出する．
評価指標	第一種の過誤確率，条件付き検出力，Bias，MSE，被覆確率など
その他	中止・脱落および欠測データは発生しないものとする 両群ともに奏効例がいない状況が発生した場合は，統計量（Z 値）を 0 として取り扱う（rpact の仕様に倣った）

パラメータ設定

N_1, N_2	試験計画時の第 1，2 ステージにおける 1 群あたりの症例数
$\widehat{N}_{final}$	症例数再推定後の試験全体の 1 群あたりの症例数
$N_{initial}$	試験計画時の試験全体の 1 群あたりの症例数
N_{max}	試験全体の 1 群あたりの最大症例数
δ_{min}	臨床的に意味のある最小の治療効果
CP_{min}	Promising zone 内の条件付き検出力が満たすべき最低要件
CP_{max}	Promising zone 内の条件付き検出力の最大値

中間解析の設定条件

中間解析の回数及び目的	1 回（無効中止），症例数再推定
中間解析の時期	全体症例の 50%の集積時
無効中止の基準	症例数再推定後の条件付き検出力（CP）が 20%未満の場合に中止
症例数再推定の方法	Unfavorable zone： 20%（無効中止基準）≦CP＜80%（CP_{min}） 症例数は変更しない Promising zone：80%≦CP＜90%（CP_{max}） 症例数を追加する（1 群あたりの最大例 392 例） Favorable zone：90%≦CP 症例数は変更しない
推定・信頼区間の算出方法	最終解析時の群間差の点推定値として，各ステージの計画時の症例配分を重みとして各ステージの群間差の点推定値の加重平均を算出する．信頼区間についても重みを考慮した Wald 型信頼区間を算出する．

シミュレーション結果の評価項目

Average of conditional power achieved	症例数再推定を行った場合の条件付き検出力の平均値
Average of conditional power achieved without SSR	症例数再推定を行わなかった場合の条件付き検出力の平均値
Overall Reject／Power／Type I Error Rate	最終解析において、帰無仮説が棄却された割合
Bias	推定された治療効果の平均値と真の治療効果との差の平均値
MSE	推定された治療効果の平均値と真の治療効果との差の二乗の平均値

Coverage Probability	推定された治療効果の 95%信頼区間が真の治療効果を含む割合
Combined (stage 1 + stage 2) bias	第 1 ステージと第 2 ステージのデータを単純に統合して推定した効果量における Bias（いわゆるナイーブ解析）
Combined (stage 1 + stage 2) mse	第 1 ステージと第 2 ステージのデータを単純に統合して推定した効果量における MSE
Combined (stage 1 + stage 2) Coverage Probability	第 1 ステージと第 2 ステージのデータを単純に統合して推定した効果量における Coverage Probability
Futility Proportion	中間解析で「無効中止」と判断され試験が終了した割合
Proportion of Sample Size Added	症例数再推定により症例追加が判断された割合
Expected total number of subjects (N)	最終症例数（両群合計）の平均値
Overall Reject (Power of Fixed Design)	固定デザイン（症例数追加なし）における検出力
N #%	最終症例数の分布における#パーセンタイル値

シナリオ設定

シナリオ	Placebo	Active	N_1	N_2	N_{max}	δ_{min}	CP_{min}	CP_{max}	説明
基本設定	0.2	0.33	120	120	392	0.1	0.8	0.9	各群 240 例となる設定
①	*	*	120	120	392	0.1	0.8	0.9	全奏効率のみを変えて比較

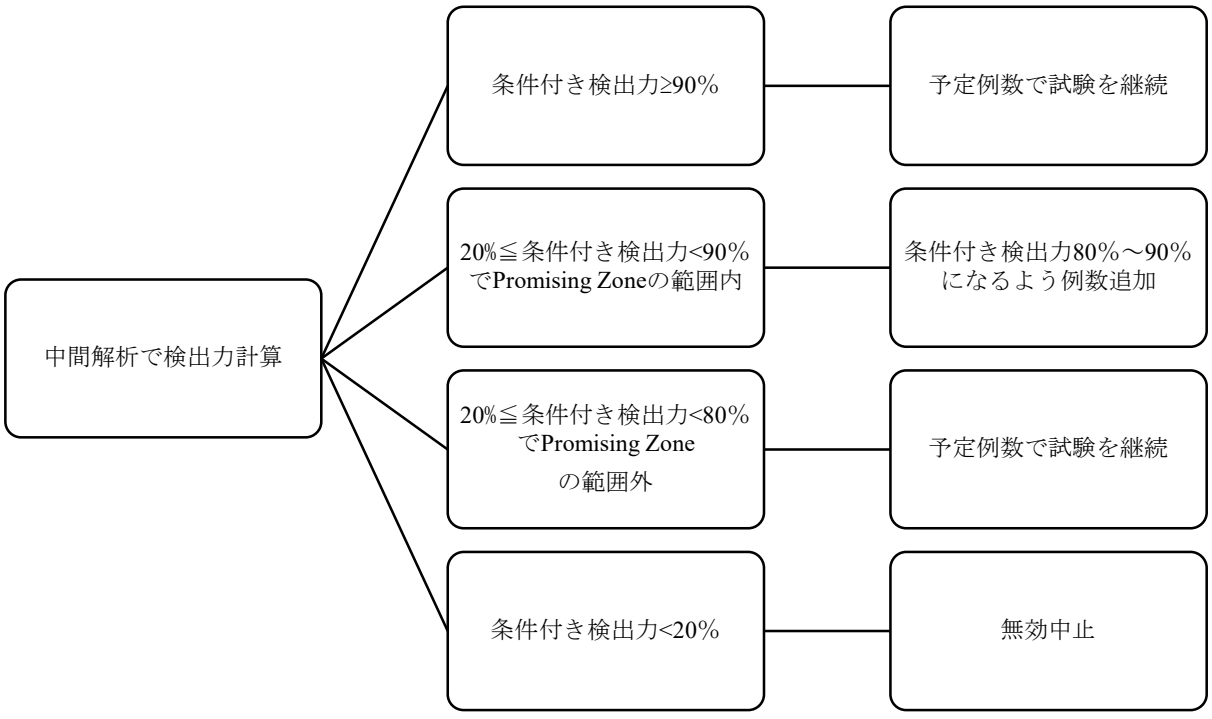


図 4. CPZ デザインのフロー

5.2. シミュレーション結果

まず，全奏効率の群間差（Active - Placebo）を 0 から 0.3 まで変化させた場合のシミュレーション結果を表 1 と表 2 に示す.

表 1. シミュレーション結果①-1

Placebo	Active	Average of conditional power achieved	Average of conditional power achieved without SSR	Overall Reject (Power)	Bias	MSE	Coverage Probability	Combined (stage 1 + stage 2) bias	Combined (stage 1 + stage 2) mse	Combined (stage 1 + stage 2) Coverage Probability
0.2	0.2	0.2801	0.2407	0.0235	-0.0106	0.0019	0.9515	-0.0125	0.0018	0.9516
0.2	0.3	0.7990	0.7203	0.7980	-0.0021	0.0018	0.9385	-0.0010	0.0019	0.9319
0.2	0.33	0.8844	0.8271	0.9378	-0.0004	0.0016	0.9444	0.0017	0.0016	0.9395
0.2	0.4	0.9697	0.9576	0.9988	-0.0008	0.0017	0.9453	0.0003	0.0016	0.9521
0.2	0.5	0.9981	0.9979	1.0000	0.0005	0.0017	0.9500	0.0005	0.0017	0.9512

シミュレーション回数：各条件 10,000 回
Maximum total number of subjects stage1: 240 stage2: 544
Bias represents the average difference between the estimated and true treatment effect, indicating systematic error in estimation.
MSE (Mean Squared Error) quantifies the overall estimation accuracy by incorporating both variance and bias squared.
Coverage Probability refers to the proportion of simulations in which the 95% confidence interval contains the true treatment effect, serving as an indicator of the reliability of the interval estimates.
Smallest clinically meaningful difference for sample size re-estimation: 0.1 (Active: 0.3 vs Placebo: 0.2)

表 2．シミュレーション結果①-2

Placebo	Active	Futility Proportion	Proportion of Sample Size Added	Expected total number of subjects (N)	Overall Reject (Power of Fixed Design)	N 5%	N 25%	N 50%	N 75%	N 80%	N 90%	N 95%
0.2	0.2	0.5358	0.1707	394.5	0.0264	240	240	240	480	480	772	784
0.2	0.3	0.0466	0.4867	559.9	0.7253	480	480	480	671	717	784	784
0.2	0.33	0.0118	0.4037	544.7	0.9086	480	480	480	595	641	763.5	784
0.2	0.4	0.0006	0.1159	494.7	0.9977	480	480	480	480	480	502	609
0.2	0.5	0.0000	0.0023	480.2	1.0000	480	480	480	480	480	480	480

シミュレーション回数：各条件 10,000 回
Maximum total number of subjects stage1: 240 stage2: 544
Smallest clinically meaningful difference for sample size re-estimation: 0.1 (Active: 0.3 vs Placebo: 0.2)

症例数再推定を用いた場合，両群の奏効率が 0.2 と等しいシナリオを除くすべてのシナリオで条件付き検出力および最終的な検出力（Power）が症例数再推定なしと比較して増加した．一方，バイアスは全奏効率にかかわらずおおむね±1%以内に収まり，MSE や被覆確率についても症例数再推定の有無による大きな差は認められなかった．

また，第一種の過誤確率の制御状況をより詳細に評価するため，両群の真の全奏効率が等しいシナリオについて，全奏効率を 0.05 から 0.5 まで変化させた場合のシミュレーション結果を表 3 に示す．

表 3．シミュレーション結果①-3

Placebo	Active	Average of conditional power achieved	Futility Proportion	Type I Error Rate	Proportion of Sample Size Added	Expected total number of subjects (N)	Bias	MSE	Coverage Probability
0.05	0.05	0.2854	0.5583	0.0242	0.1790	390.8	-0.0057	0.0006	0.9528
0.1	0.1	0.2801	0.5443	0.0244	0.1641	390.1	-0.0078	0.0011	0.9494
0.15	0.15	0.2799	0.5332	0.0254	0.1641	392.9	-0.0088	0.0015	0.9487
0.2	0.2	0.2781	0.5420	0.0220	0.1692	392.0	-0.0112	0.0019	0.9499
0.25	0.25	0.2796	0.5550	0.0240	0.1698	388.8	-0.0113	0.0022	0.9502
0.3	0.3	0.2797	0.5837	0.0257	0.1646	380.7	-0.0116	0.0025	0.9490
0.4	0.4	0.2842	0.5746	0.0256	0.1767	385.9	-0.0118	0.0028	0.9507
0.5	0.5	0.2802	0.5733	0.0267	0.1582	380.4	-0.0120	0.0030	0.9507

シミュレーション回数：各条件 10,000 回
Maximum total number of subjects stage1: 240 stage2: 544
Bias represents the average difference between the estimated and true treatment effect, indicating systematic error in estimation.
MSE (Mean Squared Error) quantifies the overall estimation accuracy by incorporating both variance and bias squared.
Coverage Probability refers to the proportion of simulations in which the 95% confidence interval contains the true treatment effect, serving as an indicator of the reliability of the interval estimates.
Smallest clinically meaningful difference for sample size re-estimation: 0.1 (Active: 0.3 vs Placebo: 0.2)

いずれの奏効率においても，第一種の過誤確率は 0.025 以下にほぼ制御されていた．また，群間差の推定値のバイアスは，全奏効率の増加に伴って増加するものの 1.2%以下であった．

5.3．グラフによる CPZ デザインの動作特性の評価

この節では，5.1.で用いた基本設定に対して，３つのパラメータをそれぞれ変更した場合の，CPZ デザインの動作特性への影響をグラフを用いて評価する．

① N_{max} を変えて比較した場合

試験全体の 1 群あたりの最大症例数（N_{max}）を 300 から 500 まで変化させた場合の条件付き検出力のグラフを図 5 に示す.

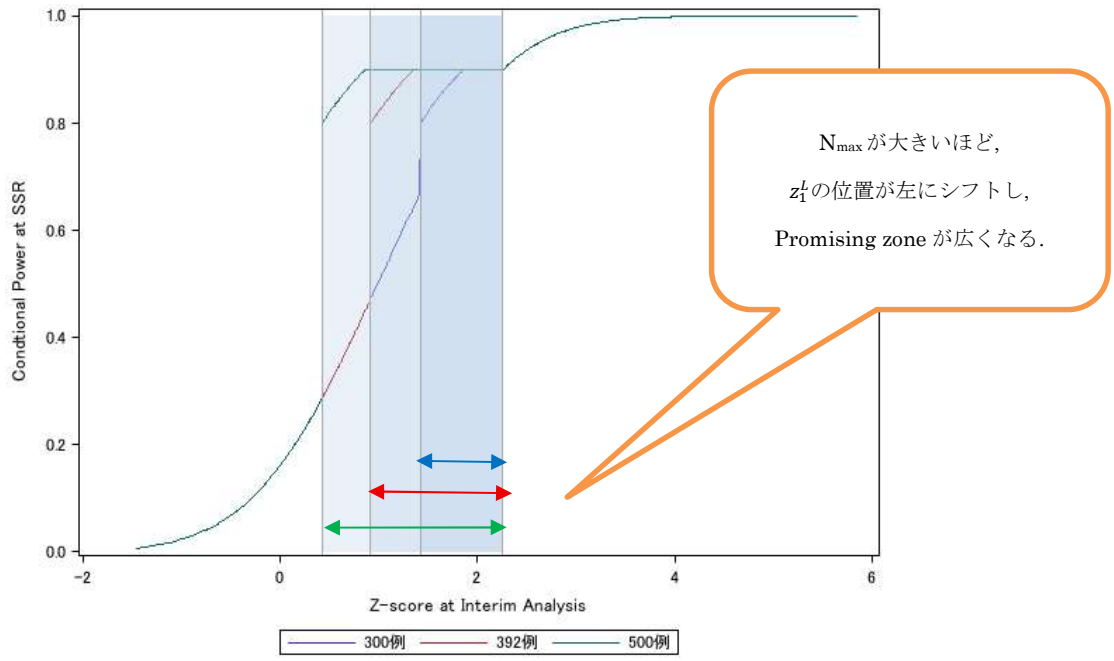


図 5. 中間解析時 Z 値と条件付き検出力の関係①

N_{max} が大きいほど Promising zone は広くなる. これは，最大症例数を増やすことで中間解析後の追加症例数が多くなり，CP_{min} に到達できる中間解析の Z 値の範囲が広くなるためである.

また試験全体の 1 群あたりの最大症例数（N_{max}）を 300 から 500 まで変化させた場合のシミュレーション結果を表 4 に示す.

表 4. シミュレーション結果①

N_{max}	Placebo	Active	Average of conditional power achieved	Futility Proportion	Power	Proportion of Sample Size Added	Expected total number of subjects (N)	Bias	MSE	Coverage Probability
300	0.2	0.33	0.8476	0.0163	0.9122	0.3048	502.5	-0.0011	0.0018	0.9362
392	0.2	0.33	0.8819	0.0163	0.9348	0.4037	543.0	-0.0011	0.0017	0.9423
500	0.2	0.33	0.9120	0.0163	0.9592	0.4599	578.9	-0.0011	0.0017	0.9426

シミュレーション回数：各条件 10,000 回
Bias represents the average difference between the estimated and true treatment effect, indicating systematic error in estimation.
MSE (Mean Squared Error) quantifies the overall estimation accuracy by incorporating both variance and bias squared.
Coverage Probability refers to the proportion of simulations in which the 95% confidence interval contains the true treatment effect, serving as an indicator of the reliability of the interval estimates.
Smallest clinically meaningful difference for sample size re-estimation: 0.1 (Active: 0.3 vs Placebo: 0.2)

N_{max} が 392 例よりも低い場合，Power および期待される症例数は N_{max} が 392 例の場合よりも減少した. 一方，N_{max} が 392 例よりも高い場合，Power および期待される症例数は N_{max} が 392 例の場合よりも増加した .

② δ_{min} を変えて比較した場合

臨床的に意味のある最小の治療効果（δ_{min}）を 0.05 から 0.15 まで変化させた場合の条件付き検出力のグラフを図 6 に示す．なお，δ_{min} の変更にあたっては，Placebo 群の想定する奏効率を調整した．

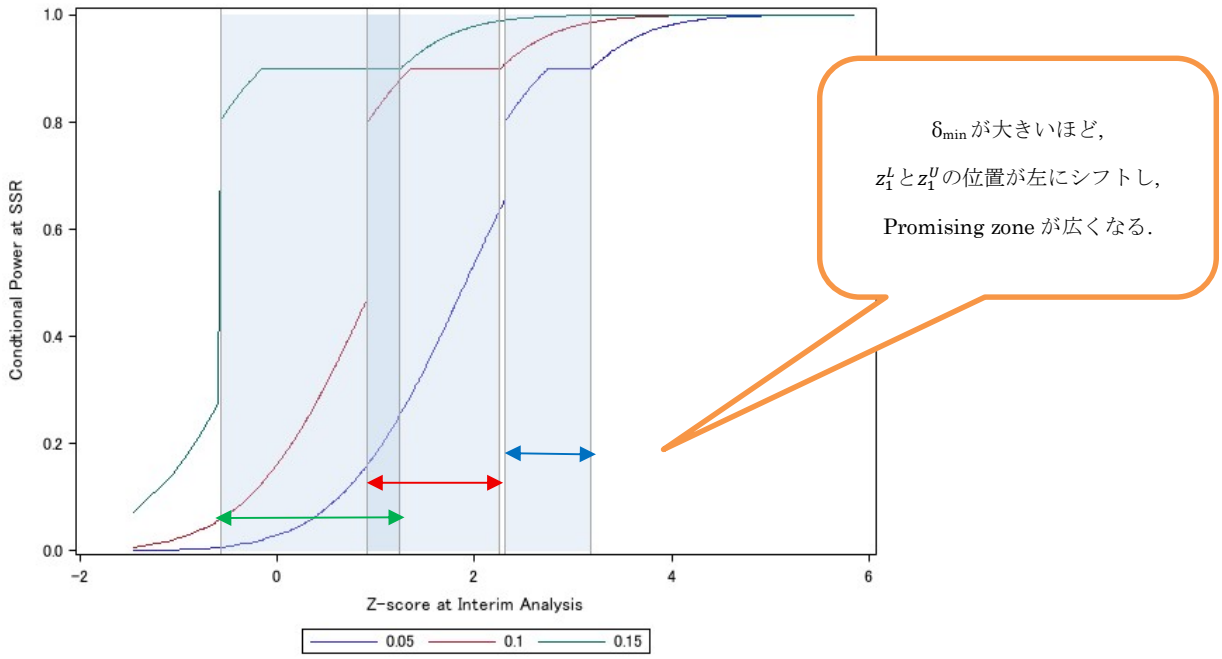


図 6．中間解析時 Z 値と条件付き検出力の関係②

δ_{min} が大きいほど，中間解析時に低めの Z 値であっても，条件付き検出力は高くなり，Promising zone は広くなる．これは，想定効果が大きければ必要症例数が少なくても所定の条件付き検出力（CP_{min}）に到達できるためである．一方，δ_{min} が小さい場合は，同じ Z 値であっても条件付き検出力が低くなり，Promising zone は狭くなる．これは，CP_{min} を達成するにはより多くの症例と高い Z 値が必要になるためである．

臨床的に意味のある最小の治療効果（δ_{min}）を 0.05 から 0.15 まで変化させた場合のシミュレーション結果を表 5 に示す．

表 5．シミュレーション結果②

δ_{min}	Placebo	Active	Average of conditional power achieved	Futility Proportion	Power	Proportion of Sample Size Added	Expected total number of subjects (N)	Bias	MSE	Coverage Probability
0.05	0.2	0.33	0.6364	0.1091	0.8439	0.3186	529.0	-0.0053	0.0022	0.9422
0.1	0.2	0.33	0.8819	0.0163	0.9348	0.4037	543.0	-0.0011	0.0017	0.9423
0.15	0.2	0.33	0.9698	0.0006	0.9192	0.1414	493.7	0.0000	0.0016	0.9406

シミュレーション回数：各条件 10,000 回
Maximum total number of subjects stage1: 240 stage2: 544
Bias represents the average difference between the estimated and true treatment effect, indicating systematic error in estimation.
MSE (Mean Squared Error) quantifies the overall estimation accuracy by incorporating both variance and bias squared.
Coverage Probability refers to the proportion of simulations in which the 95% confidence interval contains the true treatment effect, serving as an indicator of the reliability of the interval estimates.

δ_{min} が 0.1 よりも低い場合，Power は減少し，群間差の推定値のバイアスは増加した．一方，δ_{min} が 0.1 よりも高い場合，Promising zone は拡大したもの，症例数追加が起こる割合が減少し，結果的に Power および群間差のバイアスも減少した．

③　CP_{min} を変えて比較した場合

CP_{min} を 0.33 から 0.85 まで変化させた場合の条件付き検出力のグラフを図 7 に示す．

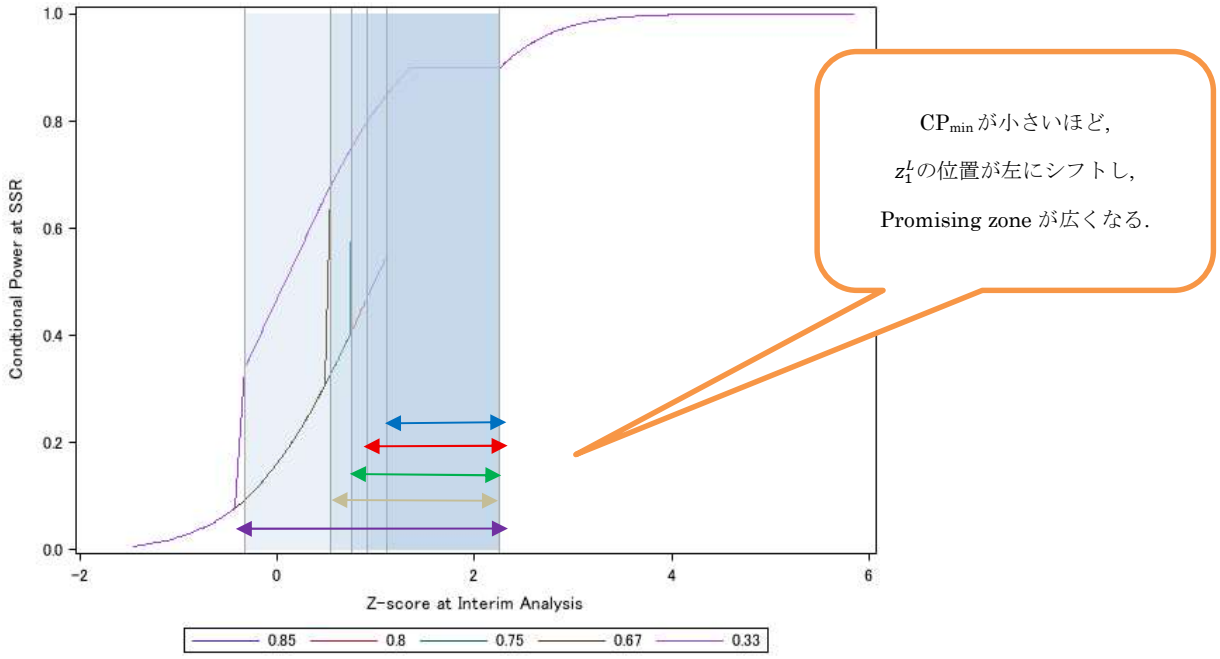


図 7．中間解析時 Z 値と条件付き検出力の関係③

CP_{min} を低く設定すると，中間解析時の Z 値が比較的低い場合でも症例数追加が行われ，高めの条件付き検出力が得られやすくなる．つまり，曲線が立ち上がる位置が左側にシフトし，Promising zone が広くなる．一方，CP_{min} を高く設定すると，追加症例数の増加が行われる範囲が限定され，Promising zone が狭くなる．また，①のパターンと異なり，z_1^C の値が変わらない影響により，症例数の上限となる N_{max} まで追加される中間解析時の Z 値の範囲が広くなるのも特徴である．

CP_{min} を 0.33 から 0.85 まで変化させた場合のシミュレーション結果を表 6 に示す．

表 6．シミュレーション結果③

CP_{min}	Placebo	Active	Average of conditional power achieved	Futility Proportion	Power	Proportion of Sample Size Added	Expected total number of subjects (N)	Bias	MSE	Coverage Probability
0.85	0.2	0.33	0.8729	0.0163	0.9274	0.3749	534.2	-0.0012	0.0017	0.9406
0.8	0.2	0.33	0.8819	0.0163	0.9348	0.4037	543.0	-0.0011	0.0017	0.9423
0.75	0.2	0.33	0.8902	0.0163	0.9435	0.4284	550.5	-0.0011	0.0017	0.9405
0.67	0.2	0.33	0.8983	0.0163	0.9484	0.4516	557.5	-0.0011	0.0017	0.9407
0.33	0.2	0.33	0.9091	0.0035	0.9669	0.4847	570.7	-0.0003	0.0015	0.9427

シミュレーション回数：各条件 10,000 回
Maximum total number of subjects stage1: 240 stage2: 544
Bias represents the average difference between the estimated and true treatment effect, indicating systematic error in estimation.
MSE (Mean Squared Error) quantifies the overall estimation accuracy by incorporating both variance and bias squared.

Coverage Probability refers to the proportion of simulations in which the 95% confidence interval contains the true treatment effect, serving as an indicator of the reliability of the interval estimates.
Smallest clinically meaningful difference for sample size re-estimation: 0.1 (Active: 0.3 vs Placebo: 0.2)

CP_{min} が 0.8 よりも低い場合，最終解析の Power は CP_{min} が 0.8 の場合よりも増加した．一方，CP_{min} が 0.8 よりも高い場合，Power は CP_{min} が 0.8 の場合よりも減少した．また，群間差の推定値のバイアスは，CP_{min} の増減にかかわらず値はほぼ同じであった．

6. 考察

本稿では，CPZ デザインに基づく症例数再推定を取り入れたアダプティブデザインの実装および統計的評価について検討した．特に，現時点で，国内での報告事例が見られない CPZ デザインの実際的な適用例を想定し，SAS によるシミュレーションと可視化の実装例を提示した．

CPZ デザインの利点として，従来の方法に比してより直感的かつ臨床的納得性の高い試験設計が可能であること，また，条件付き検出力に基づいて試験成功の可能性を柔軟に評価できる点が挙げられる．実際に，シミュレーション結果からは，効果の大きさや最大症例数（N_{max}），臨床的に意味のある最小の治療効果（δ_{min}）などのパラメータ設定に応じて，検出力の増加や推定精度の維持といった望ましい性質が確認された．

近年，希少疾病用医薬品の開発が活発化する一方で，症例数設計時に十分な情報が得られていない状況も増加していると推察される．このような場合，試験規模を抑えて開始し，中間解析において有望な結果が得られた際に追加の被験者登録を行うことで，投資規模を段階的に拡大できる CPZ デザインは，実務的にも魅力的な選択肢となり得るだろう．

一方で，本デザインの実装にはいくつかの課題も存在する．特に，中間解析で条件付き検出力の計算に用いる臨床的に意味のある最小の治療効果（δ_{min}）の設定が Promising Zone の範囲に影響するため，信頼できる推定値が必要となる点がある．また，条件付き検出力の計算や SAS による柔軟な実装に際しては一定のプログラミング技術が要求される．さらに，CPZ デザインでは，各ステージの結果の重みの統合において，計画時の例数配分を用いる．仮に，症例数の追加が発生した場合，相対的にステージ 2 の被験者の寄与がステージ 1 の被験者より小さくなる．このため，「すべての被験者を平等に取り扱うべきである」という “one patient, one vote” の原則に反するという批判がある（Hung et al.， 2014）[17].

本研究の限界としては，あくまでシミュレーションに基づく評価であるため，実際の臨床試験データにおける適用性や，第一種の過誤確率の制御，推定値のバイアスや信頼区間の被覆率の低下といった統計的妥当性の維持については，個別の試験設定に応じて評価を行う必要がある．また，中間解析で非盲検情報を利用する際には，試験運営や意思決定への影響といった運用上のリスクも考慮が必要である．さらに，本研究で設定したパラメータ範囲や再推定手法は一例に過ぎず，他の再推定手法や条件との比較検討は今後の研究に委ねられる．

7. まとめ

本稿では，症例数再推定を伴うアダプティブデザインの一種である CPZ デザインについて，理論的枠組みと実装上のポイントを整理した上で，SAS を用いたシミュレーションおよび可視化による性能評価法を提示した．

CPZ デザインの実務応用にあたっては，条件付き検出力の解釈や症例数の上限設定など，設計段階での合意形成が不可欠であり，視覚的な表現による直感的な理解の促進はその鍵を握ると考えられる．SAS などの商用ソフトウェアにおいても，アダプティブデザインの設計支援や可視化機能がより標準化され，ユーザーにとってより扱いやすい形で展開されることを望みたい．

連絡先：naito.toshiki@nobelpharma.co.jp

Appendix 1：二値変数における CPZ デザインを可視化した図

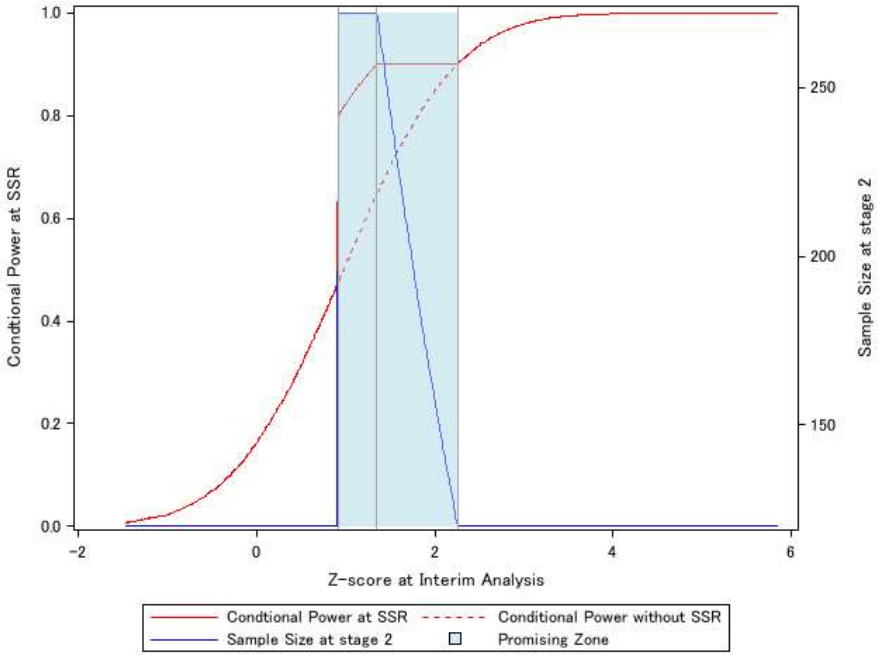


図 8．条件付き検出力と症例数の関係図

5.1.の基本設定のパラメータにおけるシミュレーション

Appendix 2：本稿に添付する SAS プログラム

プログラム名	内容
ssr_cpz_simulation6x_5_2_①.sas	5.2 章の全奏効率の差のみを変えてシミュレーション比較に用いたプログラム（表 1，表 2）
ssr_cpz_simulation6x_5_2_②.sas	5.2 章の全奏効率の差のみを変えてシミュレーション比較に用いたプログラム（表 3）
ssr_cpz_simulation6x_5_3_①.sas	5.3 章の N_{max} を変えてシミュレーション比較に用いたプログラム（図 5，表 4）
ssr_cpz_simulation6x_5_3_②.sas	5.3 章の δ_{min} を変えてシミュレーション比較に用いたプログラム（図 6，表 5）
ssr_cpz_simulation6x_5_3_③.sas	5.3 章の CP_{min} を変えてシミュレーション比較に用いたプログラム（図 7，表 6）
ssr_cpz_simulation6x_app1.sas	Appendix 1 の基本設定における CPZ デザインを可視化した図（図 8）

なお，プログラムの動作確認は SAS9.4 TS1M7 で実施した．

引用文献・サイト

[1] European Medicines Agency. Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design. Committee for Medicinal Products for Human Use, 2007.

[2] U.S. Food and Drug Administration. Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry. Center for Drug Evaluation and Research & Center for Biologics Evaluation and Research, 2019.

[3] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. Complex Innovative Trial Design の現状と事例の整理 - FDA ガイダンス邦訳付き. 日本製薬工業協会, 2024. 東京都. Web. 28 Aug. 2025. https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm00000010pp-att/DS_202404_CID.pdf

[4] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E20: Adaptive Designs for Clinical Trials – Step 2 Draft Guideline. 25 June 2025. Web. 28 Aug. 2025. https://database.ich.org/sites/default/files/ICH_E20EWG_Step3_DraftGuideline_2025_0625.docx

[5] Bothwell, Laura E., et al. "Adaptive Design Clinical Trials: A Review of the Literature and ClinicalTrials.gov." BMJ Open, vol. 8, no. 2, 2018, e018320. https://doi.org/10.1136/bmjopen-2017-018320

[6] Hsiao, S. T., Liu, L., and Mehta, C. R. Optimal Promising Zone Designs. Biometrical Journal, vol. 61, no. 5, 2019, pp. 1175–1186.

[7] Zhao Yang, Shivani Nanda. Promising Zone Designs for Sample Size Re-estimation in Clinical Trials: Graphical Approaches Using SAS. PharmaSUG 2025, San Diego, CA, 2025.

[8] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. アダプティブデザインの統計的推測に関する検討. 日本製薬工業協会, 2023.https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202302_adaptive.html

[9] Lehmacher W, Wassmer G. Adaptive sample-size calculations in group sequential trials. Biometrics 1999;55:1286-1290.

[10] Bauer P and Kohne K. Evaluation of experiments with adaptive interim analyses. Biometrics 1994;50:1029-1041.

[11] Chang M. Adaptive design method based on sum of p-values. Statistics in Medicine 2006;26(14): 2772-2784.

[12] Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. Biometrics 1995;51:1315-1324.

[13] Liu Y, Xu H. Sample size re-estimation for pivotal clinical trials, Contemporary Clinical Trials 2021;102:106215.

[14] Jennison C, Turnbull BW. Adaptive sample size modification in clinical trials: Start small then ask for more? Statistics in Medicine 2015;34:3793-3810.

[15] Wassmer, Gernot, and Friedrich Pahlke. f_simulation_base_rates.R. rpact: Confirmatory Adaptive Clinical Trial Design and Analysis. R package source code. 2025. Web. 28 Aug. 2025. https://rdrr.io/cran/rpact/src/R/f_simulation_base_rates.R

[16] Wolbers, Marcel, Gernot Wassmer, and Friedrich Pahlke. Simulation of a Trial with a Binary Endpoint and Unblinded Sample Size Re-Calculation with rpact. 25 July 2025. Web. 28 Aug. 2025. https://www.rpact.org/vignettes/planning/rpact_sample_size_reassessment_examples/

[17] Hung HMJ, Wang SJ, Yang P. Some challenges with statistical inference in adaptive designs. Journal of Biopharmaceutical Statistics 2014;24:1059-1072.

非臨床統計解析ソフトウェアEXSUSとStat Preclinicaの比較検討

○本田主税[1,2]

（小野薬品工業株式会社　[1]研究プロジェクト統括部　研究推進部、

[2]ファーマコヴィジランス統括部　PVシステム部　PVデータ戦略室）

A comparative study of EXSUS and Stat Preclinica, SAS-based non-clinical statistical analysis software

Chikara Honda[1,2]

[1] Research Promotion, Research Project Management Div.,

[2] Data Strategy Office, System Dept., PV Div.,

ONO PHARMACEUTICAL CO., LTD.

要旨

生物実験データ統計解析システム EXSUS （イグザス）ver.10.2（イーピーエス株式会社、2024 年 11 月リリース）と非臨床開発者向け統計解析ツール Stat Preclinica ver.4.2（株式会社タクミインフォメーションテクノロジー、2025 年 7 月リリース）は Microsoft Excel からデータを取り込み、SAS で解析した結果を出力する非臨床統計解析ソフトウェアである。両ソフトウェアともに、プログラムコードを記述することなく SAS の解析が実行できるため、医薬品の研究者の間で活用されてきた。本発表では、両ソフトウェアの機能、操作、信頼性面などについて、ユーザー視点で比較検討した結果を紹介する。

キーワード : 薬理試験、非臨床統計、EXSUS、Stat Preclinica、対応のあるデータの解析、逐次割付け、信頼性

はじめに

Graphical User Interface の非臨床パッケージの需要と変遷

生物実験データ統計解析システム EXSUS（イグザスと読む）は、浜田知久馬先生監修のもと、当時の有限会社アーム（山本典子社長）により開発されたソフトウェアである。2004 年にユーザー会が創設され、ユーザーの意見を取り入れながら機能拡張が進められてきた。1995 年の発売当初は EXSAS という名称であったが、事情により EXSUS に改称された。非臨床開発者向け統計解析ツール Stat Preclinica（以下、SPC）は、SAS Institute Japan が開発する SAS 前臨床パッケージの開発中止を受け、搭載されていた手法の一部を継承する形で、2008 年株式会社タクミインフォメーションテクノロジーにより開発された。解析結果の出力や機能仕様書の様式は類似しているものの、データ取込み方法や取込み後データの加工ができない点は異なっている。発売当初は SAS BI Server 版のみであったが、2014 年より Windows Server 及び Client 版がリリ

ースされ、導入のハードルが下がった。EXSUS、SPC ともに生物統計を必要とする研究部門、生物実験を取り扱う CRO（Contract Research Organization）などで活用されている。

EXSUS と SPC の比較検討

操作性とアウトプット

EXSUS は、リリース当初から ver.8 まで解析手法の選択方法が直列的（データの型依存）であり、解析者が実施したい手法を選択するまでのステップが分かりにくいという課題があった。しかし ver.10 以降（ver.9 は欠番）では並列的な（目的別の）選択が標準となり、目的とする手法の選択が容易となった。EXSUS は解析方法に応じ、Excel Sheet に予め決められた様式でデータを入力しておく必要がある。読み込み列が指定されているため、想定通りに入力しないとデータ読み込みができない。データ読み込みステップで Excel Book を指定した後、Sheet を読み込むためステップは多いが、列の役割指定は読み込んだ Sheet で完了しているため、改めて指定する必要はない。解析結果は Excel Book に新しい Sheet として追加される。解析結果の引用や二次利用がしやすい反面、意図せぬ Sheet 上書きなどが起こる可能性がある。そのため、解析を実行した Excel を編集不可にする機能が別途用意されている。解析結果には実行した解析手法などが出力されないため、解析情報付加機能が別途用意されている。グラフは PNG 形式で、Excel ファイルとは別に出力される。

SPC はリリース当初より並列的な選択メニューとなっている。データは Excel Sheet に予め縦持ちで入力しておき、読み込む領域をコピーして、SPC のデータ入力画面にペーストする。データを読み込んだ後、SPC 側で役割指定のステップがあるため、読み込むデータに解析に用いない列が入っていても除去できる。解析結果は PDF に出力される。実行した解析手法とその条件、解析に用いた入力データ、グラフも一つの PDF として出力される。実行した時間や PC にログインしたユーザー名が Footer に出力されるため管理もしやすい。

EXSUS、SPC に共通する事項として、読み込み後のデータはソフトウェア側で積み重ねなどの構造変化ができないことがある。一般に生物系研究者は横持ちでデータをまとめるが、解析用データは基本的に縦持ちであるため、データの並び替えは事前に行っておく必要がある（SPC の機能外となるが、SPC では構造変更の Excel マクロが提供される）。また、手法毎に入力フォーマットに違いがあり使いにくかったが、最新バージョンでは手法毎に対応した出力フォーマットがすぐ呼び出せるようになり、操作性は向上している。

	EXSUS	SPC
データ読み込み	Excel Book 、Sheet の選択（2 ステップ）	Excel データをコピー（1 ステップ）
変数指定	（なし）Sheet 列で指定	読み込んだデータ列から指定するが 読み込み方法により、自動設定可
数値変換（連続値）	対数変換、基準値からの差、変化率	対数変換
解析結果出力	Excel	PDF（補助的に CSV）
グラフ体裁	グラフメニューで軸の変更可	軸変更不可　デフォルト出力のみ
グラフ出力	PNG	解析結果とともに PDF
元データとの対応	出力 Sheet に入力 Sheet 名が印字	解析結果とともに PDF
解析情報付加	Professional 機能（有償オプション）	解析結果とともに PDF
出力の桁数制御	Professional 機能（有償オプション）	不可
Excel Book プロテクト	プロテクトモード（有償オプション）	ー
縦⇒横持ちデータ変換	なし	Excel マクロで提供

解析手法、機能

EXSUS、SPC 両者で同じ、または類似の解析手法は、比較のため可能な限りまとめて表記した。（そのため、ソフトウェアに搭載されている手法名と一致しない場合があることにご留意頂きたい）。表中の

『○』は搭載されている解析手法、『×』は搭載されていない（別の手法の計算結果から容易に代替できない）解析手法、空欄はメニュー化されていないが、得られた出力から四則演算や代表的な Excel 関数で比較的簡易に検討可能な方法及び類似の解析手法が搭載されている場合とした。それぞれの最新版で搭載された新規手法は『#』で記した。EXSUS、SPC ともに、生物実験（薬理試験、安全性試験）の比重が高く、これらの分野で汎用されない重回帰、ロジスティック回帰、主成分分析などの多変量解析の手法及び薬物動態関連の手法は採用されていない。CMC（Chemistry, Manufacturing and Control）の実験計画法では JMP（JMP Statistical Discovery LLC）、薬物動態の解析では Phoenix WinNonlin（Certara Inc）などが用いられる。

○データの要約

	EXSUS	SPC	解析内容・コメント
要約統計量	○	○	主な基本統計量　EXSUS：項目選択可　SPC：項目固定
外れ値の検出	○		studentized residual SPC：単変数の個体除去にて検出可
多変数の相関	○	○	

○一因子の解析（パラメトリック）

	EXSUS	SPC	解析内容・コメント
正規性	○	○	EXSUS：Shapiro-Wilk （SW）、Kolmogorov-Smirnov（KS） SPC：SW、KS、Cramer-von Mises、Anderson-Darling
等分散性	○	○	EXSUS：Bartlett、Levene（誤差平方） SPC：Bartlett、Levene（誤差平方、誤差絶対値）、O'Brien、Brown-Forsythe
2 標本 Kolmogorov-Smirnov 検定	○	×	
t 検定	○	○	
Welch 検定	○	○	
一元配置分散分析	○	○	
Least Squares Difference 検定（LSD 検定）	○	○	多重性は無調整
LSD 検定（基準群との比較）		○	片側、両側検定を選択可能
Bonferroni 検定	○		
Holm 検定	○		
固定順序法（閉手順）		○#	
Dunnett 検定	○	○	
Dunnett 逐次検定	○	×	
Dunnett 検定（不等分散考慮）		○	
Tukey-Kramer 検定	○	○	
Williams 検定	○	○	

○一因子の解析（ノンパラメトリック）

	EXSUS	SPC	解析内容・コメント
Wilcoxon 検定 *1	○	○	
Kruskal-Wallis 検定*1	○	○	ノンパラ版の一元配置分散分析
ノンパラ LSD 検定	○	×	Joint ranking 無調整
ノンパラ Holm 検定	○	×	Separate ranking Holm 型
Steel 検定	○	○	Separate ranking Dunnett 型
Steel 逐次検定	○	×	
ノンパラ Dunnett 検定	○	○	Joint ranking Dunnett 型
Steel-Dwass 検定	○	○	Separate ranking Tukey 型
ノンパラ Tukey 検定	○	○	Joint ranking Tukey 型
Shirley-Williams 検定	○	○	Separate ranking Williams 型

Shirley 検定	○	○	Joint ranking Williams 型
False Discovery Rate	○		Benjamini & Hochberg （BH） （Separate ranking のみ）基準群型、総当たり型

*1 EXSUS：近似法、正確法、simulation 法が選択できる。SPC：近似法、正確法が選択でき、正確法を選択した場合、計算時間の上限を指定できる（Kruskal-Wallis 検定は χ^2 近似のみ）。

○p 値の調整

	EXSUS	SPC	解析内容・コメント
p 値の調整	○		Bonferroni、Holm、FDR（BH、Two-stage linear step-up procedure）

データの要約、一因子の解析については EXSUS、SPC ともに十分な手法が搭載されている。ノンパラメトリック検定では、EXSUS、SPC ともに近似法のほかに正確法も選択できるようになっている。EXSUS は Dunnett と Steel に逐次検定を搭載している。逐次検定は下降手順で逐次比較していくため、2 つ目以降の検出力が高くなる。有効用量を検討する試験では通常の検定よりも有利となる。SPC は ver.4.2 から固定順序法を搭載した。固定順序法は実務で利用する場面が多く、メニュー化されていると便利である。

○対応のあるデータの解析

	EXSUS	SPC	解析内容・コメント
対応のある t 検定	○	○	
対応のある Wilcoxon 検定	○	○	
Signed 検定	×	○	
共分散分析（基準群比較）	○	○#	EXSUS： LSD、Dunnett SPC： LSD、Dunnett (simulation 型含む)、Dunnett 逐次(simulation 型含む)、固定順序法、Bonferroni
共分散分析（総当たり比較）	○	○	EXSUS：Tukey SPC：Tukey、Bonferroni
乱塊法（Dunnett）	○		GLM procedure 乱塊法の分散分析表とブロック内の誤差分散を用いた Dunnett 結果を出力
対応のあるパラメトリック Dunnett 検定		○	MIXED procedure
対応のあるパラメトリック Tukey 検定	×	○	MIXED procedure
対応のあるパラメトリック固定順序法	×	○#	MIXED procedure
対応のあるパラメトリック Williams 検定	×	○#	
Friedman 検定	○		対応のある多群ノンパラ SPC：メニューとしては存在しないが、Cochrane-Mantel-Haenzel のブロック要因が 1 つの場合と等価
McNemar 検定	○		2×2 分割表の対応のある検定

対応のあるデータの解析手法について、EXSUS、SPC ともに主要な解析が搭載されている。共分散分析は解析の前提条件を確認して使用する必要があり、SPC は解析対象の項目と共変量との関係、すなわち、実験群それぞれの傾き、実験群共通の傾きで回帰直線をプロットする機能を備えている。一方、EXSUS にはその機能が搭載されていないため、別のソフトウェア（JMP、GraphPad Prism など）でグラフを描いて検討する必要がある。SPC は、同一個体（又は同一サンプル）に対照群と各用量群を割付けるクロスオーバーデザインに対応する手法を充実させている。SPC は ver.4.2 から対応のあるパラメトリック検定を拡張し、固定順序法と Williams 検定を追加している。Williams 検定を対応のあるデータの解析に（理論的に）拡張し、シミュレーションによる性能比較を行った上で実装した [1)]。薬理試験では、本来個体の対応を考慮して

解析すべき場面でも、ソフトウェアに搭載されていないことから対応を無視して解析せざるを得ないことがあった。利用者（試験実施者）は、実施時期が群間でアンバランスにならないような試験デザインを行うことが必要である。

○分割表（対応なし）

	EXSUS	SPC	解析内容・コメント
χ^2 検定	○	○	2×2、n×m 表に対応可
Fisher exact 検定	○	○	EXSUS：2 群のみ対応 SPC：n×m 表に対応可

○用量反応関係、傾向性、相関性

	EXSUS	SPC	解析内容・コメント
用量反応関係	○	×	用量を一次成分、二次成分、高次成分（一次、二次以外の回帰成分）に分解した分散分析表
回帰分析	○		傾きの有意性 p 値、モデル適合度%（直線性）、適合度検定 p 値（Lack of fit） 用量は実用量ではなく対比に変換される。傾きの有意性は非直線性成分を分離した誤差分散を使用
Lack of Fit 検定		○	直線性と非直線性（Lack of fit）
パワーモデル回帰	○		傾きとその信頼区間、切片、寄与率
直線回帰分析	○	○	回帰の切片と傾き
平行線検定・効力比（直線）	○	○	
平行線検定・効力比（シグモイド）	○	×	収束アルゴリズム選択可（Marquardt、Gauss-Newton、Newton、Gradient）
Dx 計算（直線）	○	○	
Dx 計算（シグモイド）	○	○	EXSUS：収束アルゴリズム選択可
酵素阻害実験	○	○	SPC：酵素阻害様式の統計的選択法（赤澤らの方法）を搭載
受容体結合実験	○	○	
受容体置換実験	×	○	
Cochran-Armitage 検定 *2	○	○	2 値データの反応割合に対する相関性
Cochran-Mantel-Haenzel 検定	×	○	多値応答の分割表の相関性
Jonckheere-Terpstra 検定 *2	○	○	ノンパラ 3 群以上の傾向性
Tarone 検定	○	×	生存時間の傾向性

*2 EXSUS、SPC ともに近似法、正確法が選択できる（EXSUS の Jonckheere-Terpstra 検定は simulation 法も選択できる）。SPC は正確法の計算時間の上限を指定できる。

EXSUS の用量反応関係及び回帰分析の出力は、用語含めてやや特殊（例：高次成分）であり、回帰の傾きの検定統計量は誤差分散から非直線成分を分離して計算されている。ユーザーはその前提を理解しておく必要があるが、オンラインヘルプ（HTML）や簡易のユーザーズガイド（PDF）を参照しただけでは分からないため、追加説明が必要である。EXSUS はシグモイド曲線を用いた平行線検定・効力比の手法を搭載している。SPC は酵素阻害実験において、赤澤らが提案した酵素阻害様式の統計的選択法のアルゴリズム [2)]を搭載している。SAS 前臨床パッケージで搭載されていた Tarone 検定については、未搭載である。

○多元配置型分散分析（GLM procedure）

	EXSUS	SPC	解析内容・コメント
二元配置分散分析	○	○	EXSUS：交互作用モデルのみ、多重比較なし SPC：交互作用の有無を選択 Dunnett、Tukey を選択

三元配置分散分析	○	×	交互作用モデル（フルモデル）のみ
ラテン方格法（Dunnett）	○	×	個体内、個体間、時期、薬剤に分割した主効果モデルの分散分析表と薬剤の効果の多重比較 Dunnett を出力

EXSUS、SPC ともに一般線型モデル（GLM）を採用している。EXSUS の二元配置分散分析は交互作用を含むモデルのみ、三元配置分散分析は三要因の三次の交互作用項まで含むフルモデルのみである。これらの手法は生物統計の知識を持つ実務経験者の助言をもとに使用することが望まれる。

○経時測定型分散分析

	EXSUS	SPC	解析内容・コメント
一般線型混合効果モデルを用いた経時型分散分析（投与前値を共変量とする解析含む）	○#	○	MIXED procedure EXSUS：共分散構造　Compound Symmetry （CS）、Auto Regressive lag1 （AR1）、Unstructured （UN）、Variance Components （VC） 自由度推定　Satterthwaite、Kenward-Roger1・2、Between-Within　投与前値を共変量とする場合、オプションで選択 SPC：共分散構造　CS、AR1、UN 自由度推定 Kenward-Rogers 投与前値を共変量とする解析は別メニュー
投与群間差の時点毎の比較（2 群）	○	○	SPC：MIXED procedure、共分散構造　CS のみ EXSUS：一般線型混合モデルを用いた経時型分散分析の 2 群の場合に相当
経時型分散分析・多重比較 「投与群間差の時点毎の多重比較（多群）」		○	MIXED procedure 自由度推定　Kenward-Rogers 共分散構造　CS、AR1、UN Dunnett、Tukey を選択
経時型分散分析・多重比較 「経時型分散分析」、「経時型多重比較」	○		GLM procedure 共分散構造 CS のみ、多重比較 Dunnett のみ 経時型多重比較に分散分析表は出力されないため、経時型分散分析と合わせて検討する。結果に Dunnett 型の注釈は付記されないため注意が必要
直交分解型分散分析	○	×	GLM procedure 経時型分散分析の時間、時間×群の交互作用を一次、二次、高次成分に分解した分散分析表が出力される。

EXSUS は従来、一般線型モデル（GLM）を採用していたが、ユーザーの要望を受けて、ver.10.2 より一般線型混合効果モデル（MIXED）を用いた解析が追加された。これにより、ランダム効果を考慮した制限付き最尤法（REML）を用いた推定が可能となり、共分散構造も柔軟に検討できるようになった。投与前値を共変量とする解析、時点毎の多重比較も解析プラットフォーム中のオプションとして採り入れている。時点毎の解析では多重比較法は採用されていない。GLM ベースの手法は修正や削除などを行わず従来通り残している。SPC はリリース当初から SAS 前臨床パッケージの考え方を継承し、一貫して MIXED で統一されている。創薬研究者は一般に高度な統計モデルの利用経験が十分ではないため、生物統計の知識を持つ実務経験者の助言をもとに使用することが望まれる。

○生物学的同等性試験、がん原性試験

	EXSUS	SPC	解析内容・コメント
2 剤 2 期生物学的同等性試験	○	○	
がん原性試験（Peto 検定）*3	○	○	FDA Draft ガイダンスの推奨の方法

			EXSUS：独自フォーマット、FDA Draft ガイダンス記載のフォーマット選択可
がん原性試験（Poly-k 検定）	○	×	FDA Draft ガイダンスの推奨の方法
がん原性試験（χ^2、Cochran-Armitage）	○		FDA Draft ガイダンス発出前の従来法

*3 EXSUS、SPC ともに近似法、正確法が選択できる

2 剤 2 期生物学的同等性試験は「後発医薬品の生物学的同等性試験ガイドライン」をもとに設計されている。今後、ICH-M13A（即放性経口固形製剤の生物学的同等性）ガイドライン（案）に合わせた変更が必要となるかもしれない（イーピーエス社は、同社の生物学的同等性試験システム BESTS での修正・機能拡張を予定している）。がん原性試験は、FDA Draft ガイダンス [3]の手法をもとに設計されている。それ以来、ガイダンスの更新や業界から新たな解析方法の提案はなく、同手法の修正・変更は行われていない。がん原性試験は、専門技術を持つ CRO に委託することが一般的となっており [4]、解析用データセット（SEND format、tumor.xpt など）の作成も委託に含まれる場合、委託者が EXSUS や SPC を用いて解析する機会は少なくなっていく可能性がある。SPC には Peto 検定のみ搭載されている。Peto 検定以外の解析については、FDA ドラフトガイダンスの邦訳「医薬品のがん原性試験における統計的側面」（サイエンティスト社 2002）の付属ファイルとしてマクロが提供されている。

○生存時間

	EXSUS	SPC	解析内容・コメント
Kaplan Meier 法による生存推定	○		EXSUS：生存関数の推定（時点毎の At Risk、打切り）#、メディアン生存時間とその信頼区間（信頼区間推定法は選択可） SPC：生存関数の推定（打切り）、四分位の生存時間とその信頼区間
生存時間の違いの検定（総当たり）	○	○	EXSUS：logrank 検定、一般化 Wilcoxon 検定、likelihood ratio 検定 SPC：logrank 検定、一般化 Wilcoxon 検定
生存時間の違いの検定（多重比較）	○	○	logrank 検定、一般化 Wilcoxon 検定 EXSUS：LSD、Dunnett、Tukey、Bonferroni を選択 SPC： LSD、Dunnett、Tukey、Bonferroni、固定順序法（Joint rank、Separate rank）を選択
Cox 比例ハザードモデル		○	ハザード比

EXSUS、SPC ともに logrank 検定及び一般化 Wilcoxon 検定の多重比較が充実している。SPC は更に固定順序法を搭載している。固定順序法は実務でよく利用する方法であり、メニュー化されていると便利である。 EXSUS は Cox 比例ハザードモデルを搭載していないが、ユーザーにヒヤリングの上、採否を検討し、現在は搭載していないようである。EXSUS は時点毎の At risk を解析結果及び Kaplan-Meier 曲線のグラフ中に表示するが、SPC はいずれも表示されない。At risk の表記は論文投稿の際に求められることもあり、表示したほうがよいと思われる。

○群分け

	EXSUS	SPC	解析内容・コメント
個体除去	○	○	EXSUS：割付け手法に包含（マハラノビス距離、第一因子から選択） SPC：個別メニューとして搭載（マハラノビス距離）
完全無作為割付け	○	○	
層別無作為割付け	○	○	

逐次割付け	○	○	EXSUS：アルゴリズムは最小値法と閾値設定法、閾値設定（因子毎→説明省略）から選択する。最小化法：平均に関する分散分析 p 値、分散に関する Levene or Bartlett 検定 の p 値の最小値が最大の群分けを採用する。閾値設定法：分散に関する p 値に対して設定閾値を設定し、分散に関する p 値が設定閾値を上回る条件で、平均に関する p 値の最小値が最大になる群分けを選択する。シミュレーション回数のデフォルトは 1000 だが、自由に設定可能 SPC：第一因子の平均に関する分散分析 p 値、分散に関する Levene 検定 の p 値が設定閾値（デフォルトは 0.4） 以上の割付けから最良の割付けを探し、p 値の積もしくは和（デフォルトは積）が最大の群分けを採用する。シミュレーション回数のデフォルトは 100 だが、自由に設定可能
逐次層別割付け	○	×	層別割付を逐次に行う。シミュレーション回数のデフォルトは 50

EXSUS、SPC ともに、実験クールが複数回に渡り、実験を通じて複数の測定項目の平均と分散が揃うように逐次割付けを行える機能が搭載されている。逐次割付けは動的割付け（最小化法）の発展版であり、実験全体として測定項目の偏りを最小限に抑えることができる。一方で、無作為性が多少なりとも犠牲になる。逐次割付けが必要な実験としては、大動物の病態モデルや希少なヒト臨床サンプルを用いた繰り返し実験などが考えられる。薬理実験の場合、使用する動物やサンプル数が制限される一方で、対象とする背景要因の異質性は臨床研究のそれと比べて小さく、未測定要因の影響は少ないと考えられる。そのため、無作為性が損なわれることよりも、測定項目が群間でバランスしたデザインにするメリットが上回る。逐次割付けのアルゴリズムの詳細は紙面の都合で割愛するが、EXSUS、SPC ともにユーザーが設定する項目が多く、どの条件で良い割付けができるのかの情報はない。過去のシミュレーション検討[5)]を深く掘り下げて、条件設定の参考情報が得られると有益である。

○例数設計・検出力算出

	EXSUS	SPC	解析内容・コメント
例数設計・検出力	○	○	EXSUS：t、paired t、Dunnett、Tukey、χ^2、Fisher、logrank、一般化 Wilcoxon 検定 SPC：t、paired t、Dunnett、Tukey、Williams、χ^2、Fisher、logrank 検定

EXSUS、SPC ともに代表的な手法の例数・検出力の算出が可能である。EXSUS は一般化 Wilcoxon 検定、SPC は Willams 検定の例数・検出力の算出も可能である。経験的には、一般化 Wilcoxon 検定は logrank 検定、Willams は t 検定の例数・検出力で概ね代用できるようであるが、マニュアルなどには情報や参考文献の記載はない。代用ではなく、手法として搭載されているほうがよいと思われる。

○ツリー型アルゴリズム

	EXSUS	SPC	解析内容・コメント
ツリー型アルゴリズム	○	○	EXSUS：Dunnett type、Tukey type、LSD type （以下有償オプション機能）Type H、O 、M 、F 、J 、K SPC：t/Welch、Dunnett/Steel、Williams/Shirley Williams、Tukey/Steel Dwass

EXSUS と SPC のアルゴリズムは異なる。EXSUS は Bartlett 検定にて不等分散($p< 0.05$)かつ第 1 群の例数が 5 例以下の場合データを対数変換した後、パラメトリック法で、不等分散かつ第 1 群の例数が 6 例以上の場合 Joint ranking ノンパラメトリック法、等分散の場合パラメトリック法で計算される。EXSUS に有償オプション機能として搭載されているアルゴリズムは、過去にガイドラインで群構成や例数設定が決められ、試験系として固定された毒性分野で用いられてきた。薬理実験ではあまり使用されていないと考えられる。

SPC は Bartlett 検定で等分散性の有意水準を解析者が設定し、p 値が設定水準以下、もしくは分散が 0 またはオブザベーション 1 の群が存在する場合ノンパラメトリック法、それ以外はパラメトリック法が選ばれる。

EXSUS、SPC いずれのアルゴリズムにおいても、得られたデータからパラ・ノンパラ法を選択するため、時点が複数存在する場合に時点毎にパラ・ノンパラ解析が混在し、後に問題となることがある。そのため、ツリー型アルゴリズムの機械的な適用には注意が必要である。

ユーザーズガイド（機能仕様）、ヘルプ機能

EXSUS、SPC ともに、手法選択メニューには、その手法に関係するヘルプが用意されている。EXSUS はオンラインヘルプとして HTML ページに集約されている。SPC は各手法に対して機能仕様書の PDF と、出力結果サンプルの PDF がリンクされている。報告者が考える Pros Cons を表にまとめた。

	EXSUS	SPC
Pros	HTML 形式で検索しやすい。	手法毎に個別の解説と出力サンプルが用意されており（別々の PDF）、詳しい解説と更新履歴があるため、内容を深く堀り下げやすい。
Cons	EXSUS が搭載されている環境でなければ、オンラインヘルプを参照できない。十分な解説がない手法や、他のメニューに解説があるにもかかわらずリンクがない手法がある（新しく搭載された手法ほど例示や解説が少ない）。	個別の PDF のため総覧的に参照できない。内容の詳しさを優先しているため分量が多く、手法全体を捉えにくい。昔から搭載している手法と新しく搭載した手法で記載の粒度が異なる（ユーザーが迷うほどではない）。

サポート体制

EXSUS は問い合わせ窓口として、SPC は保守サポート窓口として、ソフトウェアに関する技術的な質問を受け付けている。EXSUS は開発当初より浜田知久馬先生がソフトウェアを監修し、現在は魚住龍史先生が引継いでいる。加えて、EXSUS では 2004 年にユーザーが発起となりユーザー会が発足し、現在も年 2 回の頻度で開催されている。ユーザーはユーザー会で自施設での課題や悩みを発表することにより様々な助言が得られる。ユーザー会は相互の情報交換やソフトウェアの改善提案の場としても活用されている。一方、SPC は非臨床統計のエキスパートが複数名で開発・サポートにあたっており、ユーザーの意見を取り入れることに加え、独自に潜在ニーズを汲み取り、ソフトウェアに搭載する体制が整っている。また、非臨床統計に関するコンサルティングサービス（別料金）があり、個別に種々の相談が可能である。

薬理研究者の使用感アンケート

EXSUS、SPC の使用感をユーザー視点で比較するため、弊社内の薬理研究者及び薬理研究者の統計アドバイザーを対象に、実機でのテスト使用依頼とアンケートを実施した。アンケート結果は当日紹介する。

ソフトウェア導入・活用時に留意すべきこと

EXSUS、SPC ともに医薬品の承認申請に用いる試験の解析に使用する場合、所属組織の方針に沿った検証（バリデーションなど）を行う必要がある。どの程度の信頼性レベルで維持・管理するかについては、組織のポリシーに則り、研究管理部門あるいは IT 部門と相談の上で導入することが望ましい。本発表の主題から逸れるため深堀りはしないが、導入後に考慮すべきことについて報告者の考えを記す。

試験実施時

EXSUS、SPC ともに出力データを保存しただけでは信頼性確保の面で不十分であり、試験責任者や試験実施者が二次データ（個表）に必要な記述を残し、結果の追跡性を確保することが必要である。EXSUS は、別料金の Professional 機能（解析情報付加機能 、出力桁数指定機能 、印刷コントロール機能 、履歴管理機能）及びプロテクトモードを導入することで、個表の作成を容易にし、解析結果の上書きを防止できるため、信頼性・管理レベルともに向上する。SPC は解析の設定、使用データ、解析結果、実施日時、実施者が一つの PDF に集約されるため、個表は簡潔に作成でき、記録の保存・管理面で省力化が期待できる。

環境変更時

解析環境が変更となった場合、導入時に立ち返り、必要に応じて追加のエビデンス（IQ、OQ、PQなど）を残す必要がある。なお、会社の組織体制が変更となった場合、管理部門・管理者があいまいになることがあり注意が必要である。解析環境の維持管理は研究データの信頼性確保に不可欠である。

おわりに

本項では、主要な非臨床統計ソフトウェアであるEXSUS及びStat Preclinicaをユーザー目線で比較し、特徴と課題を挙げた。すでに導入済み、あるいはこれから導入・置換えを検討しているユーザーの参考となれば幸いである。非臨床統計は研究者コミュニティーが小さいうえに、薬理、安全性（毒性）、CMC、薬物動態分野で興味のある事象や規制要件などが異なり、議論の場が分かれている。SASユーザー総会など、分野の垣根なくオープンな場にてコミュニケーションが行われ、情報交換やレベルアップの場となっていくことを期待する。

参考文献

1) 橋本 敏夫、斎藤 和宏、福島 慎二、河口 裕、中西 展大. Williams検定の対応のあるデータへの拡張. 計量生物学. Vol. 44, No. 2, 83-106 (2024)
2) 赤澤 理緒、中山 勝、山崎 亜紀子、橋本 敏夫、浜田 知久馬. 酵素阻害様式の選択のための統計的アプローチ. 日本薬理学雑誌. Vol.133 No.2,313-318 (2009)
3) Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals. Center for Drug Evaluation and Research.FDA-2001-D-0043 (2001)
4) 嶋本 敬介. がん原性試験の実際～CRO担当者の立場から～. 谷本学校 毒性質問箱. Vol. 2023 No.25, 43-48 (2023)
5) 本田主税、中西展大、橋本敏夫、魚田孝志. EXSUS群分け機能拡張の歩み、現在の取り組み、そして今後の展望. 第27回 EXSUSユーザー会 (2019) （閲覧にはEXSUSユーザー向けサイトへの登録が必要）

階層型クラスタリングを活用したヒートマップのSAS実装

○高田 浩成

（イーピーエス株式会社）

Implementation of Heatmap utilizing Hierarchical Clustering in SAS

Hiroshige Takata

EPS Corporation

要旨

マイクロアレイや RNA シーケンスにより測定されたサンプル毎の遺伝子発現量をデンドログラム及びヒートマップに描画した資料は、バイオインフォマティクス分野の学術論文等で多用されている。ヒートマップの縦軸・横軸の片方または両方に対して階層型クラスタリングが用いられている。多変量解析におけるクラスタリングには、ウォード法のような階層型クラスタリングと k-means 法のような非階層型クラスタリングが存在する。前者は近距離とされたデータから次々に結合していき、後者は予めクラスター数を指定した上で分割することで、クラスターを形成する。本稿では、SAS において、階層型クラスタリングには CLUSTER プロシジャ・TREE プロシジャを使用し、クラスタリングによるカテゴリー順序を保持しつつ、SGPLOT プロシジャによりヒートマップ及びデンドログラムを描き、GTL によりそれらを結合出力した。その上で、ランダムな順序と比較した利点を示し、デンドログラムとの参照関係における SAS の課題も考察する。

キーワード：階層型クラスタリング、ヒートマップ、デンドログラム、ウォード法、ゲノムデータベース

1. 背景

1.1 バイオインフォマティクス

ヒートマップとは、ビジュアライゼーション（可視化）手法の一種であり、地理情報システム(GIS)、マーケティング戦略、心理アセスメントなどの幅広い分野で利用される。特にバイオインフォマティクス（生物情報科学）において多用されている。これはライフサイエンス（生命科学）において、次世代シーケンサー等の急速な進展により、データが大量に生成されるようになり、コンピュータによる解析や統計的なアプローチが活躍するようになったためである。ゲノムに限らず生体を構成している各階層（タンパク質、代謝物、表現型など）をそれぞれ網羅的に解析し、生命現象を包括的に理解しようとすることをオミクス解析という [1]。その中でも、遺伝子解析は基盤となる位置付けであり、治療法や医薬品の開発、さらに個別化医療にあたるプレシジョンメディシン（高精度医療）にも寄与する [10]。

クラスタリングはバイオインフォマティクスで広く活用される手法であり、それと組み合わせたヒートマップによる遺伝子発現量の描画は学術論文等で目を引く。発現量については、例えば、疾患患者のみに特定

遺伝子が強く発現することがあったり、がん領域ではがん遺伝子（例：RAS, EGFR, MYC）やがん抑制遺伝子（例：p53, Rb, APC）に大きく分類されたり、対象部位、各因子、シグナル伝達のしくみ、といった複数の異常が積み重なり、複雑な反応をする。このようなパターンの把握はメカニズムの解明に重要になる。SAS においては、それを直接的に一度に実装する機能は存在しないが、ヒートマップを描画するための各ステップを組み合わせることで実現可能である。本稿では、そのプロセスや技法の特徴を紹介し、理解を深める。

1.2 オープンデータソース

バイオインフォマティクスの分野において、様々なカテゴリーで数多くのデータベースが公開され、その大半は無償で利用できる。ゲノムの塩基配列をはじめとし、国際的な人類共通財産のオープンデータソースとして、様々な形式で膨大なデータが蓄積されており、検索・ダウンロードが可能である。カテゴリーごとの分子生物学・ゲノムデータベースの例を Table 1 に示す[1][2]。これらは論文発表を行う場合には、登録機関への登録が義務付けられており、匿名化技術により処理もされる[11]。とくに本稿に関係するゲノムデータベースとして、マイクロアレイ（既知遺伝子を安価に測定）や RNA シーケンス（未知も含め網羅的に高精度で測定）による遺伝子発現量の解析プロジェクトに特化した米国立生物工学情報センター(NCBI)の Gene Expression Omnibus (GEO)が挙げられる[4]。遺伝子発現のカテゴリーでは世界最大で、あらゆる生物種が登録されている。データベースとともにクラスタリングされたヒートマップも示されている。

近年は、次世代シークエンサーに代表される実験装置の飛躍的な発展により、これらのデータベースに登録されるデータ量は爆発的に増加している。その結果、利用者が研究目的に応じて必要なデータを取得するのが困難な状況にもなってきた。そこで本国では、分子生物学分野での多種多様なデータベースを統合し、有機的に関連付けることによってデータの価値を高めることを目的として、バイオサイエンスデータベースセンター(National Bioscience Database Center : NBDC)が設立された。NBDC においては、Integbio データベースカタログなどが公開されており、そこでは、データベースが生物種やデータの種類別などによって分類され、さらに、データベースの概要などが日本語で説明されているため、どのようなデータベースが利用できるかを容易に検索することができる[2]。

カテゴリー	データベース
塩基配列	DDBJ(DNA Data Bank of Japan), ENA(European Nucleotide Archive)
タンパク質立体構造	PDBj(Protein Data Bank Japan), SCOP(Structural Classification of Proteins)
化合物	ChEMBL, PubChem, CSD(Cambridge Structural Database)
ゲノム	UCSC Genome Browser, MBGD(Microbial Genome Database for Comparative Analysis)
遺伝子配列	RefSeq(Reference Sequence), UniProt(Universal Protein Resource)
遺伝子発現	NCBI GEO(Gene Expression Omnibus), Array Express
遺伝子オントロジー	GO(Gene Ontology)

Table 1　分子生物学・ゲノムデータベースの例

2. 手法

本稿における階層型クラスタリングを活用したヒートマップのグラフ作成の一連の手順の概要を示す

[6][7][8][9]。

2.1 スケーリング

発現量の場合、遺伝子ごとにその基準は異なる。母集団における相対的な位置付けを知る場合の特徴量スケーリングとして、データについて、最小値 0・最大値 1 に変換する正規化(Normalization)（式 1）と平均 0・標準偏差 1 に変換する標準化(Standardization)（式 2）がある。データ範囲に明確な決まりが無く、外れ値も考慮し、データが平均からどれだけ離れているかを指標とする場合は、標準化が望ましく、とくにそれによって変換された値を Z スコア（Z 得点）という。なお、一般的に使われる偏差値は、式 3 の通りこの Z スコアを利用して算出している。データの状態によっては、別途、欠測補完やクリーンアップが必要になることもある。

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

$$Z = \frac{x - \mu}{\delta} \tag{2}$$

$$T = Z \times 10 + 50 = \frac{x - \mu}{\delta} \times 10 + 50 \tag{3}$$

（x：任意のデータにおける任意の観測値、μ：任意のデータの平均、δ：任意のデータの標準偏差、x_{min}：任意のデータにおける最小値、x_{max}：任意のデータにおける最大値）

2.2 クラスタリング

多変量解析におけるクラスタリング（クラスター分析）は、機械学習の一種である教師なし学習として、正解ラベルを与えずにデータのパターンや構造を発見する手法である。クラスタリングにおいて、データの集まりをクラスターと捉えるが、階層構造を構築して分類する階層型クラスタリング (Hierarchical Clustering) と、あらかじめ設定したクラスター数に基づいてデータを直接分割する非階層型クラスタリング (Non-Hierarchical Clustering) に大別される。

本稿で用いるのは階層型クラスタリングで、データ間の距離（非類似度）を指標とし、類似しているものほど近い位置にまとめられる。手法には、ウォード法(Ward's Minimum-Variance Method)、群平均法(Average Linkage)、最短距離法(Single Linkage)、最長距離法(Complete Linkage)などの種類があり、一般的にヒートマップに適用されるのはウォード法である。ウォード法は式(4)、群平均法は式(5)が最小になる処理を繰り返す[16][22][23]。ウォード法と群平均法の大きな違いとして、ウォード法はクラスター内の分散（誤差平方和）の増加を最小限に抑えるように計算が複雑になっているが、分類感度は高い。一方、群平均法は計算がシンプルで、柔軟性があり、外れ値を吸収しやすい。

$$D_{wt} = \frac{(n_u + n_t){D_{ut}}^2 + (n_v + n_t){D_{vt}}^2 - n_t{D_{uv}}^2}{n_u + n_v + n_t} \tag{4}$$

$$D_{wt} = \frac{n_u D_{ut} + n_v D_{vt}}{n_u + n_v} \tag{5}$$

（D：非類似度、n：データ数、u, v：結合前のクラスター、w：結合後のクラスター、t：任意のクラスター）

非階層型クラスタリングの代表例は k-means 法であり、アルゴリズムが単純かつ計算効率が高いため広く利用されている。とくに大規模データでは、全体的な傾向を把握する初期解析として用いられることが多い。一方で、非線形かつ複雑なクラスター構造にも対応可能な手法として、スペクトラルクラスタリングが提案されている[14]。いずれの手法においても、クラスター数の指定が必要となる。

計算量はアルゴリズムの効率を示す尺度としてオーダー記法(Big-O notation)で表される。これは入力データ数 N に対して、処理に要する計算量が漸近的にどの程度増加するかを示す表記であり、例えば$O(N)$はデータ数に比例して計算量が増加することを、$O(N^2)$はデータ数が 2 倍になると計算量がおおよそ 4 倍になることを意味する。階層型クラスタリングは標準的な実装で$O(N^2)$を要し、データ数が増えると計算量が急速に増大する。一方、非階層型クラスタリングの代表である k-means 法は k をクラスター数とし、$O(kN)$と表され、階層型に比べて計算効率が高い[12][13]。

2.3 デンドログラム

デンドログラム(Dendrogram)は樹形図(Tree Diagram)の一種であり、階層型クラスタリングの結果をグラフ化したもので、観測データやクラスター間の類似性を表す。ノード（節点）は葉(leaf)・根(root)・クラスター形成を含めた地点であり、非階層型とは異なり、葉から根に向かってクラスターが結合する処理過程の階層構造を俯瞰して見ることができる。葉がペアとして、クラスターを形成し、そのクラスターもペアで結合していくトーナメント表のようなグラフになる。

デンドログラムの横軸にはクラスターを構成する要素（例：カテゴリー）が並び、縦軸の高さは距離（非類似度）を表す。高さが高いほどカテゴリー間の非類似度が大きいと解釈できる。ウォード法の場合、この高さはクラスター結合に伴う分散の増加量を距離として算出している。

2.4 ヒートマップ

ヒートマップは、縦軸と横軸にカテゴリー（例：サンプルや遺伝子）を配置し、各セルの数値を色の濃淡やグラデーションで表現することで、2 次元上にデータの大きさを可視化する手法である。これは三次元グラフではないが、色を第三の情報次元として用いる点で類似する。バイオインフォマティクスの分野では、マイクロアレイ解析では緑・黒・赤のカラースキームが古典的に用いられ、近年の RNA シーケンス解析では視認性の向上から青・白・赤の組み合わせも広く利用されている。さらに、クラスタリングと組み合わせることで、遺伝子やサンプル間のパターンや類似性を視覚的に把握することが可能となる。

2.5 グラフレイアウト

複数のグラフを使用するため、レイアウトを考慮し、視覚的に画像を 1 つの画像に結合する。プログラミング言語の機能にもより、各グラフに連携の機能が備わっていれば別であるが、個別にデンドログラムやヒートマップのグラフを描画した場合は、それらの結果を結び付ける必要がある。ヒートマップの横軸・縦軸 2 方向の各カテゴリーにデンドログラムが対応する。その際、画像の配置やサイズを適切に設定するために必要になる。

3. SAS による実装

前章に対応する、様々な SAS プロシジャについて、主要部分のプログラム等を例示しながら説明する。実際は遺伝子数が数千・数万になることも多いが、説明の便宜上、遺伝子の数を絞ったテストデータとした（5 サンプル・10 遺伝子）。なお、遺伝子毎に発現量の基準は異なることを考慮し、遺伝子毎に発現量を平均 10～100、標準偏差 1～10 の一様分布にランダムで生成した。Table 2 のような行：遺伝子、列：サンプルのワイド形式（横持ち）を仮定する。

Gene	Sample1	Sample2	Sample3	Sample4	Sample5
Gene01	82.9	77.9	79.6	83.3	83.6
Gene02	39	50.3	49.1	32	50.3
Gene03	98.4	97.5	92.6	89.3	92
Gene04	20.9	23.1	19.8	22.5	18.4
Gene05	78	76.4	75.7	78.1	70.8
Gene06	93.4	85.6	96	88.4	93.9
Gene07	25.1	39.5	28.9	40	40.2
Gene08	61	63.8	61	61.7	68.2
Gene09	10	16.7	7.8	0.8	8.4
Gene10	52.7	46	52	49.2	48.5

Table 2　対象とするデータセット

3.1 STANDARD プロシジャ／STDIZE プロシジャ

Table 2 のテストデータは行間（遺伝子間）の差が大きいため、基準について、STANDARD プロシジャまたは STDIZE プロシジャにより Z スコアのスケーリングを実装する[15]。ワイド形式のデータセットに対して、var ステートメントで変数を指定する必要があるため、行：遺伝子・列：サンプルに対して、遺伝子ごとにスケーリングするには、行方向になるため、それらを変数とするように、TRANSPOSE プロシジャ等で事前に転置する必要がある。

STANDARD プロシジャでは平均値・標準偏差の設定が必要である、平均値 0、標準偏差 1 とすることで Z スコアを算出できる。STDIZE プロシジャでは method=std で Z スコア変換となる。上記の STANDARD プロシジャと STDIZE プロシジャによる結果は同じになる。なお、STDIZE プロシジャではより詳細な設定として、add=, mult=オプションにより、追加計算で偏差値のような算出もプロシジャ内で可能である。

Table 2 と同じ形式にするには、変換後、再度転置する必要がある。その結果を Table 3 に示す。行間（遺伝子間）の差が抑えられている。

```
/*  事前転置  */
proc transpose data=EXPRESSION_WIDE out=EXPRESSION_TRNS(rename=(_NAME_=Sample)) ;
   id Gene ;
run ;

/* STANDARD プロシジャ：　平均値・標準偏差の設定  */
proc standard data=EXPRESSION_TRNS out=EXPRESSION_TRNS_SCORE mean=0 std=1 ;
   var Gene01--Gene10 ;
run ;
```

```
/* STDIZE プロシジャ：　手法・オプションの設定 */
proc stdize data=EXPRESSION_TRNS out=EXPRESSION_TRNS_SCORE method=std ;
   var Gene01--Gene10 ;
run ;

/* 事後転置 */
proc transpose data=EXPRESSION_TRNS_SCORE out=EXPRESSION_SCORE(rename=(_NAME_=Gene)) ;
   id Sample ;
run ;
```

Program 1　STANDARD プロシジャ及び STDIZE プロシジャによる実装と転置対応

VIEWTABLE: Work.Expression_score_round

	Gene	Sample1	Sample2	Sample3	Sample4	Sample5
1	Gene01	0.56	-1.39	-0.73	0.72	0.84
2	Gene02	-0.62	0.74	0.6	-1.47	0.74
3	Gene03	1.15	0.92	-0.35	-1.21	-0.51
4	Gene04	-0.02	1.12	-0.59	0.81	-1.32
5	Gene05	0.74	0.2	-0.03	0.77	-1.68
6	Gene06	0.45	-1.36	1.06	-0.71	0.57
7	Gene07	-1.34	0.66	-0.81	0.73	0.76
8	Gene08	-0.7	0.22	-0.7	-0.47	1.66
9	Gene09	0.22	1.4	-0.17	-1.4	-0.06
10	Gene10	1.11	-1.35	0.85	-0.18	-0.43

Table 3　Table 2 を各行に Z スコアでスケーリングしたデータセット（小数第 2 位まで表示）

3.2 CLUSTER プロシジャ

階層型クラスタリングは CLUSTER プロシジャにより実装される[16]。ヒートマップに対応する縦軸・横軸 2 方向について行う。こちらもスケーリング時と同様に、var ステートメントで変数指定を行うため、行については事前転置が必要である。手法は method=オプションにより設定が必要である。SAS では多数用意されており、代表例として、ウォード法は「ward」、群平均法は「average」、最短距離法は「single」、最長距離法は「complete」で実装できる。本稿では同解析で一般的に使用されているウォード法を使用する。

outtree=オプションにより Table 4 のデータセットが出力できるが、親クラスターや高さ関係は、デンドログラムの葉の順序そのものには対応していない。高さの意味は手法(method=)により異なり、デンドログラムの_HEIGHT_変数（反転表示箇所）は基本的に距離といえ、(a)のウォード法では SAS では調整済み決定係数(adjusted R^2)との記載があるが、分散の増加量とみなせる。また、群平均法の結果を(b)に示す。CLUSTER プロシジャには、デンドログラムを描画する機能もあるが、SAS 環境や条件により出力されない場合もあるため（Warning 出力）、以降の描画専用の TREE プロシジャを使用するのが確実である。

なお、非階層型クラスタリングの k-means 法は FASTCLUS プロシジャにより実装でき、スペクトラルクラスタリングは SAS ではプロシジャが用意されていない。

```
/* 横軸用（上向きグラフ） */
proc cluster data=EXPRESSION_TRNS_SCORE outtree=SAMPLE_CLST method=ward ;
   id Sample ;
   var Gene01 -- Gene10 ;
run ;
/* 縦軸用（右向きグラフ）：事前転置必要 */
```

```
proc transpose data=EXPRESSION_TRNS_SCORE out=EXPRESSION_SCORE(rename=(_NAME_=Gene)) ;
   id Sample ;
run ;
proc cluster data=EXPRESSION_SCORE outtree=GENE_CLST method=ward ;
   id Gene ;
   var Sample1 -- Sample5 ;
run ;
```

Program 2　2 方向の階層型クラスタリング

VIEWTABLE: Work.Sample_clst (Ward の最小分散法によるクラスター分析)

	OBS またはクラスターの名前	OBS またはクラスターの親	クラスター数	クラスターの度数	半偏 R2 乗	標準偏差の平均平方平方根	半偏 R2 乗	R2 乗
1	Sample1	CL4	5	1	0	0	0	1
2	Sample3	CL4	5	1	0	0	0	1
3	CL4	CL3	4	2	0.0898734395	0.5995779833	0.0898734395	0.9101265605
4	Sample4	CL3	5	1	0	0	0	1
5	CL3	CL2	3	3	0.2598467374	0.8363255071	0.2598467374	0.6502798231
6	Sample5	CL2	5	1	0	0	0	1
7	CL2	CL1	2	4	0.3140906402	0.9407874837	0.3140906402	0.3361891829
8	Sample2	CL1	5	1	0	0	0	1
9	CL1		1	5	0.3361891829	1	0.3361891829	0

(a)ウォード法(method=ward)

VIEWTABLE: Work.Sample_clst (群平均法によるクラスター分析)

	OBS またはクラスターの名前	OBS またはクラスターの親	クラスター数	クラスターの度数	クラスター間の平均距離	標準偏差の平均平方平方根	半偏 R2 乗	R2 乗
1	Sample1	CL4	5	1	0	0	0	1
2	Sample3	CL4	5	1	0	0	0	1
3	CL4	CL3	4	2	0.5995779833	0.5995779833	0.0898734395	0.9101265605
4	Sample4	CL3	5	1	0	0	0	1
5	CL3	CL2	3	3	0.9324235366	0.8363255071	0.2598467374	0.6502798231
6	Sample5	CL2	5	1	0	0	0	1
7	CL2	CL1	2	4	1.0347568819	0.9407874837	0.3140906402	0.3361891829
8	Sample2	CL1	5	1	0	0	0	1
9	CL1		1	5	1.0827642244	1	0.3361891829	0

(b)群平均法(method=average)

Table 4　CLUSTER プロシジャによる出力データセット

3.3 TREE プロシジャ／SGPLOT プロシジャ

TREE プロシジャで CLUSTER プロシジャによる出力データセットを読み込むことで、同様に 2 方向に、Figure 1 のデンドログラムを自動的に直接描くことができる[17]。やはり高さにラベル記載があるが、(a)のウォード法では分散の増加量にあたる。デフォルトでは垂直方向のデンドログラムを horizontal オプションで水平方向にも表示できる。ただし、葉を基準に右向きのみで、左向きは不可である。軸やラベルの設定も不可である。

なお、ヒートマップと関連付けるには、順序データを保持する必要があり、要素名を昇順としてペアの代表を結合するように、データセットを整形することで CLUSTER プロシジャの結果から再現することも可能であるが、list オプションを付け、ODS OUTPUT により、容易にそこから葉の順序を再現したデータセットを出力できる(Table 5)。これはヒートマップに適用するために重要な要素である。CLUSTER プロシジャの outtree=オプションや TREE プロシジャの out=オプションとは異なり、描画されるデンドログラムに対応する葉の順序になっている。親クラスター、その距離・高さも各レコードに示されている。高さは Table 4 の _HEIGHT_変数が用いられている。

ただ、とくにヒートマップと組み合わせる場合、通常はデンドログラムの形状を重視し、軸やラベルは表示しないため、本来は削除する必要がある。そこで、TREE プロシジャでは不可であるため、SGPLOT プロシジャでデンドログラムを再現するプログラムを作成した。整形過程は**＜別途添付プログラム＞**に示す。概要

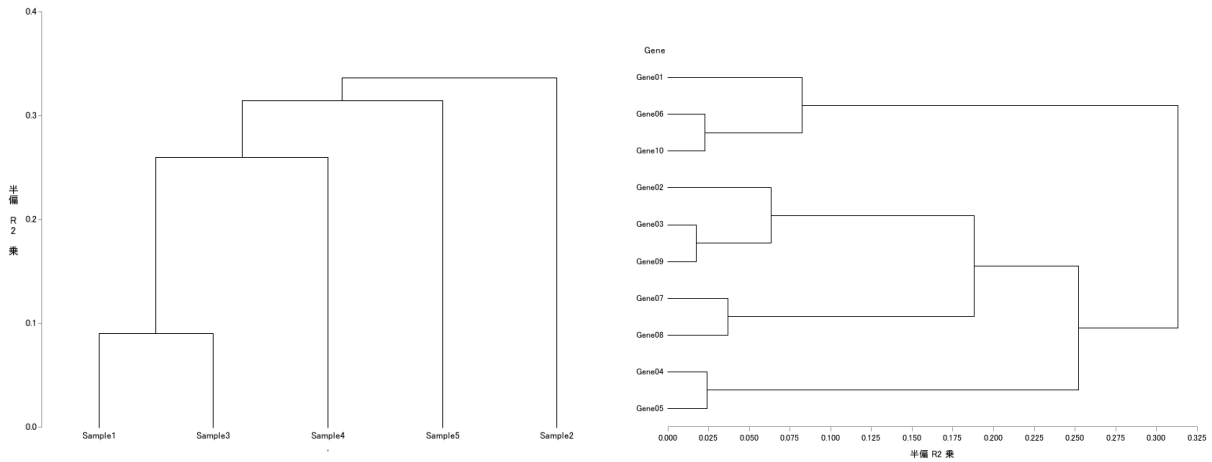
としては、葉の部分から親クラスターへ遡っていき、クラスターの構成カテゴリーのペアの中央の座標を計算して POSITION 変数とし、高さ HEIGHT 変数を保持した（2 変数は座標に関係するがグラフの向きにより回転や反転を行うこともあるため直接 X, Y 変数の表記は避けた）。SGPLOT プロシジャへの入力データセットに向けて整形し、ロング形式（縦持ち）に縦積みした。線分のペアの GROUP 変数と座標となる POSITION 変数・HEIGHT 変数の 3 変数に集約した。Program 4 の SGPLOT プロシジャを経て Figure 2 のグラフを描画する。

SGPLOT プロシジャによって、グループ毎にペアを作り黒色実線で直線を描いていく。グラフ描画において自由が利き、メリットが大きい。基本的なことでは、軸・ラベルの削除が可能になり、series ステートメントにより実線で描き、lineattrs=(thickness=)オプションにより線の太さを調整できる。xaxis, yaxis ステートメントにおいて、reverse オプションにより各軸の回転・反転の設定が容易に可能になる。カテゴリー数が多くなる場合は、グループ数、描く線分の数も増えるため、ODS GRAPHICS で、groupmax=オプションでグループ数の上限を増やしておく必要もある。他グラフと結合するため、norborder オプションで外枠を削除しておく。後にグラフ結合に使用するため、imagename=オプションでファイル名を指定して保存しておく（画像形式はデフォルトでは PNG）。

縦軸用・横軸用ともに葉の順序、クラスターの組み合わせや高さを再現することができた。

```
/* 横軸用（上向きグラフ） */
ods output treelisting=TREELISTING_SAMPLE ;
proc tree data=SAMPLE_CLST list ;
   id Sample ;
run ;
/* 縦軸用（左向きグラフ） */
ods output treelisting=TREELISTING_GENE ;
proc tree data=GENE_CLST list horizontal ;
   id Gene ;
run ;
```

Program 3　TREE プロシジャによるデンドログラム直接描画とデータセット出力



(a)横軸対応（サンプル）　　(b)縦軸対応（遺伝子）

Figure 1　TREE プロシジャによるグラフ出力

VIEWTABLE: Work.Treelisting_sample

	type	batch			
1	h				TREE プロシジャ
2	h	の最小分散法によるクラスター分析			Ward
3	h	'	Height	Parent	Children
4	h				
5	d	CL1	0.336189		CL2	Sample2
6	d	CL2	0.314091	CL1	CL3	Sample5
7	d	CL3	0.259847	CL2	CL4	Sample4
8	d	CL4	0.089873	CL3	Sample1	Sample3
9	d	Sample1	0	CL4		
10	d	Sample3	0	CL4		
11	d	Sample4	0	CL3		
12	d	Sample5	0	CL2		
13	d	Sample2	0	CL1		

(a)横軸対応（サンプル）

VIEWTABLE: Work.Treelisting_gene

	type	batch				
1	h					TREE プロシジャ
2	h					
3	h	Gene	Height	Parent	Children	
4	h					
5	d	CL1	0.313009		CL4	CL2
6	d	CL4	0.082489	CL1	Gene01	CL8
7	d	Gene01	0	CL4		
8	d	CL8	0.022618	CL4	Gene06	Gene10
9	d	Gene06	0	CL8		
10	d	Gene10	0	CL8		
11	d	CL2	0.252288	CL1	CL3	CL7
12	d	CL3	0.187898	CL2	CL5	CL6
13	d	CL5	0.063309	CL3	Gene02	CL9
14	d	Gene02	0	CL5		
15	d	CL9	0.017566	CL5	Gene03	Gene09
16	d	Gene03	0	CL9		
17	d	Gene09	0	CL9		
18	d	CL6	0.036881	CL3	Gene07	Gene08
19	d	Gene07	0	CL6		
20	d	Gene08	0	CL6		
21	d	CL7	0.023941	CL2	Gene04	Gene05
22	d	Gene04	0	CL7		
23	d	Gene05	0	CL7		

(b)縦軸対応（遺伝子）

Table 5　TREE プロシジャの ODS GRAPHICS による出力データセット

VIEWTABLE: Work.Ds_tree_coordinate

	GROUP	POSITION	HEIGHT
1	1	5	0
2	1	5	0.336189
3	2	5	0.336189
4	2	3	0.336189
5	3	4	0
6	3	4	0.314091
7	4	2.5	0.314091
8	4	2.5	0.336189
9	5	4	0.314091
10	5	2.5	0.314091
・・・・・・			
133	67	2	0.314091
134	67	2.5	0.314091
135	68	2.5	0.336189
136	68	3	0.336189

VIEWTABLE: Work.Ds_tree_coordinate

	GROUP	POSITION	HEIGHT
1	1	1	0
2	1	1	0.082489
3	2	2	0.082489
4	2	2	0.313009
5	3	1	0.082489
6	3	2	0.082489
7	4	2	0.313009
8	4	5.5	0.313009
9	5	9	0
10	5	9	0.023941
・・・・・・			
141	71	6	0.252288
142	71	7	0.252288
143	72	7	0.313009
144	72	5.5	0.313009

(a)横軸対応（サンプル）

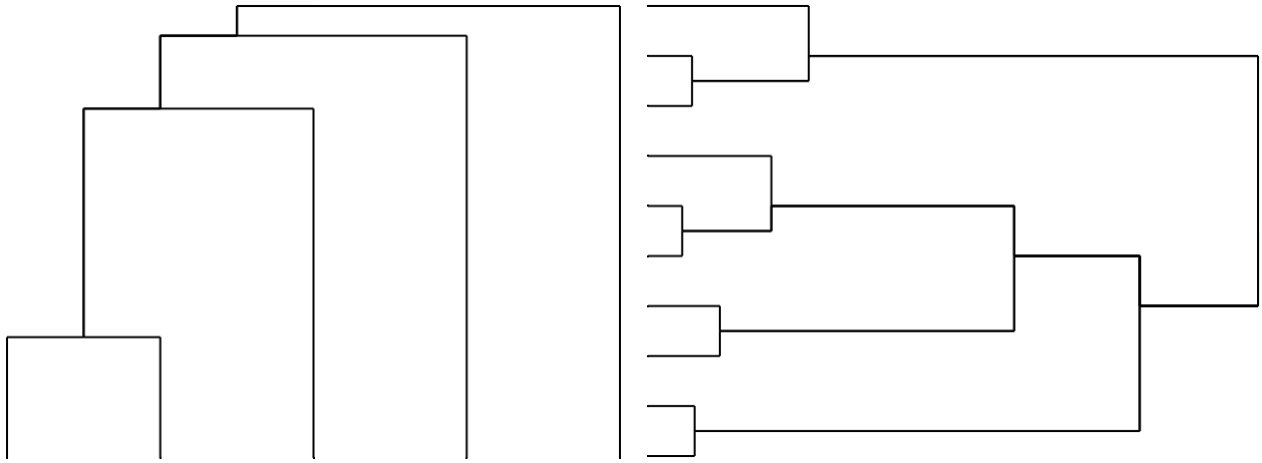
(b)縦軸対応（遺伝子）

Table 6　SGPLOT プロシジャに読み込むために整形したデータセット

```
/* 上向きデンドログラム描画 */
ods graphics on / imagename='DENDROGRAM_X' groupmax=5000 noborder ;
proc sgplot data=DS_TREE_COORDINATE noautolegend noborder ;
   styleattrs datacontrastcolors=(black) datalinepatterns=(1) ;
   series x=POSITION y=HEIGHT / group=GROUP lineattrs=(thickness=2) ;
   xaxis display=none ;
   yaxis display=none ;
run ;
/* 右向きデンドログラム描画 */
ods graphics on / imagename='DENDROGRAM_Y' groupmax=5000 noborder ;
proc sgplot data=DS_TREE_COORDINATE noautolegend noborder ;
```

```
    styleattrs datacontrastcolors=(black) datalinepatterns=(1) ;
    series x=HEIGHT y=POSITION / group=GROUP lineattrs=(thickness=2) ;
    xaxis display=none ;
    yaxis display=none reverse ;
run ;
```

Program 4　SGPLOT プロシジャによるグラフ描画部分



(a)横軸対応（サンプル）　　(b)縦軸対応（遺伝子）

Figure 2　SGPLOT プロシジャにより描画したデンドログラム

3.4 SGPLOT プロシジャ／HEATMAPPARM ステートメント

クラスタリングしたヒートマップは、Program 5 のように、処理過程を経て、SGPLOT プロシジャの HEATMAPPARM により描くことができる。各軸については、従来のグラフと同様に、xaxis, yaxis ステートメントにより設定する。カラーグラデーションは colormodel オプションにより付与でき、濃淡の各段階の色の設定も可能であり、本稿では低発現量から古典的な緑色・黒色・赤色に合わせた。なお、SGPLOT プロシジャにおいても、カラーバーの位置等の調整は不可である。

視覚的にヒートマップはワイド形式のデータセットそのものであるが、SGPLOT プロシジャで読み込むためには、TRANSPOSE プロシジャで転置しつつ、2 変数のロング形式に変形する必要がある。また、クラスタリングを適用するためには、TREE プロシジャの ODS OUTPUT の出力データセットの葉の順序を保持するように、オブザベーションを取得したりしてデータセットに事前にマージ等で紐付ける必要がある。縦軸側の横向きデンドログラムでは上、ヒートマップでは下が始点で、逆になっていることに注意し、デンドログラムから受け取ったカテゴリーは降順にソートする必要がある。

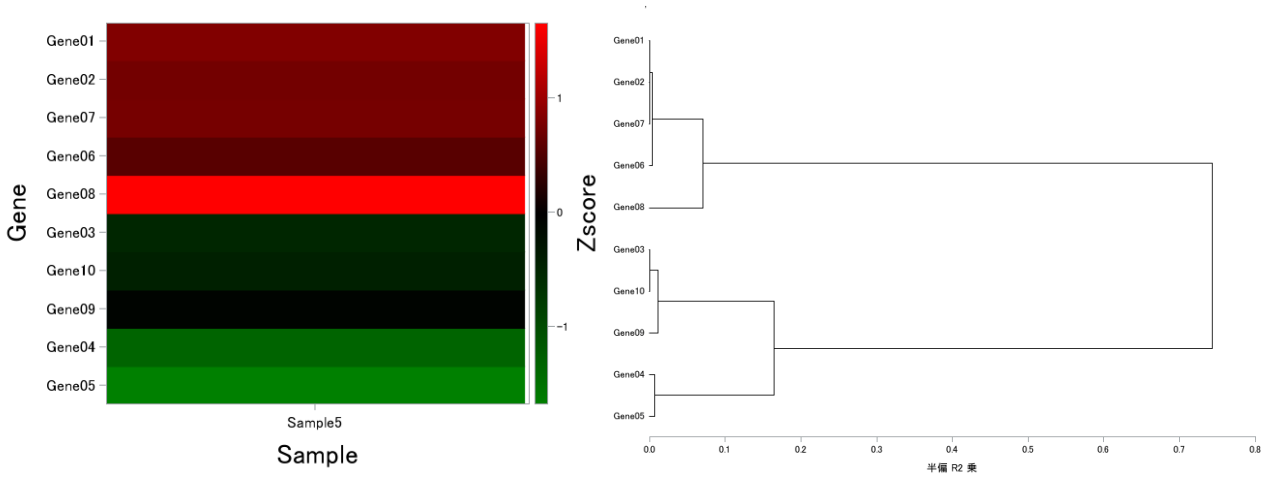
実装例として、はじめに、Figure 3 のように、サンプルを Sample5 に絞って、シンプルな状況で確認した。上半分が赤色系（高発現群）、下半分が緑色系（低発現群）として大きく分けられている。なお、最も高発現の赤色の Gene08 が上側にくるわけではないが、他の遺伝子とは距離（高さ）の差が大きいことは分かる。次に、Figure 4 について、(a)クラスタリング前で、(b)は全データで 2 軸に対してクラスタリングを行った結果では、ヒートマップの横軸・縦軸のカテゴリーがクラスタリングにより入れ替わっている。(a)よりは(b)の集合・クラスターがある程度明瞭になっており、傾向・パターンを把握しやすくなった。

```
/* 葉の順序取得（縦軸・横軸） */
data SAMPLE_CLST_SORT ;
   set TREELISTING_SAMPLE ;
   where type = 'd' and input(scan(batch, 2, ' '), best.) = 0 ;
   length Sample $8 ;
   keep Sample SORT_SAMPLE ;
   Sample = scan(batch, 1, ' ') ;
   SORT_SAMPLE = _N_ ;
run ;
data GENE_CLST_SORT ;
   set TREELISTING_GENE ;
   where type = 'd' and input(scan(batch, 2, ' '), best.) = 0 ;
   length Gene $8 ;
   keep Gene SORT_GENE ;
   Gene = scan(batch, 1, ' ') ;
   SORT_GENE = _N_ ;
run ;

/* ロング形式転置 */
proc transpose data=EXPRESSION_SCORE out=EXPRESSION_SET(rename=(_NAME_=Sample COL1=Zscore)) ;
   by Gene ;
run ;
/* 順序紐付け（縦軸・横軸） */
proc sort data=GENE_CLST_SORT ; by Gene ; run ;
data EXPRESSION_SET_SORT ;
   merge
      EXPRESSION_SET
      GENE_CLST_SORT
   ;
   by Gene ;
run ;
proc sort data=SAMPLE_CLST_SORT ; by Sample ; run ;
proc sort data=EXPRESSION_SET_SORT ; by Sample ; run ;
data DS_HEATMAP ;
   merge
      EXPRESSION_SET_SORT
      SAMPLE_CLST_SORT
   ;
   by Sample ;
run ;
proc sort data=DS_HEATMAP ; by descending SORT_GENE SORT_SAMPLE ; run ;

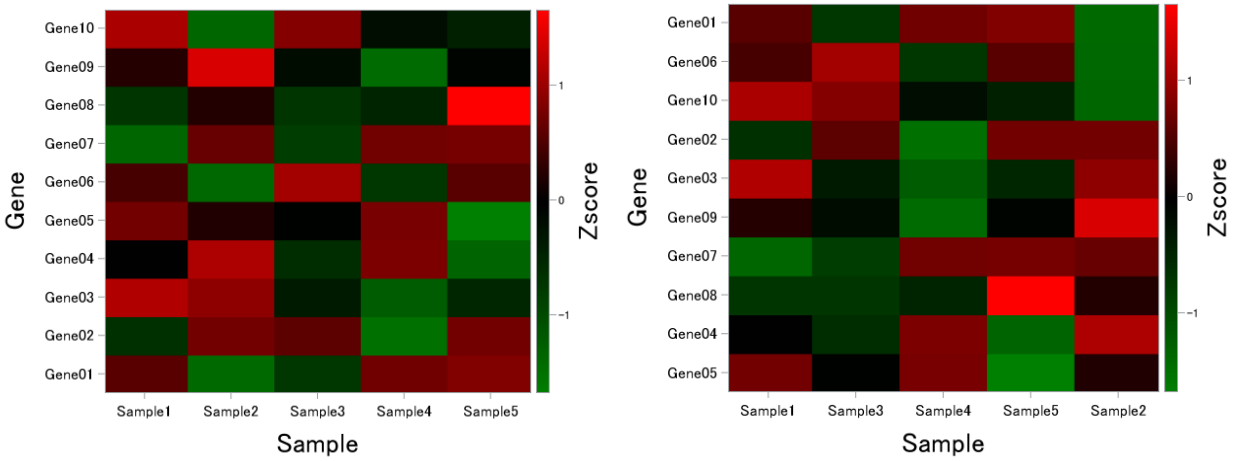
/***** ヒートマップ描画 *****/
ods graphics / imagename='HEATMAP' noborder ;
proc sgplot data=DS_HEATMAP ;
   heatmapparm x=Sample y=Gene colorresponse=zscore / colormodel=(green black red) ;
   gradlegend / titleattrs=(size=20) ;
   xaxis label='Sample' labelattrs=(size=20) valueattrs=(size=12) ;
   yaxis label='Gene' labelattrs=(size=20) valueattrs=(size=12) ;
run ;
```

Program 5　事前処理と SGPLOT プロシジャによるヒートマップ作成



(a)ヒートマップ　　(b)TREE プロシジャによるデンドログラム

Figure 3　シンプルな状況下での階層型クラスタリング例



(a)クラスタリング前　　(b)両軸でのクラスタリング後

Figure 4　全データに対してサンプル・遺伝子ともにクラスタリングを行ったヒートマップ

3.5 TEMPLATE プロシジャ／SGRENDER プロシジャ

SAS には、グラフレイアウトを支援する Graph Template Language (GTL)が備わっており、様々な機能やプロシジャが存在する[3][18]。Program 6 に示す TEMPLATE プロシジャにより Figure 5 のようにテンプレートの枠組みを取り決め、SGRENDER プロシジャによりテンプレートを読み込んだグラフを作成することをセットとし、Figure 6 のように 1 つのヒートマップと 2 つのデンドログラムを結合したグラフを出力した。

TEMPLATE プロシジャでは、Figure 5 のように各グラフのレイアウトを設定する。画像位置の座標(x, y)はその画像の中心にあたる。width, height は単位無しで、座標(x, y)は左下(0, 0), 右上(100, 100)として、割合にしておくことで、全体の画像サイズに関わらず見積もりやすい。一方で、ODS GRAPHICS 設定では単位無しでは pixel であるが、環境により変動があるため、こちらは inch, cm のようにした方が定量的である。なお、画像サイズが小さいと解像度、デンドログラムの表示が崩れたりすることがある。layout ステートメントは種類が多数存在し、layout lattice のようなグラフ行列・分割レイアウトを使用する方法もあるが、それぞれ異な

るタイプのグラフで、規則的な位置・サイズでもないため、layout overlay ステートメントを使用し、直接座標・サイズを指定する。今回出力するのは画像ファイルのみでデータセットは無いが、SGRENDER プロシジャでは原則 data=オプションでデータセット（1 オブザベーション以上）を読み込む必要があるため、ダミーのデータセットを作成して読み込む必要がある。

デンドログラムは、縦横では異なるため、別々の比でサイズを変更してもとくに問題無い。ヒートマップとデンドログラムは直接的には紐付いていないため、カテゴリー数が変われば、細部の調整が必要になる部分もある。ヒートマップに対してデンドログラムの両端を揃える。

```
/*** グラフテンプレート ***/
proc template ;
   define statgraph JOINED_GRAPH ;
      begingraph ;
         layout overlay ;
            /* ヒートマップ */
            drawimage '…¥HEATMAP_CLUSTER.png' / x=40 y=40 width=65 height=65 ;
            /* デンドログラム（上側） */
            drawimage '…¥DENDROGRAM_X.png' / x=42.5 y=80 width=40 height=15 ;
            /* デンドログラム（右側） */
            drawimage '…¥DENDROGRAM_Y.png' / x=85 y=45 width=20 height=52.5 ;
         endlayout ;
      endgraph ;
   end ;
run ;

/* ダミーデータセット作成 */
data DUMMY ;
   DUMMYVAR = . ;
run ;

/*** グラフアウトプット ***/
ods graphics on / width=20in height=20in ;
proc sgrender data=DUMMY template=JOINED_GRAPH ;
run ;
```

Program 6　TEMPLATE プロシジャによる座標設定例と SGRENDER プロシジャによる出力

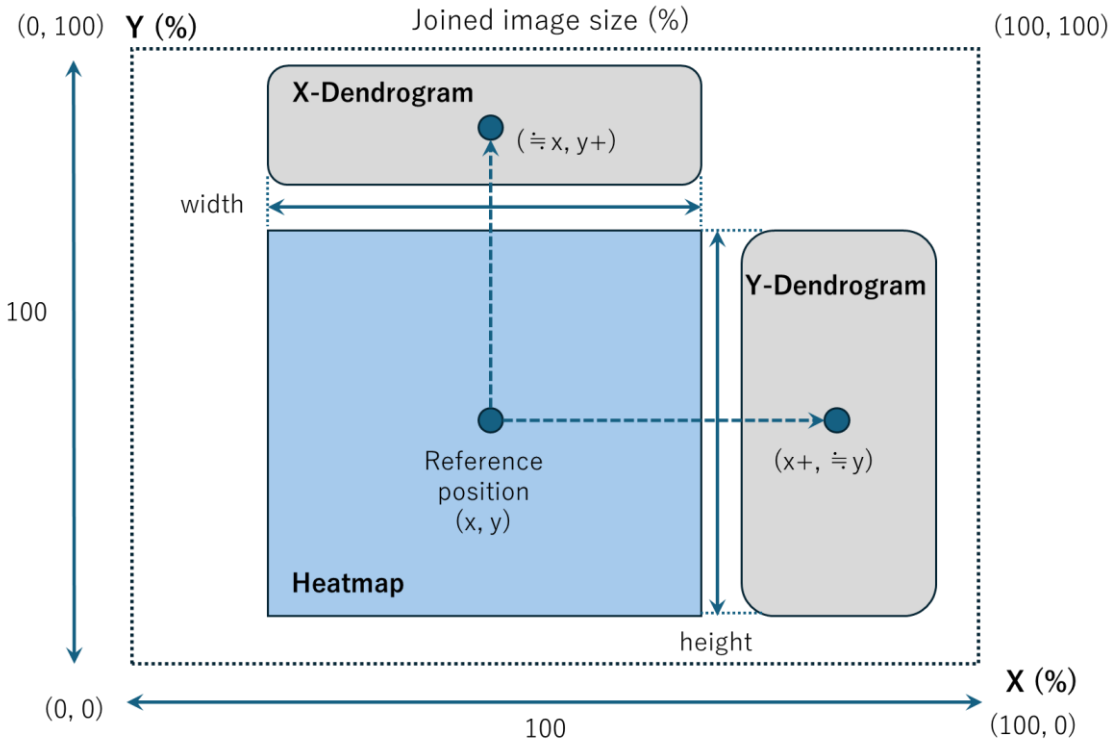


Figure 5　TEMPLATE プロシジャによる各画像出力のレイアウト例

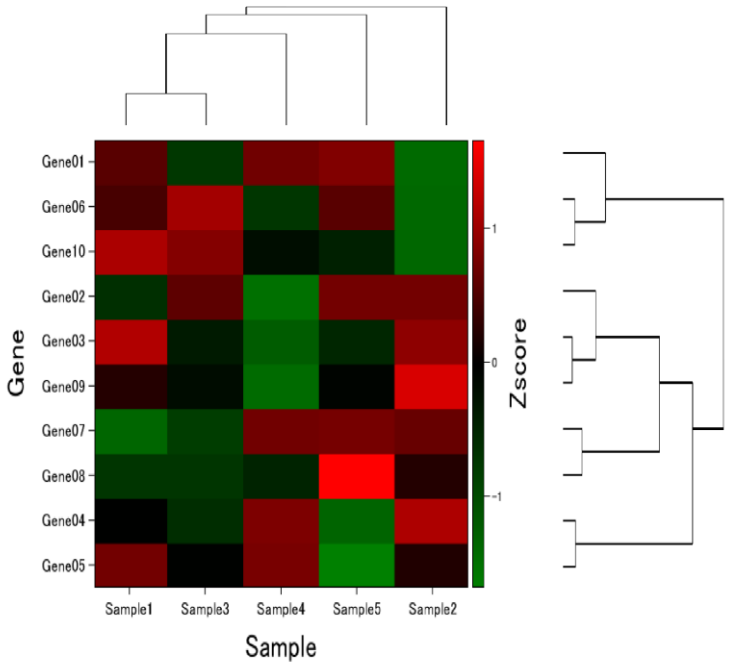


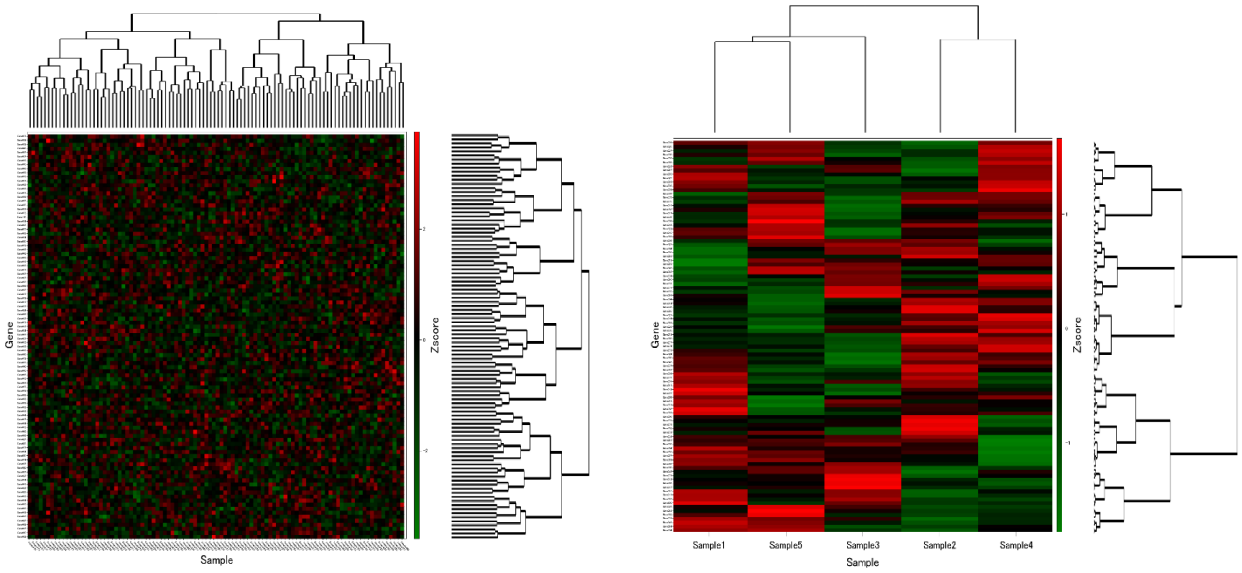
Figure 6　階層型クラスタリングを実装したヒートマップの結合グラフ

4. その他条件下での比較

4.1 カテゴリー数

実際はとくに一度に解析する遺伝子数が多いことが想定されるため､両方または片方のカテゴリー数を 100 に増やして、同様にランダムな一様分布でテストデータを生成した。このデータについても SGPLOT プロシジャでデンドログラム描画可能であった。カテゴリー数が多くなるほど、TEMPLATE プロシジャでの座標は細部の調整が必要になる。ヒートマップの目盛とデンドログラムの葉は等間隔であるため、基本的に両端を揃えれば良い。

ヒートマップのカテゴリー数は多いほど情報量があるが、視認性・解釈性が上がるとは限らず、初期解析用には、絞った方が明確なパターンを見つけやすい。Figure 7 の(a)デンドログラムから、初回のクラスター形成時点で高さ（分散増加・距離）が大きくなり、類似度の解析が困難になっていることが分かる。(b)のように片方のサンプル数だけでも絞ることで、集合が明瞭になりパターン化しやすくなる。



(a)横軸・縦軸ともにカテゴリー数 100 に増加　　(b)片方の縦軸のみカテゴリー数 100 に増加

Figure 7　カテゴリーを増やしたデンドログラム及びヒートマップ

4.2 プログラミング言語

データ分析・解析に用いられるプログラミング言語として、他には R, Python が著名で、遺伝子解析おいてはそれらが主流である。R では Base R に heatmap 関数が用意されており、Program 7 により Figure 8 を作成した。Python では seaborn ライブラリの成果物が公開されており[19]、Program 8 により Figure 9 を作成した。R, Python ともに、テストデータのデータフレームのワイド形式のままスケーリングの向き（行・列）は設定可能で、クラスタリング方法にウォード法等を指定することで、簡潔なコードで、スケーリングから、クラスタリング・デンドログラム・ヒートマップ・グラフテンプレートまでを実装できた。

ヒートマップとデンドログラムの並び順や色合いに違いはあってもクラスターの構成要素は同じで、本質

的には変わらないため、再現できているといえる。ヒートマップとデンドログラムがデータとして紐付いているのは大きな特徴である。

```
z_scores <- t(apply(expression_wide, 1, scale))
heatmap(z_scores, col=greenred(256), cexRow=1, cexCol=1, hclustfun=function(x){hclust(x, method='ward')})
```

Program 7　R によるデンドログラム付きヒートマップの作成

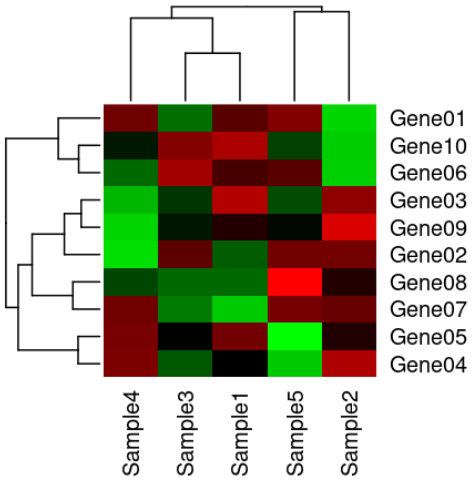


Figure 8　R により Figure 6 と同条件で描いたグラフ

```
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.colors import LinearSegmentedColormap
green_black_red = LinearSegmentedColormap.from_list('gbr', ['green', 'black', 'red'], N=256)
sns.clustermap(expression_wide, method='ward', cmap=green_black_red, z_score=0)
plt.show()
```

Program 8　Python によるデンドログラム付きヒートマップの作成

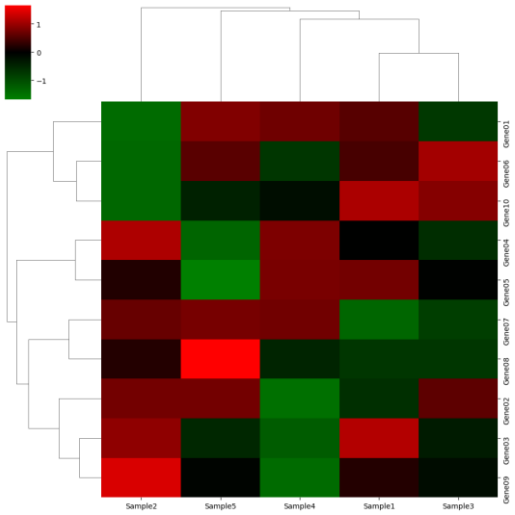


Figure 9　Python により Figure 6 と同条件で描いたグラフ

5. まとめ

階層型クラスタリングを活用したヒートマップを作成する過程として、スケーリング・クラスタリング・デンドログラム・ヒートマップ・グラフレイアウトの一連の手順を精査し、SAS により実装した。プログラミング言語により得意分野があり、R や Python ではパッケージ・ライブラリが充実しているが、SAS でも同様の結果を得ることが可能であった。SAS では随所での設定が必要になり、デンドログラムとヒートマップの連携は想定されておらず、課題が存在する一方で、GTL を利用することにより一定の対応が可能であることを確認できた。精密なグラフ連携を求める場合は、座標の観点から、より高度な実装が必要になる。また、グラフにおいて自由度の高い SGPLOT プロシジャでデンドログラムを実装できた意義は大きい。クラスタリングの実施により、ランダムな状態と比較してデータの傾向把握やパターン抽出が容易となり、データマイニングが容易になった。このようなビジュアライゼーションの技術は、医薬品開発においても広く応用されているため、SAS における実装時に本稿が一助となれば幸いである。

参考文献

[1] 岡田随象、ゼロから実践する遺伝統計学セミナー：疾患とゲノムを結びつける、羊土社、2020 年

[2] 日本バイオインフォマティクス学会、バイオインフォマティクス入門[第 2 版]、慶應義塾大学出版、2023 年

[3] Sanjay Matange, Getting Started with the Graph Template Language in SAS: Examples, Tips, and Techniques for Creating Custom Graphs, 2013

[4] National Center for Biotechnology Information, Gene Expression Omnibus
https://www.ncbi.nlm.nih.gov/geo/

[5] 国立研究開発法人科学技術振興機構、National Bioscience Database Center
https://biosciencedbc.jp/

[6] 永廣卓哉、データ解析入門 5：階層的クラスタリングの基礎、大阪産業技術研究所、2022 年
https://orist.jp/technicalsheet/22-03.pdf

[7] 永廣卓哉、データ解析入門 6：階層的クラスタリングの実践、大阪産業技術研究所、2022 年
https://orist.jp/technicalsheet/22-04.pdf

[8] 眞鍋柊、Python による RNA-Seq (バルク) データ解析 ~階層的クラスタリングとヒートマップの作成~、Qiita、2022 年
https://qiita.com/shyu_manabe/items/35ea921470146483e53e

[9] 合同会社 BxINFO、RNA-Seq 解析におけるヒートマップ、OlvTools、2025 年
https://olvtools.com/rnaseq/help/heatmap

[10] がん情報サイト「オンコロ」、プレシジョンメディシン（Precision Medicine；高精度医療）
https://oncolo.jp/dic/precision-medicine

[11] 稲岡秀検、バイオデータベースとバイオインフォマティクス、計測と制御、53 巻 5 号 p.401-406、2014 年
https://www.jstage.jst.go.jp/article/sicejl/53/5/53_401/_article/-char/ja/

[12] 神嶌敏弘、データマイニング分野のクラスタリング手法、人工知能学会誌、18 巻 1 号 p.59-65、2003 年
https://www.jstage.jst.go.jp/article/jjsai/18/1/18_59/_article/-char/ja
[13] Murtagh, Fionn and Contreras, Pedro. Algorithms for Hierarchical Clustering: An Overview, II. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7 (6), 2017
https://eprints.hud.ac.uk/id/eprint/32552/1/DWD_Fmurtagh_31.pdf
[14] Ulrike von Luxburg, A Tutorial on Spectral Clustering, Statistics and Computing 17(4), 2007
https://arxiv.org/abs/0711.0189
[15] SAS Institute Inc., SAS/STAT 15.2® User's Guide - The STDIZE Procedure, 2020
https://documentation.sas.com/api/docsets/statug/15.2/content/stdize.pdf
[16] SAS Institute Inc., SAS/STAT 15.3® User's Guide - The CLUSTER Procedure, 2025
https://documentation.sas.com/api/docsets/statug/15.3/content/cluster.pdf
[17] SAS Institute Inc., SAS/STAT 15.1® User's Guide - The TREE Procedure, 2018
https://documentation.sas.com/api/collections/pgmsascdc/9.4_3.4/docsets/statug/content/tree.pdf
[18] SAS Institute Inc., SAS® 9.4 ODS Graphics: Procedures Guide, Sixth Edition, 2025
https://documentation.sas.com/api/collections/pgmsascdc/9.4_3.5/docsets/grstatproc/content/grstatproc.pdf
[19] Michael Waskom, seaborn.clustermap, seaborn: statistical data visualization v0.13.2, 2024
https://seaborn.pydata.org/generated/seaborn.clustermap.html
[20] Wikipedia - ヒートマップ
https://ja.wikipedia.org/wiki/ヒートマップ
[21] 株式会社社会情報サービス、BellCurve：統計 WEB：統計学の時間 - 偏差値（Z 得点）
https://bellcurve.jp/statistics/glossary/996.html
[22] 株式会社社会情報サービス、BellCurve：統計 WEB：統計学の時間 - ウォード法
https://bellcurve.jp/statistics/glossary/679.html
[23] 株式会社社会情報サービス、BellCurve：統計 WEB：統計学の時間 - 群平均法
https://bellcurve.jp/statistics/glossary/1277.html

（各 Web サイトへの最終アクセスは 2025 年 8 月）

ベイズ流試験デザインにおける
シミュレーションに基づく動作特性のポイントとその具体例

○吉本 拓矢[1,3]，田中 大樹[2,3]
([1]中外製薬株式会社バイオメトリクス部，[2]鳥居薬品株式会社 調査役，
[3]日本製薬工業協会 医薬品評価委員会 データサイエンス部会)

Key Points of Simulation-Based Operating Characteristics in Bayesian Trial Design and Its Specific Examples

Takuya Yoshimoto[1,3], Hiroki Tanaka[2,3]
[1]Biometrics Dept., Chugai Pharmaceutical Co., Ltd., [2]General Manager, Torii Pharmaceutical Co., Ltd.,
[3]Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association

要旨

ベイズ流試験デザインは，臨床試験の参加者数が少ない場合でも適切な事前分布を設定することで，事後分布を通して治療効果についての妥当な統計的推測が可能となり，Complex Innovative Trials Designsを中心に複雑化する臨床試験デザインにおいて注目されている。また，事前分布を積極的に活用せずとも，無情報事前分布を用いた場合において事後分布を用いた治療効果の定量的評価や中間モニタリングによる柔軟な意思決定が可能になる等の利点があり，特に希少疾患における開発において有用なデザインとなり得る。しかしながら，設定する事前分布や意思決定の基準によって誤った結論を導く可能性が否定できない。このため，試験実施前にその性能を慎重に評価する必要がある。本稿では，ベイズ流試験デザインにおけるシミュレーションのポイントを説明し，希少疾患における事例としてThall and Simon (1994)の方法（TSデザイン）の頻度論的動作特性の評価方法を紹介する。また，TSデザインの動作特性を評価するために開発したSASプログラムの特筆すべき処理概要についても説明する。

キーワード：IML プロシジャ，TS デザイン，シミュレーション，ベイズ流試験デザイン

1　はじめに

ベイズ流試験デザインは，臨床試験の参加者数が少ない場合でも適切な事前分布を設定することで，事後分布を通して治療効果についての妥当な統計的推測が可能となる。医療機器に関する FDA ガイダンス「Guidance for the Use of Bayesian Statistics in Medical Device in Clinical Trials」では，具体的なベイズ流試験デザインの利点と課題を以下のように整理している。

利点:

- 事前情報を取り入れることによって現行試験の情報は補強され，その精度が増す可能性がある。
- 用いる事前分布によっては，頻度流デザインと比較して試験に必要な参加者数を減らすことができる可能性がある。
- 無情報事前分布を設定した場合でも，適応的に試験途中での変更に柔軟に対応することができる。

課題:

- 事前情報の確率分布及び関連する統計モデルの設定を事前に決定する必要がある。事後的に事前情報や統計モデルを変更した場合には試験結果の科学的妥当性が説明困難となるため，これらは規制当局と事前に合意形成する必要がある。

以上の特性より，医薬品の開発においてもベイズ流試験デザインを活用することで途中の試験結果に基づく意思決定により試験に柔軟性を持たせることができ，他にも外部の情報を借用する等により効率的な開発の実現が期待される。特に，「希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針」では，「無情報事前分布を用いたとしても，事後分布を用いた治療効果の定量的評価中間モニタリングによる柔軟な意思決定が可能になる点で，ベイズ流アプローチは有用である」と言及している。無情報事前分布を活用することで，頻度流アプローチに基づく第一種の過誤確率を制御し，より柔軟なデザインを採用し得る。一方で，事前分布の設定によっては頻度流アプローチに基づく第一種の過誤確率の厳密な制御が困難であることも課題となり得る。本来，ベイズ流アプローチにおいては伝統的な頻度論的アプローチで重視される第一種の過誤確率の評価を必要としないものの，FDA は規制上の観点から性能を評価する点を重視し，第一種の過誤確率を含む頻度論的動作特性の評価を推奨している。詳細は FDA ガイダンス「Guidance for the Use of Bayesian Statistics in Medical Device in Clinical Trials」を参照されたい。なお，informative な事前分布を設定する場合における第一種の過誤確率の増大に伴う種々の課題に対して，average type I error に基づく新たな評価指標についても提案されている。詳細は Best et al. (2025)を参照されたい。

ベイズ流試験デザインを利用することで効果的な開発を促進できる可能性があるものの，事前分布や意思決定の基準によっては誤った結論を導く可能性が増加し得る。このため，試験実施前にその性能を慎重に評価することが肝要である。本稿は，①ベイズ流試験デザインにおけるシミュレーションのポイントを整理すること，②希少疾患における事例として Thall and Simon (1994a)の方法（TS デザイン）を取り上げ，シンプルな事例としてその動作特性を評価する方法を紹介すること，③TS デザインの動作特性の評価に必要となる SAS を用いたシミュレーションプログラムの特記事項を紹介することを目的とし，シミュレーションの実施にあたり必要となる考え方や技術的な課題への対処法を紹介する。なお，本稿は日本製薬工業協会医薬品評価委員会データサイエンス部会から公表されている報告書「ベイズ流試験デザインにおけるシミュレーションの利用法」（以下，製薬協報告書）の 2 章の内容に基づくものである。また，使用する SAS のバージョンは，SAS/STAT 15.3（SAS 9.4 TS1M8）である。

2　ベイズ流試験デザインにおけるシミュレーションのポイント

FDA のアダプティブデザインに関する企業向けガイダンス「Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry」には，シミュレーションの注意事項が整理されている（日本製薬工業協会

医薬品評価委員会 データサイエンス部会が発出したアダプティブデザインに関する FDA ガイダンスの邦訳も参照されたい)。また，FDA のガイダンス「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products」では，事前分布の選択に対する試験の動作特性の感度を評価することは有益であると言及されている。事前分布の設定によっては誤った結論を導く可能性が高くなるため，事前分布に関する議論は，ベイズ流試験デザインを提案した際に求められる評価において特に重要である。このため，包括的な頻度論的動作特性の結果を提示し，事前分布の設定を含めて採用するデザインの評価能を規制当局に説明する必要がある。本項では，これらのガイダンスを参考に，特に重要であると考えられるシミュレーションのポイント並びに動作特性の評価に必要となるシミュレーションのプロセスを示す。

ベイズ流試験デザインにおけるシミュレーションのポイント

- 第一種の過誤確率におけるシナリオの範囲
  第一種の過誤確率は，妥当なグリッド上（試験を実施する上で想定される様々なシナリオや条件）でシミュレーションを実施し，算出する。
- 第一種の過誤確率におけるシミュレーションの繰り返し数
  シミュレーションにより精度の高い動作特性を得るため，各シナリオで繰り返し回数 100,000 回を設定することが推奨される。各シナリオにおける第一種の過誤確率の 95%の信頼区間の幅が±0.1%に収まることを保証し，第一種の過誤確率の推定値のごくわずかな差を特定できるようにするためである。ただし，細かいグリッド上を探索し，すべてのシナリオで第一種の過誤確率の推定値が目標とする水準を下回っている場合等においては，10,000 回の繰り返し回数で十分である。
- シミュレーションで用いる疑似乱数
  シミュレーション結果が生成されたデータセットに依存することを避けるため，シミュレーションのシナリオごとに異なるシードを用いる。
- 検出力やその他の動作特性の留意事項
  第一種の過誤確率及び検出力だけでなく，その他の動作特性として期待参加者数，期待試験期間，治療効果の推定値のバイアス等も評価し，得られた結果の総体として動作特性を評価する。

動作特性に必要となるシミュレーションのプロセス

検討する試験デザインに応じてデータの生成並びに適用するベイズ流アプローチと成功の定義を変更する必要はあるものの，基本的なシミュレーションのプロセスは以下の流れに基づく。

1. シミュレーション設定に基づき，仮定した確率分布に従う疑似乱数を生成し，1回の臨床試験データを生成する。
2. 1.のデータに対してベイズ流試験デザインを適用し，事後分布に基づく要約統計量並びに試験の成功の有無を評価する。
3. 上記を複数回繰り返し行い，得られた結果に基づいて第一種の過誤確率，検出力及び推定値のバイアス等の動作特性の評価指標を算出する。

3　TS デザインを用いた頻度論的動作特性の評価例

希少癌においてベイズ流解析を用いた事例として，切除不能の明細胞肉腫及び胞巣状軟部肉腫に対するニボルマブの医師主導治験がある（Hirakawa et al., 2018）。本項では，当該治験の適用事例を参考に，2 項で述べたベイズ流試験デザインにおけるシミュレーションの留意点やプロセスに沿って，SAS を用いて TS デザインの動作特性を評価する方法を紹介する。

3.1　TS デザインの数理

TS デザインは，がん領域を対象とした抗腫瘍効果に対する二値アウトカムを評価するベイズ流適応的デザインであり，単群試験を前提に開発された手法である。本稿では，がん領域における奏効の有無を評価項目として取り扱うが，二値アウトカムを対象とした単群試験を想定する場合，特に疾患領域は問わず適用できる。TS デザインは，標準治療と試験治療における奏効割合の事前分布にベータ分布を仮定する。次に，試験治療において単群臨床試験から得られたデータ（尤度）で奏効割合の事後分布を求める。最後に，標準治療の奏効割合の事前分布と試験治療の奏効割合の事後分布を比較する。以降，具体的な数理について説明する。p_Eとp_Sをそれぞれ試験治療と標準治療の奏効割合とし，独立にベータ分布に従うものとする。

$$p_E \sim Beta(\alpha_E, \beta_E), \qquad p_S \sim Beta(\alpha_S, \beta_S)$$

ここに，α_E，β_E，α_S，β_Sは試験計画時に設定する超パラメータである。標準治療に関しても参照する奏効割合の成績の不確実性を考慮して分布を仮定する点に注意する。組み入れ予定の最大参加者数N_{max}例のうち，n例時点で中間モニタリングを実施する場合に，Xをn例の参加者に対する奏効例数とすると，確率変数Xの従う確率分布は二項分布であり，$X \sim Bin(n, p_E)$である。ベータ分布と二項分布の共役性により，$X = x$を得たp_Eの事後分布もまたベータ分布に従い，

$$p_E | x \sim Beta(\alpha_E + x, \beta_E + n - x)$$

となる。また，標準治療に対する試験治療の奏効割合の最低許容増分をδとすると，試験の成功確率は以下のように定義される。

$$\Pr(p_E > p_S + \delta | x)$$

p_Eとp_Sはそれぞれ独立であるから，p_Eの事後分布とp_Sの事前分布に関する同時分布の確率密度関数は

$$f(p_E; \alpha_E + x, \beta_E + n - x) f(p_S; \alpha_S, \beta_S)$$

となる。ただし，$f(p; \alpha, \beta)$は$Beta(\alpha, \beta)$の確率密度関数を表す。また，固定したあるp_Sに対して$p_E > p_S + \delta$となる確率は，

$$\int_{p_S+\delta}^{1} f(p_E; \alpha_E + x, \beta_E + n - x) f(p_S; \alpha_S, \beta_S) dp_E = [1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)] f(p_S; \alpha_S, \beta_S)$$

となる。ただし，$F(p; \alpha, \beta) = \int_0^p f(x; \alpha, \beta) dx$である。最後に，$p_S$に対して周辺化することで以下を得る。

$$\Pr(p_E > p_S + \delta | x) = \int_0^{1-\delta} [1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)] f(p_S; \alpha_S, \beta_S) dp_S$$

以上の数理をまとめた TS デザインのシェーマを図1に示す。

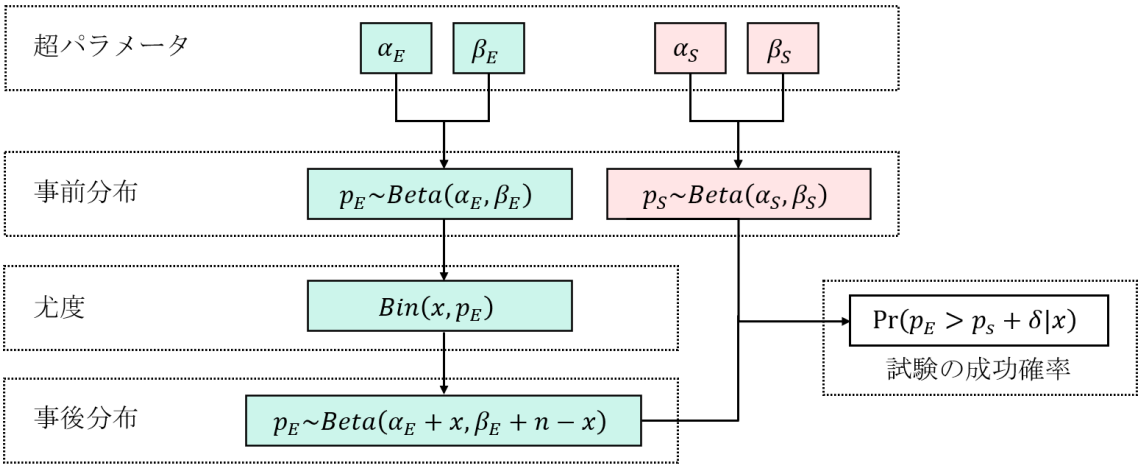


図1　TS デザインのシェーマ

θ_Uとθ_Lをそれぞれ事前に規定した事後確率の目標値とし，以下のルールに基づいてn例時点の境界奏効例数を設定する。

- $\Pr(p_E > p_S + \delta | x) \geq \theta_U$を満たす奏効例数$x$の最小の整数を$U_n$とする。
- $\Pr(p_E > p_S + \delta | x) \leq \theta_L$を満たす奏効例数$x$の最大の整数を$L_n$とする。

中間モニタリング時点での中止規則を以下のように記述できる。この操作を予定最小例数（N_{min}）からN_{max}まで評価を継続して意思決定する。

- $x \geq U_n$の場合，試験治療は有望であると判断し，有効中止する。
- $x \leq L_n$の場合，試験治療は有望でないと判断し，無益性中止する。
- $L_n < x < U_n$かつ$n < N_{max}$の場合，試験を継続して追加で参加者を組み入れる。

3.2　意思決定基準

本項では，3.1項の理論に基づき設定される意思決定基準を算出することを目的とする。デザインパラメータは以下のとおりとする。

- 試験治療の奏効割合（p_E）の事前分布：$Beta(0.5, 0.5)$
- 標準治療の奏効割合（p_S）の事前分布：$Beta(34.4, 137.6)$
- 予定登録例数（N_{max}）　　：15例

- 予定最小例数（N_{min}）　　：10例
- 最低許容増分（δ）　　：0

標準治療の奏効割合の事前分布は Thall and Simon (1994b)を参考に，平均が0.2で，かつベータ分布の5%点と95%点の幅が0.1となるような超パラメータを選択した。試験治療の奏効割合の事前分布は Jeffreys の無情報事前分布を採用した。事後確率の目標値はθ_Lを0%，θ_Uを95%とした。すなわち，無益性中止は考慮せず，有効中止のみを有するデザインを想定する。以上の設定の下，試験治療の奏効割合の事後分布（p_E）が標準治療の奏効割合の事前分布（p_S）を超える確率が95%以上となる必要最小奏効例数を表1に示す。例えば，10例時点での奏効例数が5例以上であれば試験治療は有効と判断し，試験を中止する。

表1　登録例数とその必要最小奏効例数

登録例数	10例	11例	12例	13例	14例	15例
必要最小奏効例数	5例	5例	5例	6例	6例	6例

以上の意思決定基準を算出する SAS のマクロプログラムを以下に示す。IML プロシジャを用いて成功確率$\Pr(p_E > p_S + \delta|x)$を計算し，当該マクロを用いて有効中止基準を満たす最小奏効例数をグリッドサーチにより探索する。

```
%macro simon(aS, bS, aE, bE, d0, y, n);
proc iml;
aS = &aS.;
bS = &bS.;
aE = &aE.;
bE = &bE.;
d0 = &d0.;
delta = 1 - d0;

y = &y.;
n = &n.;

start f(x) global(aS, bS, aE, bE, d0, n, y, delta);
                first = 1 - cdf('Beta', X+d0, aE+y, bE+n-y);
                second = pdf('Beta', X, aS, bS);
                fx = first * second;
                return(fx);
        finish;
```

```
        delta = 1 - d0;
        call quad(result, "f", 0 || delta);/*関数fを積分*/
        print result;

create result;
append;
quit;
%mend;
```

3.3　頻度論的動作特性

3.2項で示したデザインの動作特性を評価する。具体的には，コンピュータシミュレーションを用いて帰無仮説及び対立仮説下での頻度論的動作特性を評価する。以下に動作特性に必要となる設定を示す。

- 閾値奏効割合の真値　　：20%
- 期待奏効割合の真値　　：50%
- シミュレーション回数　　：100,000回
- 評価のタイミング　　：1例ごと（すなわち，意思決定回数は最大6回）
- 名目有意水準　　：10%以内
- 目標検出力　　：80%以上

以上の設定の下で，指定したシミュレーション回数の臨床試験を仮想的に実施し，得られた結果に基づいて第一種の過誤確率，検出力及び期待参加者数を評価する。特に，第一種の過誤確率は帰無仮説下で試験治療を有効と判断する確率，検出力は対立仮説下で試験治療を有効と判断する確率と定義する。表2に頻度論的動作特性の結果を示す。

表2　TS デザインの頻度論的動作特性の結果

帰無仮説		対立仮説	
平均参加者数	第一種の過誤確率	平均参加者数	検出力
14.69例	8.85%	11.20例	87.26%

シミュレーションに基づく第一種の過誤確率は8.85%であり，名目有意水準である10%を下回ることが確認できる。また，検出力も80%を超える十分高い確率を担保できていることから，検討した試験デザインの性能は許容範囲内であると判断できる。なお，実際の検討にあたっては，2項で示したとおり第一種の過誤確率の評価にあたってより広範なシナリオを設定し，実施可能性の観点も考慮して許与される評価能を有するもののうち最適な設定を採用する。本稿ではこのステップを省略し，あるシナリオにおける性能を評価した点に注

意されたい。

4　SAS プログラムの処理概要

シミュレーションに基づく動作特性の評価に関する SAS プログラムは，製薬協成果物を参照されたい。ここでは，3.3項の動作特性を評価するための SAS プログラムの特筆すべき処理概要を紹介する。説明を簡略化するため，仮想的な3試験分のデータを疑似乱数により発生させた場合のシミュレーションプログラムを説明する。各試験で最小組み入れ参加者数は10例，最大組み入れ参加者数は15例とし，ベータ分布の超パラメータ等の設定は3.2項に従う。また，簡単のため，本項のシミュレーションでは動作特性の評価指標として検出力のみを算出することを目的とする。以降，以下のプロセスに分けて説明する。

1. 疑似乱数に基づくデータ生成
2. TS デザインによる試験の成功確率の算出
3. 有効中止となった試験のフラグ生成と検出力の算出

1. 疑似乱数に基づくデータ生成

```
%let num = 3;/*studyの数*/
%let nmin = 10;/*最小組み入れ参加者数*/
%let nmax = 15;/*最大組み入れ参加者数*/

data D1;
length study 3 n 3 AVAL 3 sum 3;
    call streaminit(123456);
    retain sum 0 n 0;
    do study = 1 to &num.;
     do i = 1 to &nmax.;
                AVAL = rand('BERNOULLI', 0.5);
                sum = sum + AVAL;
                n = n + 1;
                if i >= &nmin. then output;
      end;
      sum = 0; n = 0;
    end;
    drop i;
run;
```

3試験分の仮想データを疑似乱数により生成するための手順を以下に示す。

Step 1.　rand 関数を用いて，真値が0.5（期待奏効割合の真値）のベルヌーイ分布に基づく疑似乱数により，試験ごとに奏効例の場合に1, 未奏効例の場合に0を取る変数 AVAL を生成する（計45オブザベーション）。

Step 2.　retain ステートメントを用いてオブザベーションごとの累積奏効例数を取得し，結果を変数 sum に格納する。

Step 3.　試験ごとに，10オブザベーション（最小組み入れ参加者数に相当）から15オブザベーション（最大組み入れ参加者数）のデータを抽出する。生成された SAS データセット D1を図2に示す。変数 study は試験 ID，n は組み入れた参加者数，AVAL は奏効例，sum は累積奏効例数を表す。例えばオブザベーション1は，study 1において10例目の評価が完了した時点で計4例が奏効したことを意味する。

	study	n	AVAL	sum
1	1	10	1	4
2	1	11	1	5
3	1	12	1	6
4	1	13	1	7
5	1	14	1	8
6	1	15	1	9
7	2	10	0	3
8	2	11	0	3
9	2	12	1	4
10	2	13	1	5
11	2	14	1	6
12	2	15	0	6
13	3	10	0	4
14	3	11	1	5
15	3	12	0	5
16	3	13	1	6
17	3	14	1	7
18	3	15	0	7

図2　生成した3試験分の仮想データ

2.　TS デザインによる試験の成功確率の算出

```
proc iml;
  aE = 0.5; bE = 0.5; aS = 34.4; bS = 137.6; d0 = 0;

  /*データセットの読み込み*/
  use D1;
  read all var {study sum n} into input;/*18行×3列を行列「input」に格納*/
  close D1;
```

```
  /* 関数の定義 */
  start f(x) global(aS, bS, aE, bE, d0, n, y, delta);
        first = 1 - cdf('Beta', X+d0, aE+y, bE+n-y);
        second = pdf('Beta', X, aS,bS);
        fx = first * second;
        return(fx);
  finish;
  delta = 1 - d0;

  /*IMLプロシジャ内にてDOループ */
  do i = 1 to nrow(input);
    y = input[i,2];
    n = input[i,3];
      call quad(result, "f", 0 || delta);/*[//]: 結合演算子 (縦) */
      t_results = t_results//result;
  end;

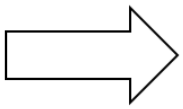
  /* データセットに出力 */
  output = input || t_results;/*[||]: 結合演算子 (横) */
  cname = {"study" "sum" "n" "result"};
  create outdata from output [colname=cname];
  append from output;
  close outdata;
quit;
```

IML プロシジャを用いて，TS デザインに基づく試験の成功確率を算出する。SAS データセットの読み込み，被積分関数の定義及びその積分計算はすべて IML プロシジャ内で実行している点に注目したい。IML プロシジャ内で積分計算を繰り返し処理させることで，関数の定義及び IML プロシジャの実行を1回のみとし，高速な処理を実現可能としている。以下の3つの手順に分けて，プログラムの詳細を説明する。

Step 1.　use ステートメントと read ステートメントを使用して，IML 内に1.で作成した SAS データセット D1 を18行3列の行列に変換して読み込み，それを行列 input に格納する。var オプションで読み込む変数を指定し，ALL オプションにより全オブザベーションを読み込んでいる。変換された行列の1列目は study，2列目は sum，3列目は n にそれぞれ対応している。行列への変換のイメージを図3に示す。続いて，プロシジャ内で試験の成功確率$\Pr(p_E > p_S + \delta|x)$の被積分関数を計算するための関数$f$を定義する。

SASデータセット

Study	n	AVAL	sum
1	10	1	4
1	11	1	5
1	12	1	6
1	13	1	7
1	14	1	8
1	15	1	9
2	10	0	5
2	11	0	5
2	12	1	6
2	13	1	7
2	14	1	8
2	15	1	9
3	10	1	7
3	11	1	8
3	12	0	8
3	13	1	9
3	14	0	10
3	15	0	10



- SASデータセットをIMLの行列として読み込む
- 変数studyがCOL1に，変数SUMがCOL2に，変数nがCOL3に格納される

IML中の行列input（18行×3列）

COL1 (study)	COL2 (sum)	COL3 (n)
1	4	10
1	5	11
1	6	12
1	7	13
1	8	14
1	9	15
2	5	10
2	5	11
2	6	12
2	7	13
2	8	14
2	9	15
3	7	10
3	8	11
3	8	12
3	9	13
3	10	14
3	10	15

図3　SAS データセットから行列への変換

IML プロシジャでサポートされている start ステートメント及び finish ステートメントを用いることで関数の定義が可能となる。ここで定義する関数fは，左累積分布関数の計算結果を返す CDF 関数及び確率密度関数を返す PDF 関数を用いて定義している。start-finish ステートメント内の「first = 1 - cdf('Beta', X+d0, aE+y, bE+n-y);」は以下に相当し，

$$[1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)]$$

「second = pdf('Beta', X, aS, bS);」は，以下に相当する。

$$f(p_S; \alpha_S, \beta_S)$$

すなわち，「fx = first * second;」は，以下に相当する。

$$[1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)]f(p_S; \alpha_S, \beta_S)$$

Step 2.　行列 input の各オブザベーションの値（2列目：COL2と3列目：COL3）を Step 1.で定義した関数fの引数とし，関数fを quad 関数により積分計算させることで試験の成功確率$\Pr(p_E > p_S + \delta|x)$を求める。この操作は以下の積分計算に対応する。

$$\int_0^{1-\delta} f(p_S; \alpha_S, \beta_S)[1 - F(p_S + \delta; \alpha_E + x, \beta_E + n - x)]dp_S$$

これを DO ループにより1オブザベーションごとに繰り返し処理する。quad 関数の結果がオブザベーションごとに result として出力されるが，演算子「//」を用いることでオブザベーションごとの result が縦結合され，18行1列の行列が生成される。その行列が t_results に格納される。本手順の処理概要を図4に示す。

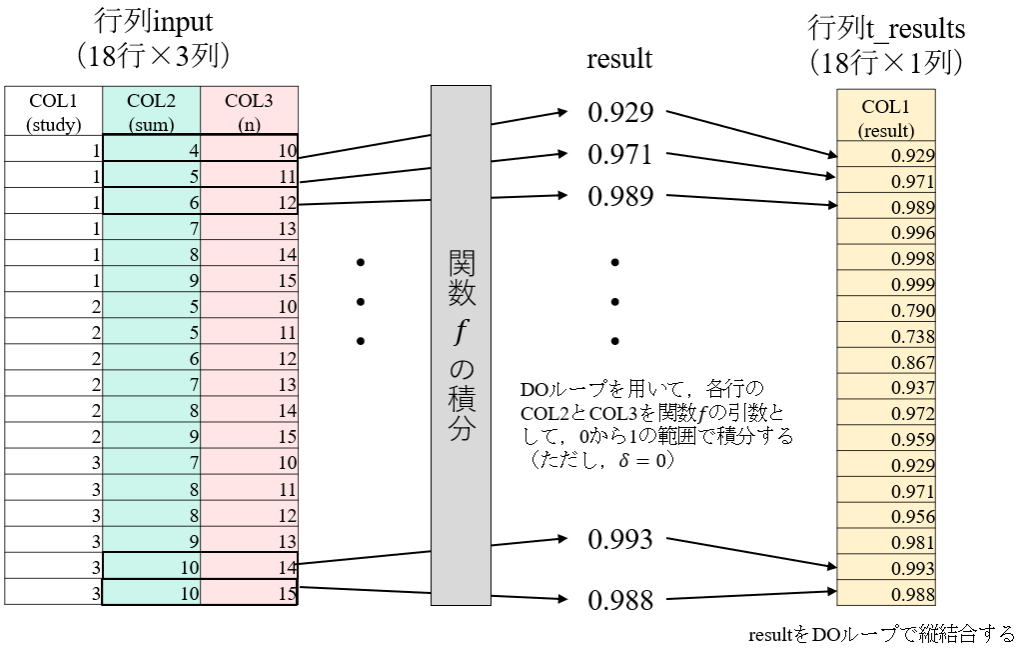


図4　IML プロシジャによる試験の成功確率の算出

Step 3.　行列 input（18行3列）と行列 t_results（18行1列）を横結合演算子「||」で横結合し，18行4列の行列 output を作成する。最後にこの行列を SAS データセット化する。処理概要のイメージを図5に示す。

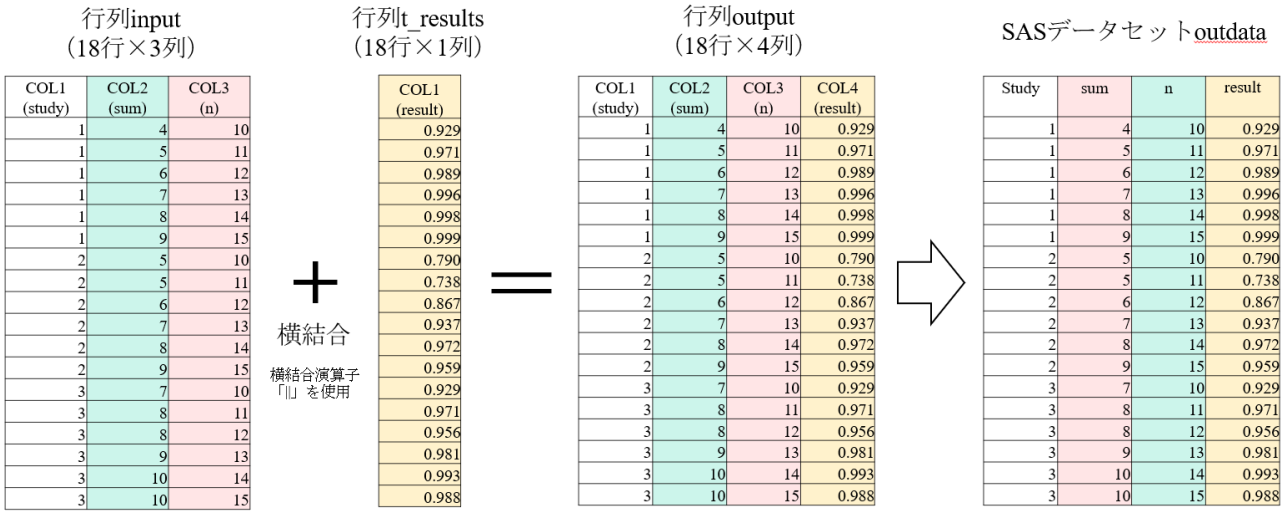


図5　試験が成功する確率を含む SAS データセットの作成

以上のように一つの IML プロシジャ内で SAS データセットの行列変換，関数の定義及び Do ループによる積分計算の繰り返し処理を実行させているため，シミュレーションの処理速度が高い。一方で，シミュレーションの繰り返しの度に IML プロシジャを起動させるアルゴリズムを採用した場合には，処理速度は極めて低下する。

3.　有効中止となった試験のフラグ生成と検出力の算出

```
data outdata2;
   set outdata;
   if RESULT >= 0.95 then flg = 1; ELSE flg = 0;
run;

proc sort data = outdata2; by study; run;
proc means data = outdata2 noprint;
  var flg;
  by study;
  output out = OUT_flg  SUM = SUM;
 run;

data flg;
  set OUT_flg;
  if SUM >=1 then ef = 1; else ef = 0;
run;

proc freq data = flg;
  tables ef/out = Power;
run;
```

5　まとめ

本稿では，ベイズ流試験デザインにおけるシミュレーションに基づく動作特性のポイントに加え，ベイズ流解析を用いた事例として TS デザインを取り上げ，その頻度論的動作特性の評価方法を説明した。試験治療における事前分布は無情報事前分布を用いたが，このような場合でもベイズ流アプローチの柔軟性が享受できるという点で登録が困難な疾患領域において活用の幅は多岐にわたることが期待される。一方，事前分布に情報を持たせる場合は，得られた結果に対するバイアスの程度を追加で評価する必要があるだろう。また，2 項で述べたとおり第一種の過誤確率の評価にはシミュレーションの繰り返し数を 100,000 回等と設定することが望ましいとされるが，複雑なベイズ流解析において事後分布が陽に解けない場合にはマルコフ連鎖モンテカルロ法等により事後分布を近似する必要もある。4 項で紹介したように繰り返して実行する処理を単純化する等，可能な限り計算負荷を最小限に抑える工夫は必至であろう。

製薬協報告書では，SAS に加えて R を用いてシミュレーションを行う方法を提示している。R コードの活用に関心がある際には，製薬協報告書を参照されたい。本稿がベイズ流試験デザインの計画及び SAS による解析の一助になれば幸いである。

謝辞

本稿の細部にわたり有益な助言をいただいた日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2024 年度タスクフォース 3-2 のメンバーに厚く御礼申し上げます。

参考文献

1. Best, N., Ajimi, M., Neuenschwander, B., Hilary, G. S. and Wandel, S. (2025). Beyond the Classical Type I Error: Bayesian Metrics for Bayesian Designs Using Informative Priors. Statistics in Biopharmaceutical Research, 17(2), 183-196.
2. Food and Drug Administration. (2010). Guidance for the Use of Bayesian Statistics in Medical Device in Clinical Trials. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials
3. Food and Drug Administration. (2019). Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry
4. Food and Drug Administration. (2020). Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/interacting-fda-complex-innovative-trial-designs-drugs-and-biological-products
5. Hirakawa, A., Nishikawa, T., Yonemori, K., Shibata, T., Nakamura, K., Ando, M., Ueda, T., Ozaki, T., Tamura, K., Kawai, A. and Fujiwara, Y. (2018). Utility of Bayesian single-arm design in new drug application for rare cancers in Japan: A case study of phase 2 trial for sarcoma. Therapeutic Innovation & Regulatory Science, 52(3), 334-338.
6. Thall, P. F. and Simon, R. (1994a). Practical Bayesian guidelines for phase IIB clinical trials. Biometrics, 50(2), 337-349.
7. Thall, P. F. and Simon, R. (1994b). A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials. Controlled Clinical Trials, 15(6), 463-481.
8. AMED 医薬品等規制調和・評価研究事業（2021）. 希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針. https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_hirakawa1.pdf
9. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会（2021）. アダプティブデザインに関する FDA ガイダンスの邦訳.
10. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会（2025）. ベイズ流試験デザインにおけるシミュレーションの利用法.

MCMCプロシジャによるcommensurate prior法を用いた外部対照データの動的借用法

○吉本 拓矢[1,4]，田中 優一[2,4]，高戸 俊一[3,4]
([1]中外製薬株式会社 バイオメトリクス部，[2]大鵬薬品株式会社 データサイエンス部，
[3]杏林製薬株式会社 開発推進部，
[4]日本製薬工業協会 医薬品評価委員会 データサイエンス部会)

Dynamic Borrowing of External Control Data Using Commensurate Prior Method with MCMC Procedure

Takuya Yoshimoto[1,4], Yuichi Tanaka[2,4], Shunichi Takado[3,4]
[1]Biometrics Dept., Chugai Pharmaceutical Co., Ltd., [2]Data Science Dept., Taiho Pharmaceutical Co., Ltd.,
[3]Clinical Sceience & Affairs, Kyorin Pharmaceutical Co., Ltd.,
[4]Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association

要旨

臨床試験において，外部データを利用することでサンプルサイズの大幅な削減，或いは統計的検出力の向上が期待される。しかしながら，利用する外部データは二次利用を前提とせずに設計・運用されていることも多く，実施する臨床試験の目的に十分に合致しないことがある。そのため，外部データとして利用する対照集団の治療効果が，臨床試験における内部対照集団と十分に類似していない状況が起こり得る。このような状況では，治療効果の推定に無視できない程大きなバイアスが生じ，第一種の過誤確率が上昇する等の問題が生じ得る。この課題に対して，Hobbs et al.（2011）は外部対照データが単一の場合に臨床試験の対照群の類似性をパラメータ化し，外部対照データと臨床試験の対照群の共通性に応じて動的に借用の程度を調整するベイズ流デザインを提案している（commensurate prior法。以下，CMP法）。本稿では，特にパラメトリックな生存時間解析においてベイズ流ワイブル回帰モデルに基づくCMP法を用いた外部データの動的借用法を紹介し，MCMCプロシジャを用いてSASで解析を行う方法を説明する。

キーワード： Commensurate prior 法，MCMC プロシジャ，ベイズ流ワイブル回帰モデル

1　はじめに

臨床試験において，外部データを利用することでサンプルサイズの大幅な削減，或いは統計的検出力の向上が期待される。一方で，利用する外部データは二次利用を前提とせずに設計・運用されていることも多く，実施する臨床試験の目的に十分に合致しないことがある。臨床試験の対照群の一部として利用するデー

タを提供可能な既存試験の条件として，Pocock（1976）は 6 つの条件を提唱している。具体的には，治療，適格基準，評価方法，患者背景，実施体制及びその他の結果の違いを生むと予想される要因の条件がすべて同じであることが求められている。しかしながら，提唱された条件をすべて満たす外部対照データが存在することは現実的にはほとんどない。そこで，すべての条件が厳密に満たされていない状況で外部対照データを活用する有効な方法を考える必要がある（武田ら，2015）。一方で，条件が満たされていない状況では，臨床試験の対照群と外部対照データの間に異質性が生じ，試験間で結果が変動してしまう可能性が否定できない。この異質性に伴う試験間の変動をドリフトと呼ぶ。ドリフトの課題に対して，Hobbs et al.（2011）は，外部対照データが単一の場合に外部対照データと臨床試験の内部対照群の類似性をパラメータ化し，外部対照データと臨床試験の内部対照群の共通性に応じて動的に借用の程度を調整するベイズ流デザインを提案している（commensurate prior 法。以下，CMP 法）。CMP 法は，臨床試験における内部対照群と外部対照群の反応が一貫しているか，どの程度異なるかを確認し，その類似性に応じて外部対照データから借用する情報量を動的に決定して解析に反映させるベイズ流接近法であり，びまん性大細胞型 B 細胞リンパ腫（DLBCL）の臨床試験において CMP 法に基づく外部対照データの活用が検討された。この検討は，Complex Innovative Trial Design（以下，CID）の Pilot program にも採択されている。詳細は，FDA のガイダンス「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products」，日本製薬工業協会 医薬品評価委員会の報告書「Complex Innovative Trial Design の現状と事例の整理」，及び Clark（2025）等を参照されたい。本稿では DLBCL の検討事例を参考に，特にパラメトリックな生存時間解析においてベイズ流ワイブル回帰モデルに基づく CMP 法に焦点を当て，MCMC プロシジャを用いて SAS で当該解析を行う方法を説明する。なお，本稿は日本製薬工業協会医薬品評価委員会データサイエンス部会から公表されている報告書「ベイズ流試験デザインにおけるシミュレーションの利用法」（以下，製薬協報告書）の 3.2.5 章の内容に基づくものである。また，使用する SAS のバージョンは，SAS/STAT 15.3（SAS 9.4 TS1M8）である。

2　本稿で使用するデータと頻度流解析

本稿で想定するデザインのシェーマを図 1 に示す。抗がん剤における time to event 型のアウトカムに対する仮想データを用いる。本データは，以下の設定値の下で生成されたものであり，製薬協報告書の別添より利用可能である（ファイル名：data_CMP.csv）。なお，ドリフト HR の参照群は内部対照群とする。

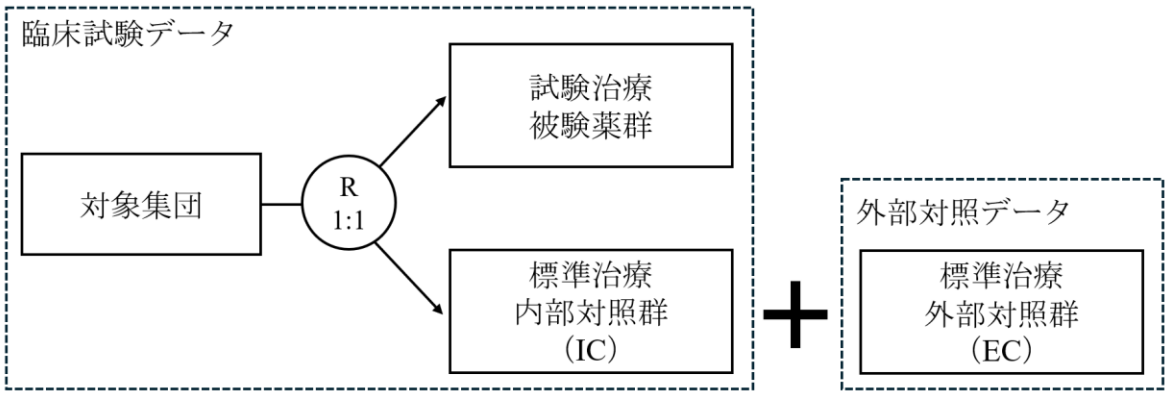


図 1　デザインのシェーマ

- 治療効果の HR　　：0.65
- ドリフト HR　　：1.20
- 臨床試験の登録期間　　：24カ月
- 臨床試験の観察期間　　：36カ月
- 臨床試験の対照群の中央値　：24カ月
- 臨床試験の総被験者数　　：300例
- 外部データの登録期間　　：24カ月
- 外部データの観察期間　　：36カ月
- 外部データの総被験者数　　：100例

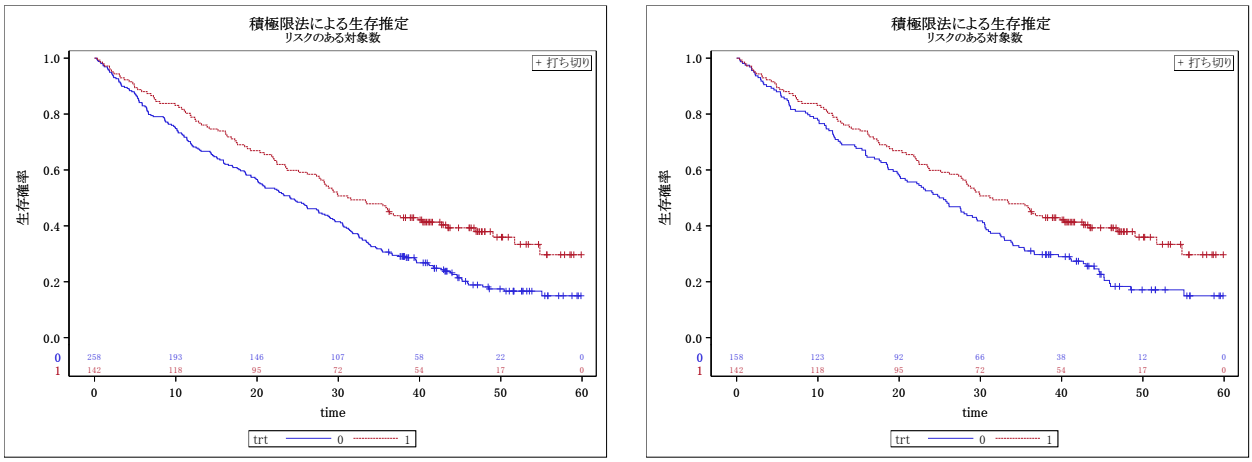
具体的なデータの構造を表 1 に示す。ただし，表 1 はデータセットの一部である。このデータは，被験者番号（patient），試験治療の有無（trt: 試験治療 1，標準治療 0），外部対照か否か（ext: 外部データ 1，臨床試験データ 0），生存時間（time），打切りの有無（cnsr: 打切り 1，イベント 0）の 5 つの変数を有し，内部対照群（IC，Internal Control）及び試験治療群を含む臨床試験データとして 300 例，外部対照群（EC，External Control）として 100 例の計 400 例からなるデータセットである。臨床試験データ及び外部対照データに対して，観察されたイベント数はそれぞれ 214 件及び 81 件である。なお，下線部は参照水準を示す。

表 1　仮想データ

patient	trt	ext	time	cnsr
1	1	0	27.86176	0
2	0	0	20.15954	0
3	1	0	39.81461	0
4	1	0	16.83621	0
5	0	0	36.20025	1

表 1 の仮想データに対して，IC 及び EC の生存時間分布を確認するため，治療効果を示すアウトカムとしてハザード比（HR）を推定する。ベイズ流の CMP 法の適用を想定するため，パラメトリックな手法の適用を考える。具体的には，ワイブル回帰モデルを用いて，IC に対する EC の HR とその 95%信頼区間を推定する。パラメトリックな生存時間解析は，SAS の LIFEREG プロシジャを用いて解析することが可能であり，HR とその 95%信頼区間（95%CI）は 1.095（0.828 - 1.448）と推定された。なお，具体的な推定方法については後述する。

本稿では，CMP 法の利用に先立ち，最も極端な状況，すなわち外部対照のデータを割り引くことなくすべて借用した場合（Full borrowing）と外部対照のデータを借用しない場合（No borrowing）のそれぞれに対して，頻度流のパラメトリックなワイブル回帰モデルに基づく解析結果を評価する。生存時間分布を確認するため，SAS の LIFETEST プロシジャを用いて推定した Kaplan Meier 曲線を図 2 に示す。



(a) Full borrowing　　(b) No borrowing

図 2　Kaplan Meier 法の適用結果

SAS の LIFEREG プロシジャを用いて推定したワイブル分布に基づく HR とその 95%CI は，Full borrowing 及び No borrowing のそれぞれで 0.649（0.506 - 0.833）及び 0.671（0.511 - 0.880）であった。なお，LIFEREG プロシジャは加速モデルの下で parameterization を行っているため，ハザード比スケールに換算する必要がある。ここに，T_iを被験者iの time to event 型アウトカムの確率変数とする。このとき，ワイブル分布に基づくパラメトリック推定として，以下の統計モデルを用いる。

$$S(t|r,\lambda_i) = \exp\{-\lambda_i t^r\},$$
$$f(t|r,\lambda_i) = \lambda_i r t^{r-1} \exp\{-\lambda_i t^r\}$$

ただし，$S(t|r,\lambda_i)$は$\Pr(T_i \geq t)$（生存関数）であり，$f(t|r,\lambda_i)$は死亡密度関数である。また，

$$r = \frac{1}{\sigma}, \quad \lambda_i = \exp\left(-\frac{(\mu + \beta X_i)}{\sigma}\right)$$

であり，X_iは被験者iの共変量（試験治療の有無），βは共変量に対応する回帰係数，μは切片項である。$\hat{\beta}$と$\hat{r}$をそれぞれβとrの推定量とすると，対数ハザード比の推定量は$-\hat{\beta}/\hat{\sigma}$であり，その近似分散の推定量は，次式で与えられる。

$$V\left(-\frac{\hat{\beta}}{\hat{\sigma}}\right) \cong \left(\frac{\hat{\beta}}{\hat{\sigma}}\right)^2 \left\{\frac{V(\hat{\beta})}{\hat{\beta}^2} - \frac{2Cov(\hat{\beta},\hat{\sigma})}{\hat{\beta}\hat{\sigma}} + \frac{V(\hat{\sigma})}{\hat{\sigma}^2}\right\}$$

詳細は，Kevin and Caroll（2003）を参照されたい。なお，LIFEREG プロシジャで推定されるワイブル分布の尺度パラメータは上式のσである点に注意が必要である。以下に，Full borrowing の例に対してハザード比とその 95%CI を算出するための SAS プログラムを示す。ただし，前処理として IMPORT プロシジャを用いて，data_CMP.csv を SAS データセット「data_CMP」に読み込ませておく。

```
/*ソートの指定*/
proc sort data = data_CMP; by descending trt ext; run;

/*ワイブル加速モデルによる推定*/
proc lifereg data = data_CMP order = data;
```

```
        class trt;
        model time * cnsr(1) = trt / dist = weibull CovB;
        ods output ParameterEstimates = estimates_full;
        ods output CovB = cov_full;
run;

/*出力結果のマクロ変数化*/
data _null_; set estimates_full;
        if Parameter = "trt" and Level1 = "1" then call symputx("trt", Estimate);
        if Parameter = "trt" and Level1 = "1" then call symputx("trt_StdErr", StdErr);
        if Parameter = "尺度" then call symputx("scale", Estimate);
        if Parameter = "尺度" then call symputx("scale_StdErr", StdErr);
run;

data _null_; set cov_full;
        if Variable = "trt1" then call symputx("cov_trt_scale", scale);
run;

/*ハザード比とその95%信頼区間の推定*/
data hazard_ratios_full;
        beta = &trt.;
        beta_StdErr = &trt_StdErr.;
        scale = &scale.;
        scale_StdErr = &scale_StdErr.;
        cov_beta_scale = &cov_trt_scale.;

        hazard_ratio = exp(-beta/scale);
        beta_var = beta_StdErr**2;
        scale_var = scale_StdErr**2;
        LnHR_var = (1/scale**2) * (beta_var - (2*beta/scale)*cov_beta_scale +
        (((beta/scale)**2)*scale_var) );
        LnHR_stderr = sqrt(LnHR_var);

        hr_lower_95 = exp(-beta/scale - quantile('NORMAL', .975)*LnHR_stderr);
        hr_upper_95 = exp(-beta/scale + quantile('NORMAL', .975)*LnHR_stderr);
        output;
run;
```

3　ベイズ流の動的借用法

3.1　ベイズ流ワイブル回帰モデルに基づく CMP 法

CMP 法は，1 項で示したとおり外部対照データが単一の場合に臨床試験の対照群の類似性をパラメータ化し，IC と EC の共通性に応じて動的に借用の程度を調整するベイズ流接近法である。借用の程度により得られる結果が変わり得るが，推定値は Full borrowing と No borrowing の幅に収まることが期待される。本項では，IC と EC のアウトカムに対する乖離度に応じて借用程度を調整することを目的として，ベイズ流ワイブル回帰モデルに基づく CMP 法の数理を説明する。なお，数理的な詳細は Hobbs et al.（2011）や野村ら（2022）を，適用事例については Lewis et al.（2019）を参照されたい。いま，外部対照データを合わせた総被験者数をnとし，Z_{1i}を外部対照データか否か，Z_{2i}を試験治療の有無を表すダミー変数とする。比例ハザードモデルでのワイブル分布に基づくパラメトリック推定として，以下の統計モデルを用いる。

$$S(t|r,\lambda_i) = \exp\{-\lambda_i t^r\},$$
$$f(t|r,\lambda_i) = \lambda_i r t^{r-1} \exp\{-\lambda_i t^r\}$$

ただし，

$$\lambda_i = \exp\{\beta_0(1 - Z_{1i}) + \beta_{\text{ext}} Z_{1i} + \beta_{\text{trt}} Z_{2i}\}$$

である。このとき，パラメータの超事前分布，事前分布を以下のように与える。ここでは，事前分布の持つ情報量を少なくするため，分散が大きい値（精度の場合は小さい値）をとるように超パラメータを指定する。

- 超事前分布

$$\beta_{\text{ext}} \sim N(0, 1000), \qquad \tau \sim gamma(1, 0.001)$$

- 事前分布

$$\beta_{\text{trt}} \sim N(0, 1000), \qquad \beta_0 \sim N(\beta_{\text{ext}}, \tau), \qquad r \sim \exp(1)$$

- n個の観測された生存時間$t_1, \ldots, t_n$の尤度（L）

$$L \propto \prod_{i=1}^{n} \{S(t_i|r,\lambda_i)\}^{\delta_i} \{f(t_i|r,\lambda_i)\}^{1-\delta_i}$$

ただし，δ_iは打切りの場合に 1，イベントの場合に 0 を取る指示関数である。

ただし，$\exp(1)$はハザードが 1 の指数分布に従うことを示す。β_{ext}は EC に対する切片項であり，β_0は IC に対する切片項である。超事前分布，事前分布及び尤度の関係を図 3 のように整理できる。事後分布は解析的に解くことができないため，マルコフ連鎖モンテカルロ法（MCMC，Markov chain Monte Carlo method）を用いて事後分布を近似する。τはβ_0の分散であり，その事前分布は commensurate prior と呼ばれ，設定する分布に応じて借用の程度が変化する。このため，設定には慎重な議論が必要であるが，ここでは commensurate prior にガンマ分布を設定する。ガンマ分布は，IC と EC の違いの程度が小さいと仮定した場合に設定されることが多い。一方，IC と EC のアウトカムの違いの程度が大きいと仮定する場合には半コーシー分布を設定することもある。これらの超事前分布の特性については，Gelman（2006）や Hobbs et al.（2012）に詳しい。以降，4 項及び 5 項では commensurate prior としてガンマ分布と半コーシー分布の双方を設定し，解析することを試みる。

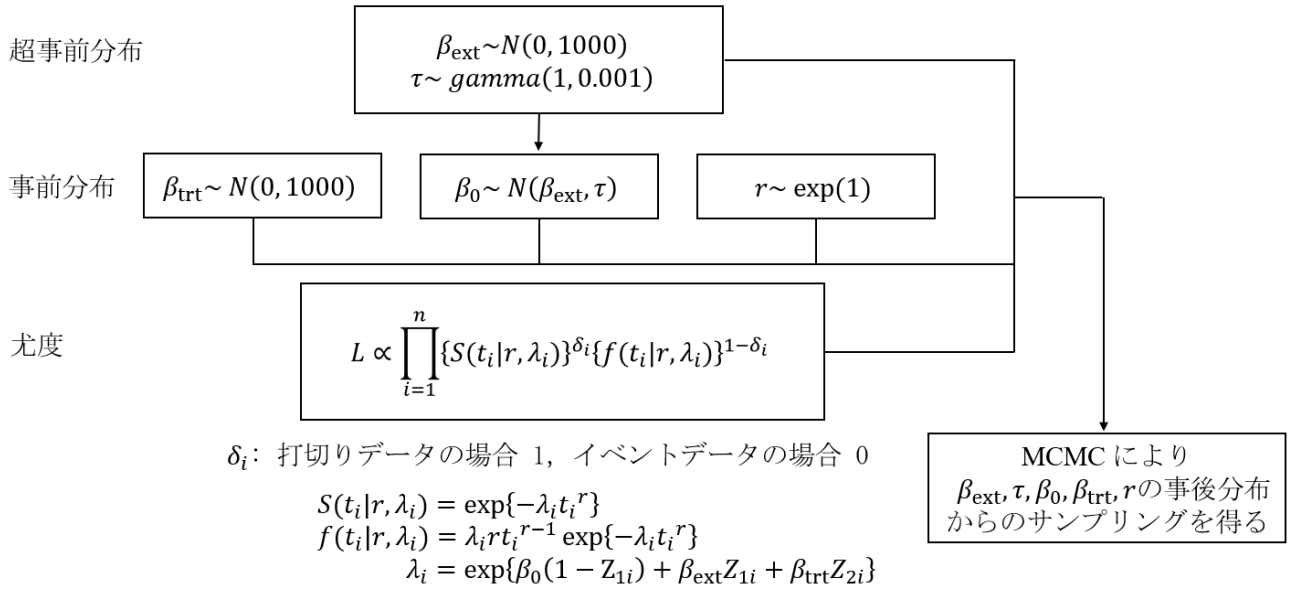


図 3　ベイズ流階層モデルの構造

3.2　Effective historical sample size（EHSS）

事後分布の推測に事前分布がどの程度影響を与えるかを評価することを目的として，Effective historical sample size（EHSS）を評価する。EHSS は，事後 Effective sample size（ESS）からデータの対象者数を差し引く方法である。本稿では対数ハザード比のスケールで EHSS を評価することとし，分散比に基づく事後 ESS を用いて EHSS を算出する。具体的な EHSS の算出式を以下に示す。なお，対照群のベースラインハザードに対して事後 ESS を算出することもしばしば行われる。EHSS はベイズ流の枠組みで評価され，外部データの実際の利用程度を被験者数のスケールで反映する。詳細は Hobbs et al.（2013），Lewis et al.（2019）及び野村ら（2022）を参照されたい。

$$N_{CT}\left(\frac{\mathrm{Prec}(\beta_{\text{trt}}|\text{all data})}{\mathrm{Prec}(\beta_{\text{trt}}|\text{clinical trial data})} - 1\right)$$

ここに，N_{CT}は臨床試験データの被験者数であり，Prec(·)は推定精度を示す（$1/\mathrm{sd}^2$）。EHSS の分子に寄与する$\mathrm{Prec}(\beta_{\text{trt}}|\text{all data})$は，CMP 法により得られた$\beta_{\text{trt}}$の事後分布の推定精度であり，分母に寄与する$\mathrm{Prec}(\beta_{\text{trt}}|\text{clinical trial data})$は，外部対照データを用いずに臨床試験データのみを用いた場合におけるベイズ流ワイブル回帰モデルのβ_{trt}の推定精度である。以降，$\mathrm{Prec}(\beta_{\text{trt}}|\text{clinical trial data})$を計算するための数理について補足する。ここでは，3.1 項と同様に比例ハザードモデルでのワイブル分布に基づくパラメトリック推定を考える。$\lambda_i = \exp\{\beta_0 + \beta_{\text{trt}} Z_{2i}\}$とし，パラメータの事前分布を以下のように設定する。

$$\beta_0 \sim N(0, 1000), \qquad \beta_{\text{trt}} \sim N(0, 1000), \qquad r \sim \exp(1)$$

3.1 項と同様に MCMC により各パラメータの事後分布を導出し，β_{trt}の事後分布に対する推定精度を得る。

本稿では被験者数に基づく EHSS に焦点を当てたが，time to event 型アウトカムにおいては ESS ではなく，イベント数に基づく Effective number of events（ENE）にも関心がある。ENE に関する詳細は，Han et al.（2017），Neuenschwander et al.（2020）及び Roychoudhury and Neuenschwander（2020）等を参照されたい。他にも，expected-local-information-ratio 法（Neuenschwander et al., 2020）等による評価も有用である。これら

の ESS に関する詳細は，野村ら（2022）を参照されたい。なお，本稿における ESS は，MCMC 及び inverse probability of treatment weighting 法の文脈で用いられる Effective sample size とは異なる点に注意されたい。

4　MCMC プロシジャによる CMP 法の解析

本項では，3 項で示したベイズ流ワイブル回帰に基づく CMP 法を SAS の MCMC プロシジャを用いて実装する方法を紹介する。図 3 の構造を直接的に記述することで，事後分布の MCMC サンプリングを容易に得ることができる。以下に SAS プログラムを示す。ただし，前処理として IMPORT プロシジャを用いて，data_CMP.csv を SAS データセット「data_CMP」に読み込ませておく。

```
proc mcmc data = data_CMP outpost = weisurvout nbi = 10000 nmc = 110000 thin = 10
                  seed = 1234 stats = (summary intervals);
   parms beta_ext 0;
   parms tau 1;
   parms beta_trt 0;
   parms beta0 0;
   parms r 1;
   hyperprior beta_ext: ~ normal(0, prec = 0.001);
   hyperprior tau:      ~ gamma(1, is = 0.001);
   prior beta_trt:      ~ normal(0, prec = 0.001);
   prior beta0:         ~ normal(beta_ext, prec = tau);
   prior r:             ~ expon(is = 1);
   lambda = beta0 * (1-ext) + beta_ext * ext + beta_trt * trt;
   mu = (1/exp(lambda))**(1/r);
   llike = (1-cnsr) * logpdf('weibull',time,r,mu) + cnsr * logsdf('weibull',time,r,mu);
   model general(llike);
run;
```

MCMC プロシジャの主なオプションを以下に示す。

- outpost　：事後分布のサンプルを格納する出力データセットの指定。
- nbi　：burn-in の回数。
- nmc　：MCMC のサンプリング回数。
- thin　：間引き（thinning）の回数（指定した値ごとのサンプルのみ保持される）。
- seed　：疑似乱数のシード値。
- stats　：信用区間等の要約統計に関する指定。

次に，MCMC プロシジャで指定するステートメントについて以下に示す。

- parm　：パラメータとその初期値の宣言。

- hyperprior　：超パラメータとその超事前分布の指定。
- prior　：パラメータとその事前分布の指定。
- model　：尤度関数の指定（対数尤度を general 関数により呼び出す）。

上記の SAS プログラムでは，対数尤度の指定にあたって MCMC プロシジャで用意されたワイブル分布の logsdf 関数及び logpdf 関数を利用している。当該関数における parametalization を確認し，適切な尤度を指定する点に注意が必要である。MCMC プロシジャでは，ワイブル分布の死亡密度関数は次式で与えられる。

$$f(t|r,\mu_i) = \exp\left(-\left(\frac{t}{\mu_i}\right)^r\right)\frac{r}{\mu_i}\left(\frac{t}{\mu_i}\right)^{r-1}$$

このため，logsdf 関数及び logpdf の第 4 引数には以下のμ_iを指定することで意図する対数尤度を構成する。なお，対数尤度の計算には，logsdf 関数及び logpdf 関数を利用せず記述することも可能である。

$$\mu_i = \left(\frac{1}{\exp\{\beta_0(1-Z_{1i}) + \beta_{\text{ext}}Z_{1i} + \beta_{\text{trt}}Z_{2i}\}}\right)^{\frac{1}{r}}$$

SAS プログラムを実行して得られる出力結果の一部を図 4 に示す。図 4(a)では，パラメータの事後分布の要約統計量の結果を示している。パラメータは対数ハザード比スケールで出力されているため，ハザード比スケールに換算するには$f(x) = \exp(x)$により変換する必要がある点に注意する。なお，本稿では事後分布に対する区間推定は，事後モードを必ず含み，$100(1-\alpha)$%信用区間の中で最も短い区間である最高事後密度（HPD）区間を用いて評価する（Gelman et al., 2013）。ここに，$1-\alpha$は信頼係数である。Commensurate prior にガンマ分布を指定した CMP 法を適用した結果，標準治療に対する試験治療のハザード比とその 95%HPD 区間は 0.650（0.510 - 0.837）と推定された。また，図 4(b)のトレースプロットにより，MCMC サンプリングに対して自己相関も高くなく安定したサンプリングが得られており，β_{trt}の事後分布は単峰かつ対称な分布であることが確認できる。このような場合には，分位点に基づく信用区間を用いても同様の区間推定の結果を得ることができる。

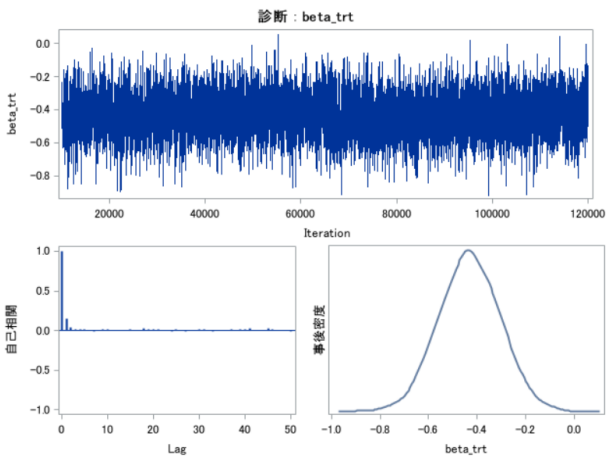
MCMC プロシジャ

事後要約

パラメータ	N	平均	標準偏差	パーセント点 25	50	75
beta_ext	11000	-3.7101	0.2094	-3.8500	-3.7041	-3.5703
tau	11000	1033.6	1004.7	326.0	729.0	1423.5
beta_trt	11000	-0.4311	0.1275	-0.5161	-0.4315	-0.3446
beta0	11000	-3.7198	0.2088	-3.8598	-3.7147	-3.5785
r	11000	1.0767	0.0556	1.0385	1.0752	1.1138

事後間隔

パラメータ	アルファ	両側対称区間		HPD 区間	
beta_ext	0.050	-4.1332	-3.3024	-4.1293	-3.3008
tau	0.050	43.6935	3705.1	1.2040	3015.2
beta_trt	0.050	-0.6806	-0.1829	-0.6735	-0.1780
beta0	0.050	-4.1386	-3.3154	-4.1274	-3.3078
r	0.050	0.9690	1.1879	0.9662	1.1845



(a) 事後分布の要約統計量　　(b) β_{trt}の事後分布及び MCMC のトレースプロット

図 4　MCMC プロシジャの出力（commensurate prior をガンマ分布とした場合）

Commensurate prior に半コーシー分布を指定する場合は，以下の SAS コードでτの指定を置き換える。半コーシー分布を設定した場合及び EHSS の算出のためのベイズ流ワイブル回帰モデルの適用結果をそれぞれ図 5(a)，(b)に示す。

```
hyperprior tau: ~ cauchy(0, 1, lower = 1e-12);
```

事後要約

パラメータ	N	平均	標準偏差	パーセント点 25	50	75
beta_ext	11000	-3.6435	0.2184	-3.7902	-3.6386	-3.4955
tau	11000	11.6138	51.9626	0.9212	2.2200	6.1962
beta_trt	11000	-0.4040	0.1370	-0.4960	-0.4034	-0.3109
beta0	11000	-3.7230	0.2081	-3.8588	-3.7196	-3.5825
r	11000	1.0704	0.0540	1.0331	1.0698	1.1049

(a) MCMC プロシジャの出力
commensurate prior を半コーシー分布とした場合

事後要約

パラメータ	N	平均	標準偏差	パーセント点 25	50	75
beta_trt	11000	-0.4008	0.1424	-0.4953	-0.3999	-0.3061
beta0	11000	-3.8068	0.2471	-3.9714	-3.8001	-3.6398
r	11000	1.0927	0.0658	1.0478	1.0913	1.1368

(b) MCMC プロシジャの出力
ベイズ流ワイブル回帰モデル

図 5　MCMC プロシジャの出力（事後分布の要約統計量）

5　データ解析の結果

表 1 の仮想データに対して，頻度流の Full borrowing 及び No borrowing，commensurate prior にガンマ分布を設定した及び半コーシー分布を設定した場合の結果の概要を表 2 に示す。「No.」列の番号は，外部データの借用の程度の順に対応している。

2 項で示したとおり，仮想事例のデータは IC と EC でアウトカムは比較的類似した傾向を示しており，Full borrowing と No borrowing におけるワイブル分布を仮定したパラメトリックな生存時間解析として，標準治療に対する試験治療の HR とその 95%CI はそれぞれ 0.649（0.506 - 0.833）及び 0.671（0.511 - 0.880）と推定された。また，4 項で示したとおり，commensurate prior にガンマ分布を指定した CMP 法を適用した結果，標準治療に対する試験治療の HR の事後平均とその 95%HPD 区間は 0.650（0.510 - 0.837）と推定された。同様に，commensurate prior に半コーシー分布を設定した CMP 法を適用した結果，HR の事後平均とその 95%HPD 区間は 0.668（0.517 - 0.881）と推定された。

表 2　仮想データに対する解析結果の概要

No.	方法	治療効果に対するハザード比	EHSS
1	Full borrowing	点推定値：0.649 95%CI：(0.506 - 0.833)	- （100に相当）
2	CMP 法（commensurate prior をガンマ分布とした場合）	事後平均：0.650 95%HPD 区間：(0.510 - 0.837)	74.2
3	CMP 法（commensurate prior を半コーシー分布とした場合）	事後平均：0.668 95%HPD 区間：(0.517 - 0.881)	24.1

4	No borrowing	点推定値：0.671 95%CI：(0.511 - 0.880)	- （0に相当）

EHSS: Effective historical sample size, HPD: Highest posterior density（最高事後密度）, CI: Confidence interval（信頼区間）

いずれも推定値は Full borrowing と No borrowing の間に収まり，EHSS より外部対照データを多く借用したことが確認できる。IC と EC にアウトカムで生存時間分布に大きな差異が見られないことから，特に commensurate prior にガンマ分布を指定した場合には Full borrowing と同様の結果が得られた。

6　共変量による調整

3 項で紹介した CMP 法では共変量に治療効果のみを指定したが，その他の重要な予後因子を共変量に含めて解析することも容易に拡張可能である。例えば，ECOG Performance Status（以下，ECOG PS）（0 か 1 か）と臨床病期（IIIb 期か IV 期か）を共変量に加えることを考える。3 項と同様に，λ_iに以下の統計モデルを立て，ベイズ流階層モデルにより各パラメータの事後分布を更新する。この操作により，重要な予後因子である ECOG PS 及び臨床病期の影響を考慮したパラメータβ_{trt}の事後分布を得ることができる。

$$\lambda_i = \exp\{\beta_0(1 - Z_{1i}) + \beta_{\mathrm{ext}}Z_{1i} + \beta_{\mathrm{trt}}Z_{2i} + \gamma_{\mathrm{ps}}X_{1i} + \gamma_{\mathrm{stg}}X_{2i}\}$$

ここに，X_{1i}を ECOG PS を表すバイナリー変数とし，X_{2i}を臨床病期を表すバイナリー変数とする。このとき，回帰係数に対してそれぞれ下記の事前分布を設定する。

・　超事前分布

$$\beta_{\mathrm{ext}} \sim N(0, 1000), \qquad \tau \sim gamma(1, 0.001)$$

・　事前分布

$$\beta_{\mathrm{trt}} \sim N(0, 1000), \qquad \beta_0 \sim N(\beta_{\mathrm{ext}}, \tau), \qquad r \sim \exp(1),$$
$$\gamma_{\mathrm{ps}} \sim N(0, 1000), \qquad \gamma_{\mathrm{stg}} \sim N(0, 1000)$$

事前分布の設定において 3 項との違いは，γ_{ps}及びγ_{stg}を設定する点のみである。以上の構造は，図 6 のように整理できる。同様に，図 6 の構造を直接的に記述することで，SAS の MCMC プロシジャにより容易に事後分布を導出することができる。

7　SAS による簡易シミュレーション

3 項で説明した CMP 法と 4 項で説明した SAS による解析方法を前提として，本項では SAS を用いて帰無仮説及び対立仮説下での簡易的な頻度論的動作特性（以下，動作特性）の評価を試みる。想定するデザインのシェーマは図 1 のとおりとし，2 項で示した仮想データの設定値を踏襲する。ここに，動作特性の評価に必要となる追加の設定を示す。ただし，time to event 型アウトカムの確率分布は指数分布に従うものする。また，簡単のため打切りについて年次脱落等は考慮せず，観察期間に依存した administrative censoring のみを考慮する。なお，簡易的なシミュレーションのため，シミュレーション回数は 1000 回としているが，第一種の過誤確率を評価する上ではシミュレーションの推定精度を向上させるため，より多くの繰り返し数による評価を行うことが望ましい。

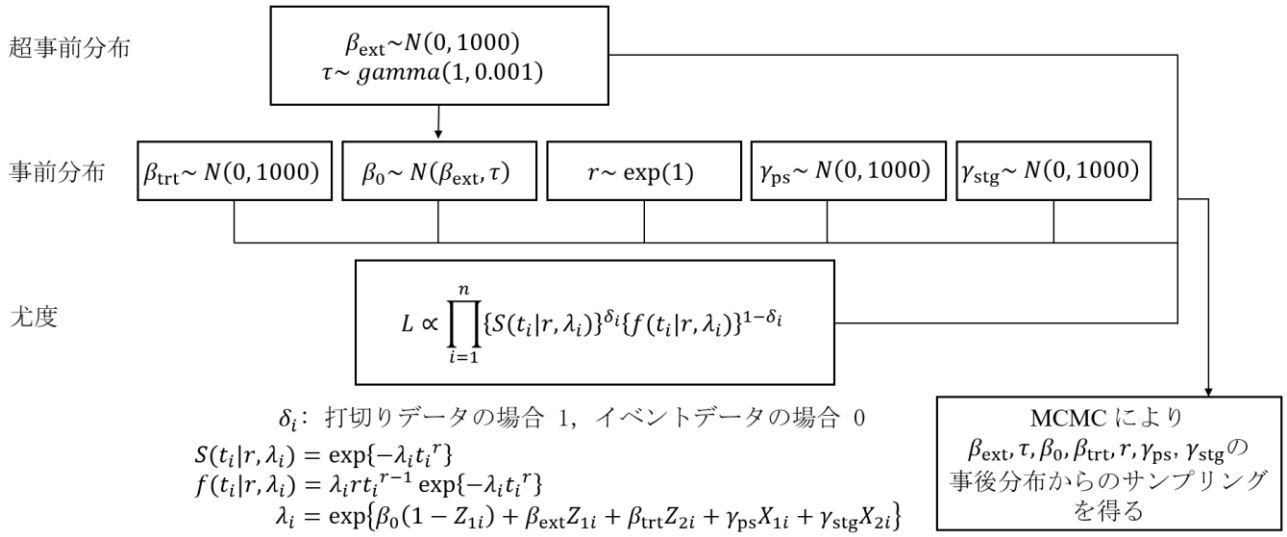


図 6　複数の共変量を含む場合のベイズ流階層モデルの構造

- 治療効果の HR　　：1.00（帰無仮説下）及び0.65（対立仮説下）
- ドリフト HR　　：1.00（ドリフト無）及び1.20（ドリフト有）
- シミュレーション回数　　：1000回

ここで，試験の成功を「対数ハザード比（治療効果）の事後分布に対して，95%HPD 区間の上限が 0 を下回ること」と定義する。以上の設定の下で，指定したシミュレーション回数の臨床試験を仮想的に実施し，得られた結果に基づいて第一種の過誤確率，検出力を評価する。特に，第一種の過誤確率は帰無仮説下で試験治療を有効と判断する確率，検出力は対立仮説下で試験治療を有効と判断する確率と定義する。表 3 に動作特性の結果を示す。ドリフトが生じていない状況では，第一種の過誤確率は 2.5%以内に制御できている一方で，ドリフトが生じている状況では第一種の過誤確率がインフレを起こしている。同様に，ドリフトが生じている状況で相対的に高い検出力を有することが確認できる。

表3　CMP 法の簡易的な動作特性の結果

帰無仮説（第一種の過誤確率）		対立仮説（検出力）	
ドリフト HR		ドリフト HR	
1.00	1.20	1.00	1.20
1.8%	6.8%	91.9%	96.3%

表 3 を得るための SAS のマクロプログラムを以下に示す。プログラムは極めてシンプルに記述でき，以下の手順に従って繰り返し処理を実行している。

1. 疑似乱数を用いて臨床試験データを生成する。
2. 疑似乱数を用いて外部対照データを生成する。
3. 1.と2.のデータを併合し，解析用データセットを生成する。

4. 3.のデータに対して4項で示したMCMCプロシジャによるプログラムを適用し，ods outputステートメントを用いて出力した結果をSASデータセットに格納する。ただし，シミュレーションによる繰り返し処理を考慮し，MCMCの間引きの回数を3にする等の変更を加えている。
5. 4.で生成したSASデータセットに対し,成功の定義を満たすか否かを判断し,結果をフラグに格納する。
6. 5.の結果を頻度集計し，結果を出力する。

```
%let acc_CT = 24;   /*臨床試験の登録期間*/
%let fol_CT = 36;   /*臨床試験の観察期間*/
%let num_CT = 300;  /*臨床試験の被験者数*/
%let median_c = 24; /*臨床試験の対照群の中央値*/
%let acc_EC = 24;   /*外部対照データの登録期間*/
%let fol_EC = 36;   /*外部対照データの観察期間*/
%let num_EC = 100;  /*外部対照データの被験者数*/

%macro CMP_OC(seed, HR, driftHR, sim);
/*1. 臨床試験データの生成*/
data monitor_CT;
        call streaminit(&seed.);
        cnsr = 0;
        ext = 0;
        do sim = 1 to &sim.;
                do n = 1 to &num_CT.;
                        rnd = rand("uniform");
                        time = rand("exponential");
                        dummy = rand("uniform");
                        if 0 <= rnd < 0.5 then do;
                                time = time / (-log(0.5)/(&median_c.));
                                trt = 0; *Internal control arm;
                        end;
                        else do;
                                time = time / (-log(0.5)/(&median_c./&HR.));
                                trt = 1; *Experimental arm;
                        end;
                        c_t = dummy * &acc_CT. + time;
                        output;
                end;
        end;
```

```
run;

proc sort data = monitor_CT; by sim; run;

data monitor_CT; set monitor_CT; by sim;
        if c_t > &acc_CT. + &fol_CT. then do;
                cnsr = 1;
                time = &acc_CT. + &fol_CT. - &acc_CT. * dummy;
        end;
        keep sim trt ext time cnsr;
run;

/*2. 外部対照データの生成*/
data monitor_EC;
        call streaminit(&seed.);
        cnsr = 0;
        ext = 1;
        do sim = 1 to &sim.;
                do n = 1 to &num_EC.;
                        time = rand("exponential");
                        dummy = rand("uniform");
                        time = time / (-log(0.5)/(&median_c./&driftHR.));
                        trt = 0; *External control arm;
                        c_t = dummy * &acc_EC. + time;
                        output;
                end;
        end;
run;

proc sort data = monitor_EC; by sim; run;

data monitor_EC; set monitor_EC; by sim;
        if c_t > &acc_EC. + &fol_EC. then do;
                cnsr = 1;
                time = &acc_EC. + &fol_EC. - &acc_EC. * dummy;
        end;
        keep sim trt ext time cnsr;
run;
```

```
/*3. 臨床試験データと外部対照データの併合*/
data monitor; set monitor_CT monitor_EC; by sim; run;

/*4. MCMCプロシジャによるCMP法の適用*/
ods graphics off;
ods exclude all;
ods noresults;
ods output PostSummaries = post_summary; /*事後分布の要約統計量*/
ods output PostIntervals = post_intervals; /*信用区間*/
proc mcmc data = monitor outpost = weisurvout nbi = 3000 nmc = 33000 thin = 3 seed = 1234
stats = (summary intervals);
        by sim;
        parms beta_ext 0;
        parms tau 1;
        parms beta_trt 0;
        parms beta0 0;
        parms r 1;
        hyperprior beta_ext: ~ normal(0, prec = 0.001);
        hyperprior tau: ~ gamma(1, is = 0.001);
        prior beta_trt: ~ normal(0, prec = 0.001);
        prior beta0: ~ normal(beta_ext, prec = tau);
        prior r: ~ expon(is = 1);
        lambda = beta0 * (1-ext) + beta_ext * ext + beta_trt * trt;
        mu = (1/exp(lambda))**(1/r);
        llike = (1-cnsr) * logpdf('weibull',time,r,mu) + cnsr * logsdf('weibull',time,r,mu);
        model general(llike);
run;
ods output close;
ods graphics on;
ods exclude none;
ods results;

/*5. 成功の定義を満たすか否かを示すフラグの生成*/
data prob_success; set post_intervals;
        if Parameter = "beta_trt";
        if HPDUpper < 0 then flg = 1; else flg = 0;
run;
```

```
/*6. フラグ変数の頻度集計*/
title "HR = &HR., driftHR = &driftHR, sim = &sim.";
proc freq data = prob_success;
        table flg;
run;
%mend;

/*seed          :シード値*/
/*HR            :ハザード比*/
/*driftHR       :ドリフトハザード比*/
/*sim           :シミュレーション回数*/

%CMP_OC(seed = 1, HR = 1.00, driftHR = 1.00, sim = 1000) /*Type I error*/
%CMP_OC(seed = 2, HR = 1.00, driftHR = 1.20, sim = 1000) /*Type I error*/
%CMP_OC(seed = 3, HR = 0.65, driftHR = 1.00, sim = 1000) /*Power*/
%CMP_OC(seed = 4, HR = 0.65, driftHR = 1.20, sim = 1000) /*Power*/
```

8　まとめ

本稿では，最も単純な状況を想定したベイズ流ワイブル回帰モデルに基づく CMP 法の数理を説明し，SAS の MCMC プロシジャを用いた解析方法を紹介した。共変量は治療効果のみを指定したが，6 項で示したとおりその他の重要な予後因子を共変量に含めて解析することも容易に拡張可能である。また，7 項ではシミュレーションにより繰り返し処理を適用することで簡易的な動作特性の評価を試みた。

外部対照データの動的借用法を検証的試験で適用する場合には，CID に関する FDA ガイダンスで言及されているとおり規制当局との合意が必要不可欠であり，検討には多くのシミュレーションの実施が必要となる。このため，既存の解析ツールでは適用できない設定下で試験デザインの動作特性の評価が求められる可能性も否定できず，関心のある試験デザインの動作特性を評価できるよう方法論並びに SAS プログラムに関する技術を整理することは重要であろう。一方，試験デザインの社内検討並びに規制当局との折衝において，効果的かつ迅速に動作特性の評価を行うことも極めて重要である。7 項で示したシミュレーション実験は，簡易的な動作特性の評価であるものの，その実行に数日単位での時間を要する。ただし，実行環境に依存する点には留意されたい。また，動作特性の評価指標は第一種の過誤確率及び検出力のみとしたが，実際には推定値のバイアスや平均二乗誤差等の指標についても評価する必要があろう。加えて，本稿で設定した commensurate prior は，MCMC プロシジャで対応した一般的な分布であるガンマ分布及び半コーシー分布を指定しているが，Spike and slab 分布（Hobbs et al., 2012）等の特殊な分布を設定する場合には，IML プロシジャを用いて解析者が MCMC サンプルを得るためのプログラムを開発する必要もある。動作特性の評価において，想定する重要な予後因子の影響を考慮することも重要である。したがって，本稿で示した概要を整理した上で，既存の解析ツール（例えば psborrow 及び psborrow2 等の R パッケージ）等（Gower-Page,

et al., 2023; Secret, et al., 2024）を積極的に活用し，より効果的かつ包括的にその動作特性を評価することが肝要である。

製薬協報告書では，SAS に加えて R（STAN，JAGS）を用いてベイズ流ワイブル回帰モデルに基づく CMP 法を用いた外部データの動的借用法を適用する方法を提示している。R コードの活用に関心がある際には，製薬協報告書を参照されたい。本稿が CMP 法に基づく動的借用法の計画及び SAS による解析の一助になれば幸いである。

謝辞

本稿の解析プログラムの作成にあたり，ご助言及びご協力いただいた中外製薬株式会社の澤本涼氏，東京理科大学の大東智洋先生に感謝申し上げます。また，本稿の細部にわたり有益な助言をいただいた日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2024 年度タスクフォース 3-2 のメンバーに厚く御礼申し上げます。

参考文献

1. Clark, E. (2025). Use of external data for efficient and robust evidence generation: A case study of dynamic borrowing with external control for overall survival. CDDF Annual Conference 2025.
2. Food and Drug Administration. (2020). Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/interacting-fda-complex-innovative-trial-designs-drugs-and-biological-products
3. Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1, 515-534.
4. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D.B. (2013). Bayesian Data Analysis (3rd ed.). CRC Press.
5. Gower-Page, C., Lu, Y. and Lin, A. (2023). _psborrow: Bayesian Dynamic Borrowing with Propensity Score_. R package version 0.2.1.
6. Han, B., Zhan, J., John, Zhong Z., Liu, D. and Lindborg, S. (2017). Covariate-adjusted borrowing of historical control data in randomized clinical trials. Pharmaceutical Statistics, 16, 296-308.
7. Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047-1056.
8. Hobbs, B. P., Sargent, D. J. and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian Analysis, 7, 639-673.
9. Hobbs, B. P., Carlin, B. P. and Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical

control data. Clinical Trials, 10, 430-440.

10. Kevin, J. and Carroll, M. (2003). On the use and utility of the Weibull model in the analysis of survival data. Controlled Clinical Trials, 24(6), 682-701.
11. Lewis, C. J., Sarkar, S., Zhu, J. and Carlin, B. P. (2019). Borrowing From Historical Control Data in Cancer Drug Development: A Cautionary Tale and Practical Guidelines. Statistics in Biopharmaceutical Research. 11(1), 67-78.
12. Neuenschwander, B., Weber, S., Schmidli, H. and O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. Biometrics, 76(2), 578-587.
13. Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. Journal of Chronic Diseases 29, 175-188.
14. Roychoudhury, S. and Neuenschwander, B. (2020). Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint. Statistics in Medicine, 39(7), 984-995.
15. Secrest, M. and Gravestock, I. (2024). _psborrow2: Bayesian Dynamic Borrowing Analysis and Simulation. https://github.com/Genentech/psborrow2, https://genentech.github.io/psborrow2/index.html.
16. 武田健太朗，大庭真梨，橋爪智行，坂巻顕太郎，田栗正隆，森田智視．（2015）．臨床試験におけるヒストリカルコントロールデータの利用．計量生物学，36(1)，25-50.
17. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会（2023）．Complex Innovative Trial Design の現状と事例の整理.
18. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会（2025）．ベイズ流試験デザインにおけるシミュレーションの利用法.
19. 野村尚吾，大東智洋，澤本涼.（2022）．Hybrid control アプローチを用いるランダム化比較試験の計画と解析：外部データが要約統計量の場合．計量生物学，43(1)，63-96.

Q-Learningに基づく最適な動的治療レジメの推定

○徳田芳稀[1]

([1]エイツーヘルスケア株式会社 データサイエンスソリューション部)

Optimal dynamic treatment regimes estimation based on Q-Learning

Yoshiki Tokuda

Data Science Solutions Department, A2 Healthcare Corporation

要旨

臨床試験において被験者に対する介入は、被験者の状態に依らず試験期間を通して同一であることが多い。しかし実際の診療において医師は、個々の治療結果を踏まえて次回以降の治療選択を動的に行い、これを繰り返すことで最適な治療方針を常に模索している。 近年、データから最適な動的治療レジメ（DTRs: Dynamic Treatment Regimes）を探索する研究が盛んに実施されており、その推定を目的とした逐次多段階割付ランダム化試験（SMART: Sequential Multiple Assignment Randomized Clinical Trials）と呼ばれる試験デザインも提案されている。最適な DTR 推定のための統計手法も多数提案されているが、その中でも手法が簡便で且つ実装が容易であることから Q-Learning が汎用されている。 そこで本稿では、まず DTR 推定の近年の動向や最適な DTR に係る試験デザインを紹介する。次いで Q-Learning に基づく最適な DTR 推定の方法を紹介し、最後にそれらを SAS で実装する方法、出力結果及びその解釈について説明する。実装の例示にあたって、R の DynTxRegime Package に用意されている bmiData を使用した。

キーワード：個別化医療、強化学習、Q-Learning、A-Learning、逐次多段階割付ランダム化試験

1. 緒言

臨床試験は、新規の治療法や医薬品、医療機器などの介入に関する有効性や安全性を確認する目的で実施される。臨床試験における被験者に対する介入は、試験デザインや対象疾患により異なるが、被験者の状態にかかわらず試験期間を通して一定であることが多い。一方、実際の診療では医師が患者の個々の治療結果を踏まえ、次回以降の治療選択を動的に行う。この一連の過程を繰り返すことで、常に最適な動的治療レジメ（DTRs: Dynamic Treatment Regimes）を模索している。[1]

近年、臨床試験や観察研究のデータから最適な DTRs を探索する研究が盛んに行われており、逐次多段階割付ランダム化試験（SMART: Sequential Multiple Assignment Randomized Trials）と呼ばれる、最適な DTR 推定に適した試験デザインも提案されている。[2] 最適な DTR 推定に関する文献について、表 1.で示した簡単な検索式を用いて PubMed で文献検索を実施したところ、2025 年 9 月 1 日時点で 642 件の文献が抽出された。

表 1. DTR 推定に関連する文献の検索式

("dynamic treatment regime"[Title/Abstract] OR "dynamic treatment regimes"[Title/Abstract] OR "dynamic treatment strategy"[Title/Abstract] OR "dynamic treatment strategies"[Title/Abstract] OR "adaptive treatment strategy"[Title/Abstract] OR "adaptive treatment strategies"[Title/Abstract] OR "adaptive treatment regime"[Title/Abstract] OR "adaptive treatment regimes"[Title/Abstract] OR "sequential multiple assignment randomized trial"[Title/Abstract] OR "sequential multiple assignment randomized trials"[Title/Abstract])

抽出された文献を出版年ごとに集計し、図 1.に示した。

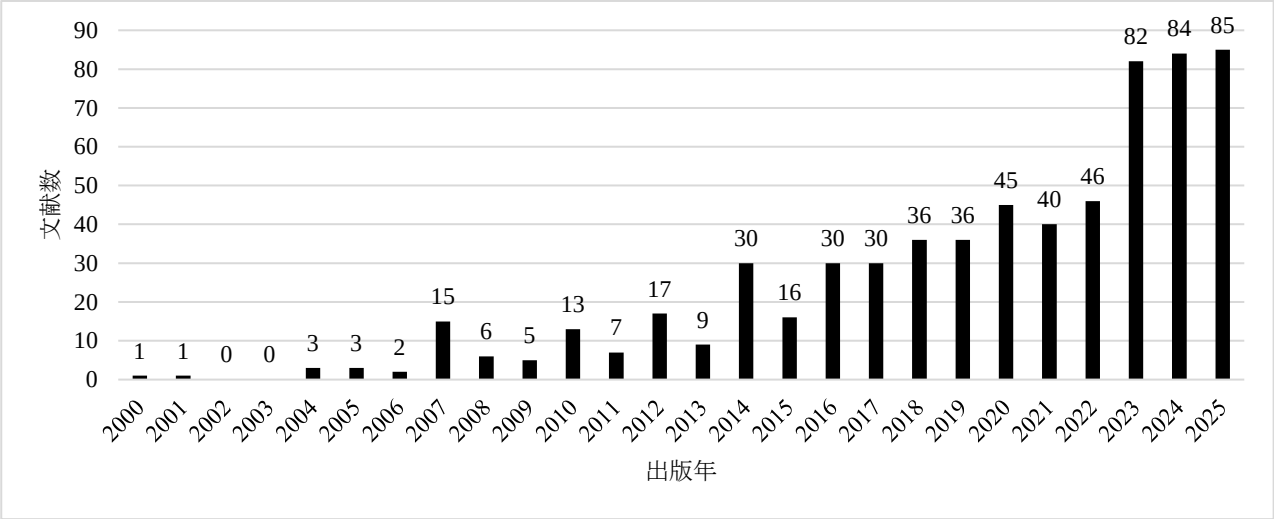


図 1. DTR 推定に関連する文献の出版推移のヒストグラム(2025 年 9 月 1 日時点)

文献数に着目すると、最適な DTR 推定に関連する文献は近年特に増加傾向にあり、関心度の高いトピックであることが推察される。抽出された文献の内容について、最適な DTR 推定を目的とした内容だけではなく、DTR 推定のための統計手法を提案している文献も含まれている。それらの統計手法の中でも、手法が簡便で且つ実装が容易な Q-Learning が汎用されている。今回抽出した文献について、タイトル及びアブストラクトに“Q-Learning”の記載がある文献を確認したところ、49 文献が該当した。また Whiston らは、Physical Activities の分野で実施された SMARTs についてシステマティックレビューを実施し、その中で DTR 推定に Q-Learning が最も使用されていたと報告している。[3]

Q-Learning の実施にあたって、R では DTRlearn2、DTRreg、DynTxRegime 等の package が利用可能である。[4][5][6] 一方 SAS では、Q-Learning を実施するためのプロシジャや参照文献が少なく、手法の簡便さとは対照的に SAS による実装の障壁が高いと考えられる。[7]

そこで本稿では、はじめに近年の DTR 推定に係る試験デザイン、統計手法の概要を紹介する。次いで最適な DTR 推定のための Q-Learning のアルゴリズムを説明し、最後に SAS を用いて Q-Learning を実装する方法をコードとともに紹介する。実装方法の例示にあたって、R の DynTxRegime Package に用意されている bmiData を使用した。[6]

2. 最適な DTR 推定のための試験デザインと統計手法

2.1. SMART

2.1.1. 概要

SMART は、多段階のランダム割付を通じて、アダプティブな介入の開発に資する情報を提供する試験デザインである。SMART という名称が初めて用いられたのは Murphy S.（2005 年）の文献であるが、SMART と同様の考え方は 2005 年以前から存在していた。[8][9][10] SMART の代表例として、アルツハイマー病患者の治療に関する CATIE 試験、うつ病治療に関する STAR*D 試験等があげられる。[11][12] 米国食品医薬品局（FDA: Food and Drug Administrations）が 2020 年 12 月に公表した「Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products」でも、SMART を重要なトピックの一つとして取り上げている。[13] この邦訳及び SMART の概要について、日本製薬工業協会 医薬品評価委員会 データサイエンス部会の成果物で紹介されているので、こちらも参照されたい。[14][15]

SMART の一例として、Nahum-Shani らの文献で説明されている例を図 2.に示した。[16] この例では、研究者が ADHD を有する子どもの学業成績向上を目的とした、研究計画の立案に興味を持っている状況を想定している。その際、「最初にどの介入を実施すべきか」「第一段階で効果不十分な場合（Non-Responder）に、次にどの介入を実施すべきか」という 2 つの重要な意思決定課題を考えている。

そこで研究者は、図 2.のような SMART を検討した。具体的には、まず第 1 段階目で低用量の薬物療法もしくは低強度の認知行動療法の 2 種類の介入を確率 0.5 でランダムに割り付け介入を実施し、介入効果に応じて Responder と Non-Responder にわける。第 2 段階の介入について、Responder は第 1 段階と同様の介入の継続、Non-Responder は第 1 段階の介入の強化、別の介入を併用するかのいずれかを確率 0.5 でランダムに割り付けたのち実施する。

このように試験をデザインすることで、2 つの重要な意思決定課題を 1 つの試験で検討できる。「最初にどの介入を実施すべきか」は、第 1 段階の介入効果を比較することで検討可能であり、「第一段階で効果不十分な場合（Non-Responder）に、次にどの介入を実施すべきか」は、Non-Responder に対して第 2 段階で実施した介入にどの程度効果があったかを、第 1 段階の介入内容及び結果も踏まえて検討可能である。（例. Subgroup 2 と Subgroup 3 の比較）

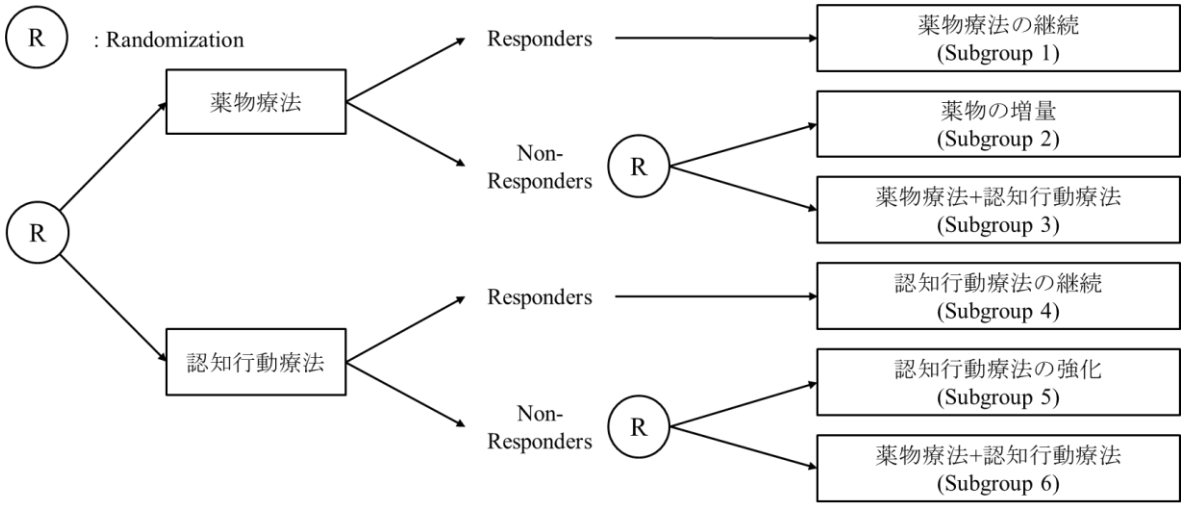


図 2. ADHD への介入に関する SMART の概略図(Nahum-Shani I らの[16]の Figure 1.を一部改変)

上記は SMART の一例であるが、研究目的に合わせて関心のある DTR 同士の比較や、ある共変量を持つ参加者に対して期待アウトカムを最大化する DTR は何かを検討可能である。

2.1.2. 利点・欠点

SMART を実施する利点として、主に①介入の相乗効果の検討、②初期の介入効果を後続する介入の個別化に利用可能、③無効なコホートの回避、が挙げられる。[17] ①について、複数のランダム化比較試験（RCT: Randomized Controlled Trials）を実施する場合と比較して、1 つの試験内で複数の介入を経時的に行うことで、初期の介入と後続する介入間の相乗効果を検討できる。②について、2 段階 SMART を例に挙げると、1 段階目の介入を実施した結果と各参加者の背景情報をもとに、2 段階目における介入の個別化に利用できる。③について、通常の RCT では、試験内で一貫した介入を受けるため、割り当てられた介入が効果不十分な場合に、悪化等による試験中止や不遵守を生じる可能性が高い。しかしながら SMART では、効果不十分な場合に介入を変更・強化等するため、参加者が試験に留まる可能性の向上が期待できる。

一方で欠点としては、多段階の割付を行うことで試験運営・管理が煩雑になる、結果の外的妥当性が低い、必要なサンプルサイズが多くなる、などが挙げられる。

2.2. 最適な DTR 推定のための統計手法

1 章で述べた通り、近年 DTR を推定するための様々な統計手法が提案されている。その中でも代表的なものについて、表 2.に手法名と概要及び関連する参照文献を示した。各統計手法の詳細な内容については参考文献だけではなく成書も参照されたい。[18,19,20] 表 2.で示した以外にも統計手法は多数あり、例えば既存手法を特定の状況に拡張した方法（小標本データ、Infinite-Horizon Settings における DTR 推定等）が挙げられる。[21,22]

表 2. 代表的な DTR 推定手法の概要

手法名	概要	参照文献
Q-learning	条件付き期待値から最適 DTR を後ろ向きに推定。データ構造に応じて、線形回帰だけではなく回帰木・カーネル等の柔軟な回帰モデルも使用可能だが、モデルの誤特定に対して脆弱。	[23,24]
A-learning	Advantage 関数を直接モデリング。治療以外の共変量効果を取り除き推定を安定化。Q-Learning よりも少ない仮定で DTR を推定するのでモデル誤特定に対して頑健である。	[25][26][27]
G-estimation	構造化ネストモデルのパラメータ推定により最適治療を導出する。時間依存性交絡にも対応可能である。	[26]
Outcome-weighted Learning (OWL)	重み付き分類問題として最適な DTR を直接推定可能。後ろ向き再帰的な BOWL(Backward OWL)や全段階を同時に推定する(Simultaneously OWL)がある。手法によってはパラメータチューニングが必要となる、計算負荷が大きくなる等の欠点がある。	[28,29]

3. Q-Learning について

3.1. 記法の整理

Q-Learning の説明に際して、以下の通り記法を整理した。

O_j: j段階目における共変量

A_j: j段階目において実施する介入$\in (0 = 非介入, 1 = 介入)$

$\bar{O}_j$: $(O_1, O_2, \dots, O_j)$

$\bar{A}_j$: $(A_1, A_2, \dots, A_j)$

Y_j: j段階目におけるアウトカム(値が大きいほど介入効果が高いと仮定)

H_j: ヒストリー$\equiv (\bar{O}_j, \bar{A}_{j-1})$ $(H_1 \equiv O_1)$

$d_j(H_j)$: ヒストリーH_jに基づく介入の意思決定ルール$\in (0, 1)$

併せてj段階 SMART と記法の関係について、図 3.に整理した。

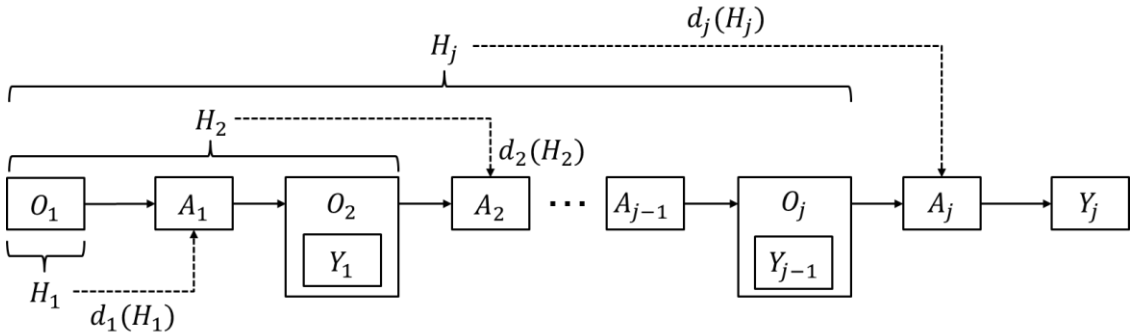


図 3. j段階 SMART と記法の関係

3.2. Q-Learning の概要

3.2.1. 強化学習のフレームワーク

強化学習（RL: Reinforcement Learning）は、エージェント（行動を選択する主体）が環境（状態の遷移や報酬を与える主体）と相互作用しながら、累積報酬の期待値を最大化するような最適な行動方策（policy）を学習する手法である。各時点でエージェントは観測した状態から行動を選択し、選択した行動に対応する報酬を環境から受け取り、次の状態に移行する。なお状態はマルコフ決定過程（MDP: Markov Decision Process）に従って遷移する。図 4.に RL のフレームワークを示した。なお、ここで示しているのは 1 段階の流れであり、これを何回も繰り返すことで累積報酬の期待値を最大化する。

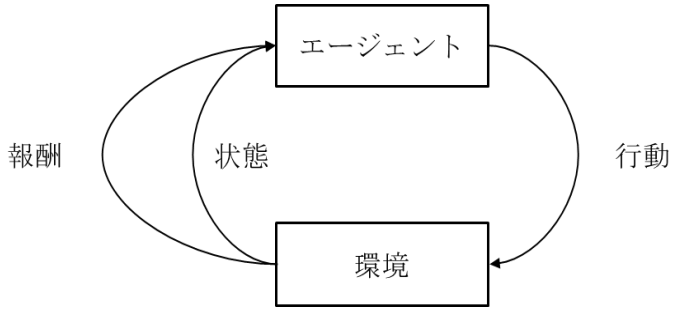


図 4. 強化学習のフレームワーク

RL における各要素を臨床現場に対応づけると、エージェントは治療方針の意思決定を行う医師、環境は患者の健康状態や臨床経過を含む医療現場そのもの、状態は各患者の臨床情報（背景因子や治療歴など）、行動は各時点での治療選択、報酬は治療後に得られる患者アウトカムに相当する。

Q-Learning は Watkins らによって提案された RL における代表的なオフポリシー型の学習アルゴリズムの一つであり、Bellman 方程式に基づく Q 関数と呼ばれる価値関数を最大化することで最適な行動方策（policy）を学習する。[23, 30]

3.2.2. Q-Learning と DTR 推定への応用

Murphy らは、RL における Q-Learning を最適な DTR 推定に応用した[2]。Q-Learning の特徴は、最終段階から逆向きに Q 関数を推定する後退帰納法（Backward Induction）を用いる点であり、この際に回帰モデル（線形モデルや非線形モデル）を利用する。

例として 2 段階 SMART における Q-Learning では、各段階に対して（疑似）アウトカムを目的変数とした Q 関数を考えた上で、まず第 2 段階における Q 関数を最大化するような DTR 推定する。その後、第 1 段階においても同様の方法で DTR を推定する。

3.3. DTR 推定のための Q-Learning

3.3.1. Q 関数の定義とモデリング

Q-Learning は、Q 関数と呼ばれる価値関数を最大化することで最適な DTR を推定する。2 段階 SMART において、各段階の Q 関数$\left(Q_2^{opt}, Q_1^{opt}\right)$は以下のように定義される。

$$Q_2^{opt}(H_2, A_2) = E[Y_2|H_2, A_2]$$

$$Q_1^{opt}(H_1, A_1) = E\left[Y_1 + \max_{a_2} Q_2^{opt}(H_2, a_2)|H_1, A_1\right]$$

上記の Q 関数が既知であると仮定した場合、各段階の最適な DTR$\left(d_1^{opt}, d_2^{opt}\right)$は以下の式で推定できる。

$$d_j^{opt}(h_j) = \underset{a_j}{\operatorname{argmax}}\, Q_j^{opt}(h_j, a_j) \qquad (j = 1,2)$$

しかしながら、実際には Q 関数が既知であることは稀なため、各段階の Q 関数について以下に示したような回帰モデルを考える。

$$Q_2^{opt}(H_2, A_2; \beta_2, \psi_2) = \beta_2^T H_{20} + (\psi_2^T H_{21}) A_2 \qquad \left(H_2 = (H_{20}, H_{21})\right) \tag{1}$$

$$Q_1^{opt}(H_1, A_1; \beta_1, \psi_1) = \beta_1^T H_{10} + (\psi_1^T H_{11}) A_1 \qquad \left(H_1 = (H_{10}, H_{11})\right) \tag{2}$$

回帰モデルの右辺について、第 1 項は共変量（性別、年齢等）の主効果を表し、第 2 項は共変量と治療の交互作用を表している。本稿では単純な線形回帰モデルを念頭に、Q-Learning に基づく最適な DTR の推定方法を説明するが、状況に応じて複雑なモデルも設定可能である。例えば、回帰モデルにロジットやベイズ加法回帰木（BART: Bayesian Additive Regression Trees）を適用した事例も存在する。[31][32]

3.3.2. Q 関数のパラメータ推定（2 段階 SMART）

2 段階 SMART における最適な DTR 推定に際して、式(1)及び(2)で規定した Q 関数のパラメータを推定する必要がある。そこで、以下の step1. ~ step3.で構成される Q-Learning のアルゴリズムにより Q 関数のパラメータを推定する。

step1. 第 2 段階における Q 関数Q_2^{opt}のパラメータ$\hat{\beta}_2, \hat{\psi}_2$の推定

$$\left(\hat{\beta}_2, \hat{\psi}_2\right) = \underset{\beta_2, \psi_2}{\operatorname{argmin}} \frac{1}{n} \sum\nolimits_{i=1}^{n} \left\{ Y_{2,i} - Q_2^{opt}\left(H_{2,i}, A_{2,i}; \beta_2, \psi_2\right) \right\}^2 \tag{3}$$

step2. 第 1 段階における疑似アウトカム$\hat{Y}_{1,i}$の計算

1.で推定した$\hat{\beta}_2, \hat{\psi}_2$を基に疑似アウトカムを計算する。ここで疑似アウトカムは、第 1 段階のアウトカムに第 2 段階で最適な DTR に従って介入した場合に得られるアウトカムを足し合わせたものである。すなわち、第 2 段階で最適な DTR に従った場合のアウトカムを計算していることに相当する。

$$\hat{Y}_{1,i} = Y_{1,i} + \max_{a_2} Q_2^{opt}\left(H_{2,i}, a_2;\ \hat{\beta}_2, \hat{\psi}_2\right) \qquad (i = 1,2, \dots, n) \tag{4}$$

step3. 第 1 段階における Q 関数Q_1^{opt}のパラメータ$\hat{\beta}_1, \hat{\psi}_1$の推定

$$\left(\hat{\beta}_1, \hat{\psi}_1\right) = \underset{\beta_1, \psi_1}{\operatorname{argmin}} \frac{1}{n} \sum\nolimits_{i=1}^{n} \left\{ \hat{Y}_{1,i} - Q_1^{opt}\left(H_{1,i}, A_{1,i}; \beta_1, \psi_1\right) \right\}^2$$

3.3.3. 最適な DTR の推定

3.3.2 では、Q 関数（Q_2^{opt}, Q_1^{opt}）のパラメータ$\hat{\beta}_2, \hat{\psi}_2, \hat{\beta}_1, \hat{\psi}_1$を step1. ~ step3.で推定した。これらを基に各参加者・段階の最適な DTR は以下の式で推定可能である。

$$\hat{d}_{j,i}^{opt}\left(h_{j,i}\right) = \underset{a_{j,i}}{\operatorname{argmax}}\, Q_j^{opt}\left(h_{j,i}, a_{j,i}; \hat{\beta}_j, \hat{\psi}_j\right) \qquad (i = 1,2, \dots, n, j = 1,2) \tag{5}$$

具体的にはある参加者について、Q 関数のパラメータ推定値と参加者のヒストリーを式に代入したうえで、Q 関数の値を最大化する介入が、その参加者における最適な DTR である。ここで式(1)及び(2)より、

いずれも Q 関数が取り得る値の大小と介入が関連する部分は第 2 項の$\left(\left(\psi_2^T H_{21}\right)A_2\right.$及び$\left.\left(\psi_1^T H_{11}\right)A_1\right)$のみである。括弧内の値が 0 未満であれば、$A_2, A_1$の値は非介入(=0)としたときに Q 関数が最大値をとる。一方で括弧内の値が 0 以上であれば、A_2, A_1の値は介入 (=1)としたときに Q 関数が最大値をとる。これを式として表したもの以下である。ここでIは指示関数である。

$$\hat{d}_{j,i}^{opt}\left(h_{j,i}\right) = \underset{a_{j,i}}{\operatorname{argmax}}\, Q_j^{opt}\left(h_{j,i}, a_{j.i}; \hat{\beta}_j, \hat{\psi}_j\right) = I\left(\psi_j^T H_{ji} > 0\right) \qquad \left(i = 1,2,\dots,n, j = 1,2. a_{j,i} \in (0,1)\right)$$

本稿では介入$A \in (0,1)$と規定しているが、$A \in (-1,1)$としている文献も一定数存在する。その場合は指示関数Iではなく、符号関数 sign を用いて各段階の最適な DTR を推定できる。

Q-Learning に基づき最適な DTR $\hat{d}_{j,i}^{opt}\left(h_{j,i}\right)$を推定する際、Q 関数のうち介入と共変量の交互作用項にのみ着目しており、共変量の主効果の項は実質的に最適な DTR 推定に寄与していない。Q 関数のうち介入Aの関数に焦点を当てた DTR 推定方法が前述の A-Learning である。

3.3.4. 多段階 SMART への拡張

3 段階以上の多段階 SMART においても、同様のアルゴリズムで Q-Learning による最適な DTR 推定が可能である。なおK段階 SMART($K \geq 3$)における Q-Learning のアルゴリズムは$(2K-1)$ステップで構成される。

K段階 SMART における各段階の Q 関数は以下のように定義される。ここで$j = K$の場合、$Q_{K+1}^{opt} = 0$となる点に留意が必要である。

$$Q_j^{opt}\left(H_j,\, A_j\right) = E\left[Y_j + \max_{a_{j+1}} Q_{j+1}^{opt}\left(H_{j+1},\, a_{j+1}\right) | H_j,\, A_j\right] \qquad (j = 1,\dots,K)$$

各段階の Q 関数の回帰モデル、パラメータ推定と疑似アウトカムの計算、及び最適な DTR 推定について、それぞれ式(6)から(8)のように構成できる。式(6)は式(1)及び(2)を一般化した式、式(7)は式(4)を式(5)に代入した形式、式(8)は式(5)のjが取りうる値をKまで拡張したものに対応する。

$$Q_j^{opt}\left(H_j,\, A_j; \beta_j, \psi_j\right) = \beta_j^T H_{j0} + \left(\psi_j^T H_{j1}\right)A_j \qquad (j = 1,\dots,K) \tag{6}$$

$$\left(\hat{\beta}_j,\, \hat{\psi}_j\right) = \underset{\beta_j,\psi_j}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}\left\{Y_{j,i} + \max_{a_{j+1}} Q_{j+1}^{opt}\left(H_{(j+1),i},\, a_{(j+1),i}; \hat{\beta}_{j+1}, \hat{\psi}_{j+1}\right) - Q_j^{opt}\left(H_{j,i},\, A_{j,i}; \beta_j,\, \psi_j\right)\right\}^2 \tag{7}$$

$$\hat{d}_j^{opt}\left(h_j\right) = \underset{a_j}{\operatorname{argmax}}\, Q_j^{opt}\left(h_j,\, a_j; \hat{\beta}_j, \hat{\psi}_j\right) \qquad (j = 1,\ \dots, K) \tag{8}$$

4. SAS による実装例

4.1. 使用するテストデータ

Q-Learning を SAS で実装するにあたり、R の DynTxRegime Package に用意されている bmiData を使用した。このテストデータは図 5.に示した、思春期の肥満に対する食事代替シェイクの介入効果を研究した 2 段階 SMART における試験フローによって生成されたダミーデータである。

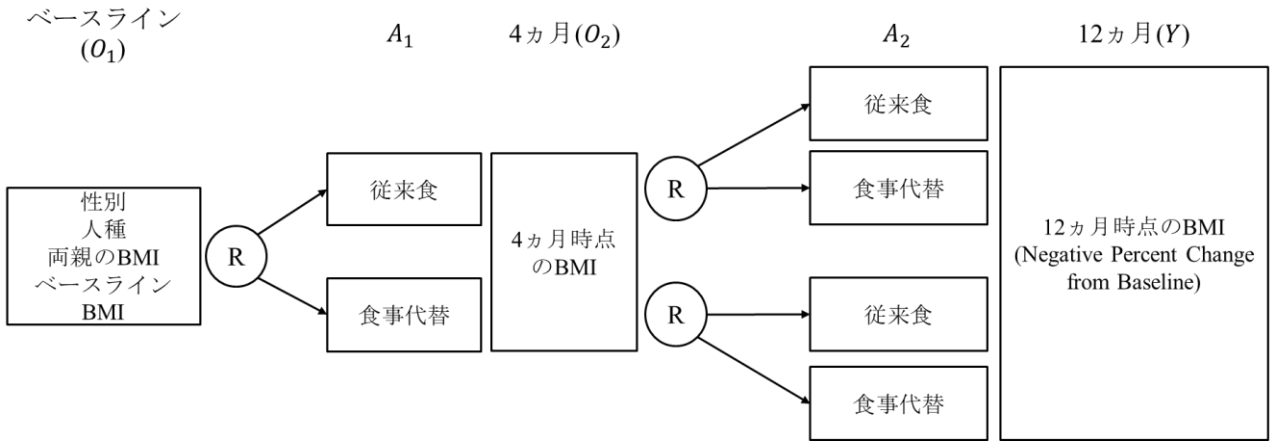


図 5. bmiData における仮想試験フロー

bmiData に含まれる変数の概要を表 4.に示した。このデータには 210 人分のデータが含まれている。なお“(追加)”とある変数は、Q-Learning 実施に際して bmiData の一部変数を用いて追加した変数である。

表 4. bmiData の概要

変数名	内容
gender	性別∈{0(=Female), 1(=Male)}
race	人種∈{0(=African American), 1(=Other)}
parentBMI	ベースライン時の両親の BMI
baselineBMI	ベースライン時の参加者の BMI
month4BMI	ベースラインから 4 ヵ月後の参加者の BMI
month12BMI	ベースラインから 12 ヵ月後の参加者の BMI
A1	1 段階目の介入 ∈ {CD(従来食), MR(食事代替)}
A2	2 段階目の介入 ∈ {CD(従来食), MR (食事代替)}
num_A1(追加)	1 段階目の介入のダミー変数 ∈ {0(=CD), 1(=MR)}
num_A2(追加)	2 段階目の介入のダミー変数 ∈ {0(=CD), 1(=MR)}
y(追加)	12 ヵ月後の参加者の BMI のベースラインからの変化率の負の値（値が大きいほど BMI の減少効果が高いことを示す指標） {(month12BMI - baselineBMI) / baselineBMI} * (-100)

BMI: Body Mass Index, CD: Conventional Diet, MR: Meal Replacement

4.2. SAS による実装

4.2.1. Q 関数の回帰モデルの設定

SAS により Q-Learning を実装する前に、まずは Q 関数の回帰モデルを仮定する必要がある。そこで本実装例では、各段階の Q 関数に対して仮定する回帰モデルは以下の通りとした。

$$
\begin{aligned}
&Q_2^{opt}(H_2, A_2; \beta_2, \psi_2) \\
&= \beta_{20} + \beta_{21} \cdot gender + \beta_{22} \cdot race + \beta_{23} \cdot parentBMI + \beta_{24} \cdot month4BMI + \beta_{25} \cdot A_1 \\
&+ (\psi_{20} + \psi_{21} \cdot A_1 + \psi_{22} \cdot parentBMI + \psi_{23} \cdot month4BMI) A_2
\end{aligned}
$$

$$
\begin{aligned}
&Q_1^{opt}(H_1, A_1; \beta_1, \psi_1) \\
&= \beta_{10} + \beta_{11} \cdot gender + \beta_{12} \cdot race + \beta_{13} \cdot parentBMI + \beta_{14} \cdot baselineBMI \\
&+ (\psi_{10} + \psi_{11} \cdot parentBMI + \psi_{12} \cdot baselineBMI) A_1
\end{aligned}
$$

4.2.2. SAS による Q-Learning と最適な DTR 推定の実装

各段階の Q 関数のパラメータ及び最適な DTR 推定について、3.3.2 及び 3.3.3 の内容を SAS で実装する方法を段階的に説明する。SAS による Q-Learning の実装方法について、日本製薬工業協会 医薬品評価委員会データサイエンス部会の成果物「連続多段階ランダム割付試験（SMART：Sequential Multiple Assignment Randomized Trials）の紹介」を参照した。[15]

4.2.2.1. テストデータの加工

bmiData に対して、表 4.で示した追加変数を SAS コード 1.により追加した。

SAS コード 1. bmiData の加工

```
data bmiData; set bmiData;
  length num_A1 num_A2 y 8.;

  /*介入のダミー変数(CD=0, MR=1)*/
  if A1 = "CD" then num_A1 = 0; else if A1 = "MR" then num_A1 = 1;
  if A2 = "CD" then num_A2 = 0; else if A2 = "MR" then num_A2 = 1;

  /*Negative percent change from Baseline*/
  y = ((month12BMI - baselineBMI) / baselineBMI) *(-100);
run;
```

4.2.2.2. Q_2^{opt}のパラメータ推定

1 つ目のステップとして SAS コード 2.により、2 段階目の Q 関数である$Q_2^{opt}(H_2, A_2; \beta_2, \psi_2)$のパラメータ$\beta_2, \psi_2$を GLM プロシジャにより推定した。$Q_2^{opt}$のパラメータ推定結果は、後ほど Q 関数の計算に利用するため、ODS OUTPUT を利用して Q2_Estimate に格納した。

SAS コード 2. Q_2^{opt}のパラメータβ_2, ψ_2の推定

```
/*Step1 - Q2のパラメータ推定*/
ods output   ParameterEstimates = Q2_Estimate;
proc glm data = bmiData;
  model y = gender race parentBMI month4BMI num_A1
            num_A2 num_A1*num_A2 parentBMI*num_A2 month4BMI*num_A2 / solution;
run;
quit;
```

実行結果を表 5.に示した。parentBMI, month4BMI 及び A1 の 3 つの項目で p 値が 0.05 未満であり、かついずれのパラメータ推定値も負であった。このことから両親の BMI、4 ヵ月目の BMI が高い参加者、そして 1 段階目で MR の介入を受けている参加者は、BMI 減少効果が減弱することが示唆された。

表 5. Q_2^{opt}のパラメータの推定結果

Coefficients	Estimate	SE	t Value	P-Value
Intercept	49.07656	5.51698	8.896	<.0001*
gender	-0.49754	0.89552	-0.556	0.57911
race	0.52277	0.91002	0.574	0.56630
parentBMI	-0.34053	0.13531	-2.517	0.01263*
month4BMI	-0.83690	0.20826	-4.019	<.0001*
A1(=MR)	-3.40834	1.25009	-2.726	0.00697*
A2(=MR)	-13.83522	7.84484	-1.764	0.07933
A1*A2	1.34430	1.78209	0.754	0.45413
parentBMI*A2	0.366645	0.20176	1.816	0.07082
month4BMI*A2	0.04123	0.27593	0.149	0.88137

*: $p < 0.05$

4.2.2.3.　2 段階目における Q 関数の計算及び最適な DTR 推定

2 段階目において、参加者ごとのアウトカムを最大化する介入a_2を推定するために、各介入を実施した場合の Q 関数の値（$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$）を計算する必要がある。$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$を計算するために、表 5.で示したパラメータを用いるため、SAS コード 3.により Q2_Estimate を加工・転置したうえで bmiData に結合した。

SAS コード 3. bmiData に対する$\hat{\beta}_2, \hat{\psi}_2$の結合

```
data Q2_Estimate; set Q2_Estimate;
  length Coef $100;
  Coef = "Q2_"||strip(Parameter);
run;
```

```
proc transpose data = Q2_Estimate out = Q2_Estimate(drop = _:);
  id Coef; var Estimate;
run;

data bmiData; set bmiData;
  /*表5.のパラメータをbmiDataに結合*/
  if _N_ = 1 then set Q2_Estimate;
run;
```

SAS コード 4.により、2 段階目にて CD, MR の各介入を実施した場合における$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$を計算し、それぞれ Q2_CD_PO(= $Q_2(h_2, a_2 = 0; \hat{\beta}_2, \hat{\psi}_2)$), Q2_MR_PO(= $Q_2(h_2, a_2 = 1; \hat{\beta}_2, \hat{\psi}_2)$)に格納した。

SAS コード 4. $Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の計算

```
data bmiData; set bmiData;
  length Q2_CD_PO Q2_MR_PO 8.;

  /*2段階目の介入をCD(num_A2 = 0)とした場合のQ関数の値*/
  Q2_CD_PO = Q2_Intercept + gender*Q2_gender + race*Q2_race + parentBMI*Q2_parentBMI
             + month4BMI * Q2_month4BMI + num_A1 * Q2_num_A1;

  /*2段階目の介入をMR(num_A2 = 1)とした場合のQ関数の値*/
  Q2_MR_PO = Q2_Intercept + gender*Q2_gender + race*Q2_race + parentBMI*Q2_parentBMI +
             month4BMI * Q2_month4BMI + num_A1 * Q2_num_A1 + Q2_num_A2 + num_A1 *
             Q2_num_A1_num_A2 + parentBMI * Q2_parentBMI_num_A2 +
             month4BMI *Q2_month4BMI_num_A2;
run;
```

SAS コード 5.より、各介入に基づく$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の大小を比較し、比較結果に基づき 2 段階目における最適な DTR を構築した。ここで Q2_CD_PO > Q2_MR_PO であれば CD、Q2_CD_PO ≤ Q2_MR_PO であれば MR が 2 段階目においてアウトカムを最大化する介入である。本実装例では、$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の値が完全に等しい（Q2_CD_PO = Q2_MR_PO）場合は MR を選択する方針としているが、実際の取り扱いは研究分野や内容に依存する。なお一般的にはリーズナブルな介入を選択することが望ましいとされている。

SAS コード 5. 2 段階目における最適な DTR の推定

```
data bmiData; set bmiData;
  length Q2_PO 8. opt_A2 $4;
  if Q2_CD_PO > Q2_MR_PO then do;
    Q2_PO = Q2_CD_PO; opt_A2 = "CD";
  end;
```

```
  if Q2_CD_PO <= Q2_MR_PO then do;
    Q2_PO = Q2_MR_PO; opt_A2 = "MR";
  end;
run;
```

表 6.に、1 段階目で実際に受けた介入a_1、第 2 段階における最適な DTR の推定結果に基づき選択された介入$\hat{d}_2^{opt}(h_2)$、及び$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の平均を示した。

前述の通り、1 段階目で MR の介入を実施している場合、CD と比較して BMI 減少効果が減弱することが、$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の平均にも反映されていることが確認できた。

表 6. 第 2 段階のみ最適な DTR に従った場合の$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の要約

a_1	$\hat{d}_2^{opt}(h_2)$	n	$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$の平均
CD	CD	57	11.008
CD	MR	52	6.616
MR	CD	42	9.492
MR	MR	59	3.626
Total		210	7.543

4.2.2.4. Q_1^{opt}のパラメータ推定

4.2.2.2 と同様の流れで、SAS コード 6.により 1 段階目の Q 関数である$Q_1^{opt}(H_1, A_1; \beta_1, \psi_1)$のパラメータ$\beta_1, \psi_1$を推定した。ここで疑似アウトカムが必要となるが、本実装例では各参加者の$Q_2(h_2, a_2; \hat{\beta}_2, \hat{\psi}_2)$を疑似アウトカムとしてそのまま利用した。

SAS コード 6. Q_1^{opt}のパラメータβ_1, ψ_1の推定

```
/*Step3 - Q1のパラメータ推定*/
ods output   ParameterEstimates = Q1_Estimate;
proc glm data = bmiData;
  model   Q2_PO = gender race parentBMI baselineBMI   num_A1 num_A1*parentBMI
                  num_A1*baselineBMI / solution;
run;
quit;
```

実行結果を表 7.に示した。parentBMI, baselineBMI, parentBMI*A1, baselineBMI*A1 の 4 項目で p 値が 0.05 未満となった。パラメータ推定値を確認すると baselineBMI*A2 のみ正であり、残りの 3 項目は負の値であった。両親の BMI、ベースライン BMI 及び両親の BMI と 1 段階目の介入(=MR)に交互作用がある参加者では、BMI 減少効果が減弱することが示唆された。一方でベースライン BMI と 1 段階目の介入(=MR)に交互作用がある場合、すなわちベースライン BMI の値が大きい参加者でかつ 1 段階目に MR を実施している参加者では、ベースライン BMI が高いほど BMI 減少効果が増強することが示唆された。

但しここで 1 つ注意すべき点がある。Q_1^{opt}のパラメータがQ_2^{opt}のパラメータに基づき計算した疑似アウトカムから推定されることである。そのため、Q_2^{opt}のパラメータ推定時の誤差がQ_1^{opt}のパラメータ推定にも伝播し、段階を経るにつれて Q 関数のパラメータ及び最適な DTR の推定精度が低下する傾向があることに留意すべきである。

表 7. Q_1^{opt}のパラメータの推定結果

Coefficients	Estimate	SE	t Value	P-Value
Intercept	40.65904	1.67045	24.340	< 0.0001*
gender	-0.46497	0.24886	-1.868	0.063153
race	0.40800	0.25012	1.631	0.104399
parentBMI	-0.14565	0.03553	-4.099	< 0.0001*
baselineBMI	-0.69651	0.04783	-14.562	< 0.0001*
A1(=MR)	-4.13445	2.58062	-1.602	0.110693
parentBMI*A1	-0.27188	0.04720	-5.761	< 0.0001*
baselineBMI*A1	0.26035	0.07192	3.620	0.000372*

*: p<0.05

4.2.2.5. 1 段階目の最適な DTR 推定

最後に SAS コード 7.から 9.により、4.2.2.3 と同様の方法で 1 段階目における最適な DTR を推定した。

SAS コード 7. bmiData に対する$\hat{\beta}_1, \hat{\psi}_1$の結合

```
data Q1_Estimate; set Q1_Estimate;
  length Coef $100;
  Coef = "Q1_"||strip(Parameter);
run;

proc transpose data = Q1_Estimate out = Q1_Estimate(drop = _:);
  id Coef;
  var Estimate;
run;

data bmiData; set bmiData;
  if _N_ = 1 then set Q1_Estimate;
run;
```

SAS コード 8. $Q_1(h_1, a_1; \hat{\beta}_1, \hat{\psi}_1)$の計算

```
data bmiData; set bmiData;
  length Q1_CD_PO Q1_MR_PO 8.;

/*1段階目の介入がCD(num_A1 = 0)のQ関数の値*/
  Q1_CD_PO = Q1_Intercept + gender * Q1_gender + race * Q1_race + parentBMI * Q1_parentBMI
                + baselineBMI * Q1_baselineBMI;

  /*1段階目の介入がMR(num_A1 = 1)のQ関数の値*/
  Q1_MR_PO = Q1_Intercept + gender * Q1_gender + race * Q1_race + parentBMI * Q1_parentBMI
                + baselineBMI * Q1_baselineBMI +Q1_num_A1 + parentBMI * Q1_parentBMI_num_A1
                + baselineBMI * Q1_baselineBMI_num_A1;
run;
```

SAS コード 9. 第 1 段階における最適な DTR 推定

```
/*1段階目における最適なDTRの推定*/
data bmiData; set bmiData;
  length Q1_PO 8. opt_A1 $4;

  if Q1_CD_PO > Q1_MR_PO then do;
    Q1_PO = Q1_CD_PO; opt_A1 = "CD";
  end;

  if Q1_CD_PO <= Q1_MR_PO then do;
    Q1_PO = Q1_MR_PO; opt_A1 = "MR";
  end;
run;
```

表 8.に、1 段階目、2 段階目ともに最適な DTR に従って介入を受けた場合に期待されるアウトカムの要約を示した。1 名を除く全ての参加者で、1 段階目に CD を実施することで期待されるアウトカムが最大化される結果を得た。

実際に受けた介入に基づいたアウトカムの平均値は 6.459（95%CI: 5.402, 7.510）であった。それに対して最適な DTR に従った場合に期待されるアウトカムは 9.251 (95%CI: 8.832, 9.670)であった。このことから、各段階の Q 関数に仮定したモデルが正しい仮定の下で、最適な DTR に従って介入を実施した場合、実際の介入よりも約 2.8%の BMI 減少効果が見込まれると解釈できる。

表 8. 最適な DTR に従った場合に期待されるアウトカムの要約

$\hat{d}_1^{opt}(h_1)$	$\hat{d}_2^{opt}(h_2)$	n	期待されるアウトカムの平均
CD	CD	99	11.312
CD	MR	110	7.479
MR	CD	1	0.310
MR	MR	0	-
Total		210	9.251

5. まとめ

本稿では、患者個々の状態変化に応じた最適な治療方針である DTR をデータから推定する手法として広く用いられている Q-Learning について、近年の DTR 推定に関する試験デザイン（SMART）や統計手法を紹介した。また、テストデータを用いた Q-Learning の SAS による実装方法についても紹介した。

今回示した実装例では、2 段階 SMART かつ Q 関数に比較的シンプルなモデルを仮定しているため、SAS でも容易に Q-Learning を実装できた。しかし、多段階 SMART や複雑な回帰モデルを用いる場合には、SAS で Q-Learning を実装するハードルは高く、場合によっては実装が困難となる可能性もある。そのため、他の推定手法の検討や、SAS 以外のソフトウェアで実行可能な方法を用意しておく必要がある。

最適な DTR 推定について、統計手法だけではなくデータソースも多様化しており、観察研究やリアルワールドデータなどが該当する。これらのデータに基づいて DTR を推定する場合、今回解説した手法をそのまま適用することはできず、仮定を置く必要があるため留意が必要である。詳細は Robert らの文献等を参照されたい。[33]

最後に、Q-Learning による最適な DTR 推定は、モデルが比較的シンプルで応用範囲も広く、SAS や R などのソフトウェアで容易に実装可能であることから、今後も幅広い分野への応用が期待される。

6. 参考文献

[1] 松井茂之. 個別化医療の臨床試験：デザインとデータ解析. 計量生物学 Vol. 43, No. 1, 97− 119 (2022). https://www.jstage.jst.go.jp/article/jjb/43/1/43_97/_pdf （最終アクセス日: 2025 年 9 月 1 日）

[2] Murphy S. An experimental design for development of adaptive treatment strategies. Stat Med. 2005;24:1455–1481.

[3] Whiston A, Kidwell KM, O'Reilly S, Walsh C, Walsh JC, Glynn L, Robinson K, Hayes S. The use of sequential multiple assignment randomized trials (SMARTs) in physical activity interventions: a systematic review. BMC Med Res Methodol. 2024 Dec 19;24(1):308.

[4] CRAN: DTRlearn2: Statistical Learning Methods for Optimizing Dynamic Treatment Regimes https://cran.r-project.org/web/packages/DTRlearn2/index.html（最終アクセス日: 2025 年 9 月 1 日）

[5] CRAN: DTRreg: DTR Estimation and Inference via G-Estimation, Dynamic WOLS, Q-Learning, and Dynamic Weighted Survival Modeling (DWSurv) https://cran.r-project.org/web/packages/DTRreg/index.html（最終アクセス日: 2025 年 9 月 1 日）

[6] CRAN: DynTxRegime: Methods for Estimating Optimal Dynamic Treatment Regimes. https://cran.r-project.org/web/packages/DynTxRegime/index.html（最終アクセス日: 2025 年 9 月 1 日）

[7] SAS PROC QLEARN User's Guide Version 1.0.0 https://d3c.isr.umich.edu/wp-content/uploads/2021/01/ug-1kus9ah.pdf（最終アクセス日: 2025 年 9 月 1 日）

[8] Lavori PW, Dawson R. A design for testing clinical strategies: biased adaptive within‐subject randomization.

[9] Dawson R, Lavori PW. Sequential causal inference: Application to randomized trials of adaptive treatment strategies. Statistics in Medicine. 2005;27:1626–45.

[10] Kubota K, Furuse K, Kawahara M, Kodama N, Yamamoto M, Ogawara M, Negoro S, Masuda N, Takada M, Matsui K, et al. Role of radiotherapy in combined modality treatment of locally advanced non-small-cell lung cancer. J Clin Oncol. 1994 Aug;12(8):1547-52.

[11] Schneider LS, Ismail MS, Dagerman K, Davis S, Olin J, McManus D, Pfeiffer E, Ryan JM, Sultzer DL, Tariot PN. Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer's disease trial. Schizophr Bull. 2003;29(1):57-72.

[12] Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA, Quitkin FM, Kashner TM, Kupfer DJ, Rosenbaum JF, Alpert J, Stewart JW, McGrath PJ, Biggs MM, Shores-Wilson K, Lebowitz BD, Ritz L, Niederehe G; STAR*D Investigators Group. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. Control Clin Trials. 2004 Feb;25(1):119-42.

[13] Food and Drug Administration. "Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry". https://www.fda.gov/regulatory-information/search-fda-guidance-documents/interacting-fda-complex-innovative-trial-designs-drugs-and-biological-products （最終アクセス日: 2025 年 9 月 1 日）

[14] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 "Complex Innovative Trial Design の現状と事例の整理 - FDA ガイダンス邦訳付き"（2024） https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm00000010pp-att/DS_202404_CID.pdf （最終アクセス日: 2025 年 9 月 1 日）

[15] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 "連続多段階ランダム割付試験（SMART：Sequential Multiple Assignment Randomized Trials）の紹介" https://www.jpma.or.jp/information/evaluation/results/allotment/od4err0000000vd0-att/DS_202211_SMART.pdf （最終アクセス日: 2025/8/27）

[16] Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, Waxmonsky JG, Yu J, Murphy SA. Experimental design and primary data analysis methods for comparing adaptive interventions. Psychol Methods. 2012 Dec;17(4):457-477.

[17] H Lei et al. A "SMART" Design for Building Individualized Treatment Sequences. Annu Rev Clin Psychol. 2012:8:21-48.

[18] B. Chakraborty, E. Moodie. Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine. (2014).

[19] Michael R. Kosorok, Erica E. M. Moodie. Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine. (2015).

[20] Anastasios A. Tsiatis, Marie Davidian, Shannon T. Holloway, Eric B Laber. Dynamic Treatment Regimes. Statistical Methods for Precision Medicine. (2019).

[21] Tamura RN, Krischer JP, Pagnoux C, Micheletti R, Grayson PC, Chen YF, Merkel PA. A small n sequential multiple assignment randomized trial design for use in rare disease research. Contemp Clin Trials. 2016 Jan;46:48-51.

[22] Ashkan Ertefaie , Robert L Strawderman. Constructing dynamic treatment regimes over indefinite time horizons. Biometrika, Volume 105, Issue 4, December 2018, Pages 963–977.

[23] Watkins CJCH. Doctoral Thesis. Learning from delayed rewards. University of Cambridge; England: 1989.

[24] Watkins CJCH, Dayan P. Q-learning. Mach. Learn. 1992;8:279–292.

[25] Murphy SA. Optimal dynamic treatment regimes (with discussion) J. Royal Statist. Soc. Ser. B. 2003;58:331–366.

[26] Robins JM. Optimal structured nested models for optimal sequential decisions. In: Lin DY, Heagerty PJ, editors. Proceedings of the Second Seattle Symposium on Biostatistics. New York: Springer; 2004. pp. 189–326.

[27] Blatt D, Murphy SA, Zhu J. Technical Report 04-63. The Methodology Center, Pennsylvania State University; 2004. A-learning for approximate planning.

[28] Zhao YQ, Zeng D, Laber EB, Kosorok MR. New Statistical Learning Methods for Estimating Optimal Dynamic Treatment Regimes. J Am Stat Assoc. 2015;110(510):583-598.

[29] Zhao YQ, Zeng D, Rush AJ, Kosorok MR. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. J Am Stat Assoc. 2012 Sep 1;107(449):1106-1118.

[30] Bellman RE, Dreyfus SE. Applied dynamic programming. Princeton University Press; Princeton, NJ: 1962.

[31] Liu D, He W. Q-Learning in Dynamic Treatment Regimes With Misclassified Binary Outcome. Stat Med. 2024 Dec 30;43(30):5885-5897.

[32] Li X, Logan BR, Hossain SMF, Moodie EEM. Dynamic Treatment Regimes Using Bayesian Additive Regression Trees for Censored Outcomes. Lifetime Data Anal. 2024 Jan;30(1):181-212.

[33] Mahar RK, McGuinness MB, Chakraborty B, Carlin JB, IJzerman MJ, Simpson JA. A scoping review of studies using observational data to optimise dynamic treatment regimens. BMC Med Res Methodol. 2021 Feb 22;21(1):39.

連絡先:
〒112-0002 東京都文京区小石川 1-4-1 住友不動産後楽園ビル
エイツーヘルスケア株式会社　データサイエンスソリューション部
E-mail: tokuda-y@a2healthcare.com

対数正規分布に従う変数に対するBland-Altman Plotにおける Limits of Agreementの描画

○徳田芳稀、田村沙弥

（エイツーヘルスケア株式会社　データサイエンスソリューション部）

Constructing the Limits of Agreement for variables following a log-normal distribution in a Bland-Altman Plot

Yoshiki Tokuda, Saya Tamura

Data Science Solutions Department, A2 Healthcare Corporation

要旨

Bland-Altman Plot は 2 つの評価方法により得られた測定値について、横軸に測定値の平均、縦軸に測定値の差をプロットしたものである。このプロットにより、測定値同士の一致性や系統誤差と偶然誤差を視覚的に確認できる。測定値及び差が正規分布に従う場合、許容限界（LoA: Limits of Agreement）は、測定値の差の分散が測定値の平均値に依存せず一様である仮定の下で、「測定値の差の平均 ± 1.96 ×測定値の差の標準偏差」で計算される。LoA を示すことで、2 つの評価方法により得られた測定値の差のばらつきが、実用上許容できる範囲内にあるかを評価する指標となる。一方で測定値が対数正規分布に従う場合、測定値の差の分散は、平均値に依存して大きくなり、測定値の差が正規分布に従わず、前述の仮定も成立しない。そのため、前述の LoA とは異なる方法で計算・描画する必要がある。そこで本稿では、Euser らによって提案された、対数正規分布に従う測定値の差に対する LoA の計算方法及び SAS による描画方法を説明する。

キーワード：偶然誤差、系統誤差、許容限界、LoA（Limits of Agreement）

1. はじめに

Bland-Altman plot とは 1986 年に J. M. Bland と D. G. Altman の 2 人によって提案された、関心のある評価対象に対して 2 つの評価方法により得られた測定値の一致性を確認するための、1 つの可視化方法である。測定値の平均値を横軸、測定値の差を縦軸にとることで図 1.のように描画できる。[1]

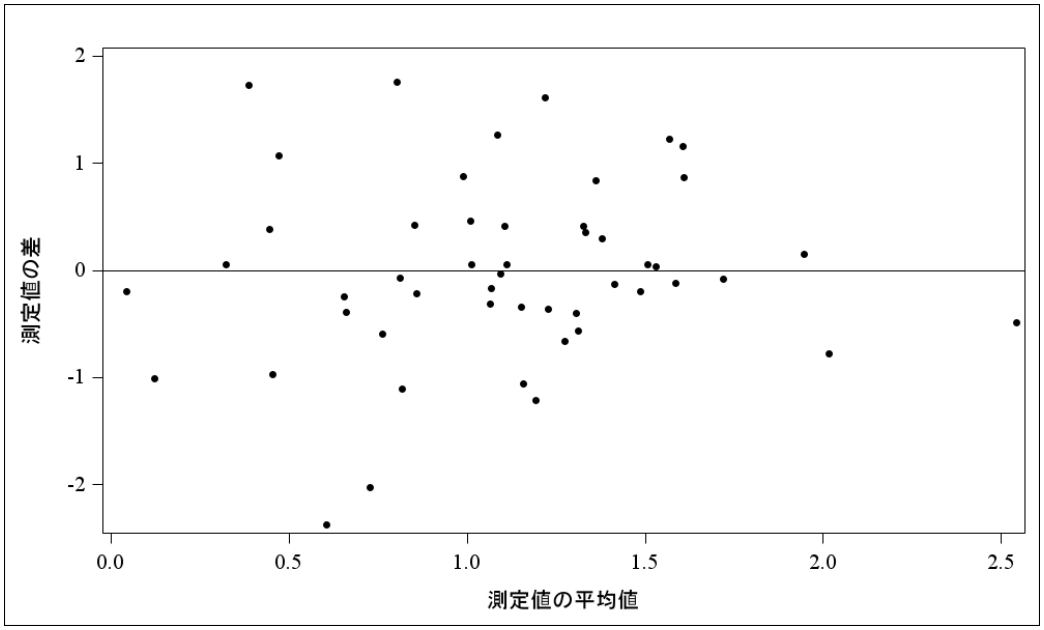


図 1. Bland-Altman Plot

Bland-Altman Plot 及び後述する LoA を描画することで、2 つのことを視覚的に確認できる。

1 つ目は、系統誤差や比例誤差の有無である。2 つの評価方法により得られた測定値の差が 0 を中心に分布している場合、評価方法間の測定値の一致性は高いと考えられる。しかしながら、例えば 2 つの評価方法のうち、いずれかの評価方法の測定値が常に低くなる傾向にある場合、測定値の差がある一定の方向に偏る、いわゆる系統誤差を生じている可能性がある。そこで Bland-Altman Plot を用いることで、図 2（左）のように系統誤差の存在を視覚的に確認できる。また別の例として、測定値の平均値が大きくなるほど測定値の差の分散が大きくなる、いわゆる比例誤差を生じている懸念がある場合も前述と同様に描画することで、図 2（右）のように比例誤差の存在を視覚的に確認できる。信頼性や一致性に関する研究の報告ガイドラインである GRRAS（Guidelines for Reporting Reliability and Agreement）の中でも、測定値の一致性の確認方法の 1 つとして挙げられている。[2]

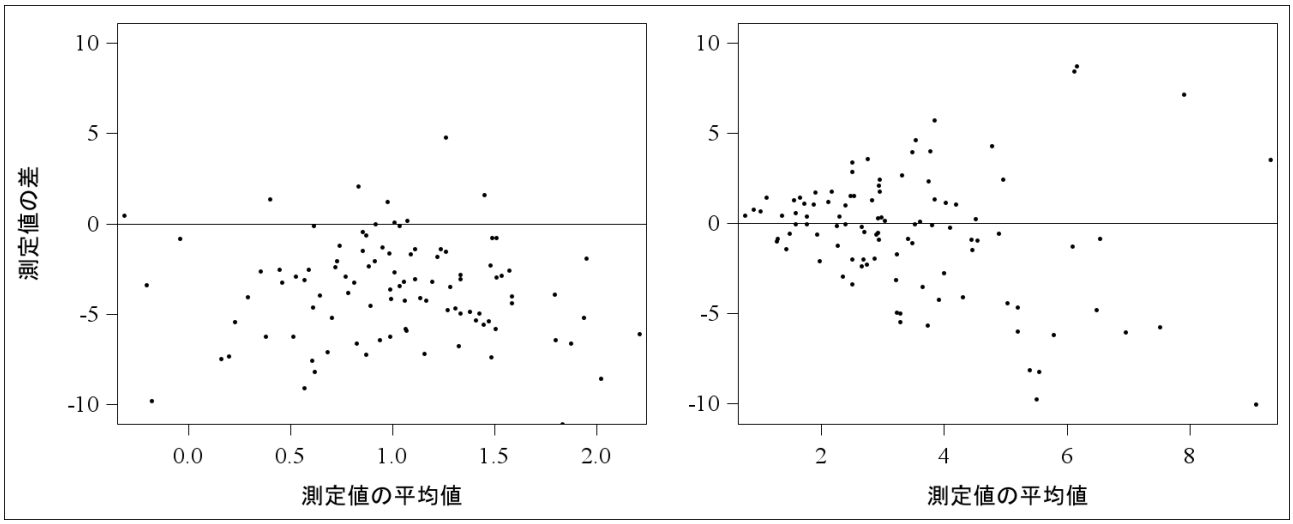


図 2. 系統誤差（左）・比例誤差（右）が存在するデータの Bland-Altman Plot

2 つ目は、2 つの評価方法により得られた測定値の差のばらつきが実用上許容できる範囲内にあるかである。ここで Bland-Altman Plot に加えて LoA の計算・描画が必要となる。LoA は、測定値の差のばらつきに係る許容範囲を表すもので、以下の式により計算される。（以下、従来法）

$$\text{測定値の差の平均} \pm 1.96 \times \text{測定値の差の標準偏差}$$

図 1.の Bland-Altman Plot に対して、従来法に基づく LoA（点線）を描画したものが図 3.である。測定値の差の平均値が正規分布に従う場合、漸近的に 95%のデータが LoA 内に含まれる。そのため、測定値の差のばらつきがこの範囲内に含まれれば、評価方法間の測定値の差のばらつきは実用上許容できる範囲内にあると解釈できる。

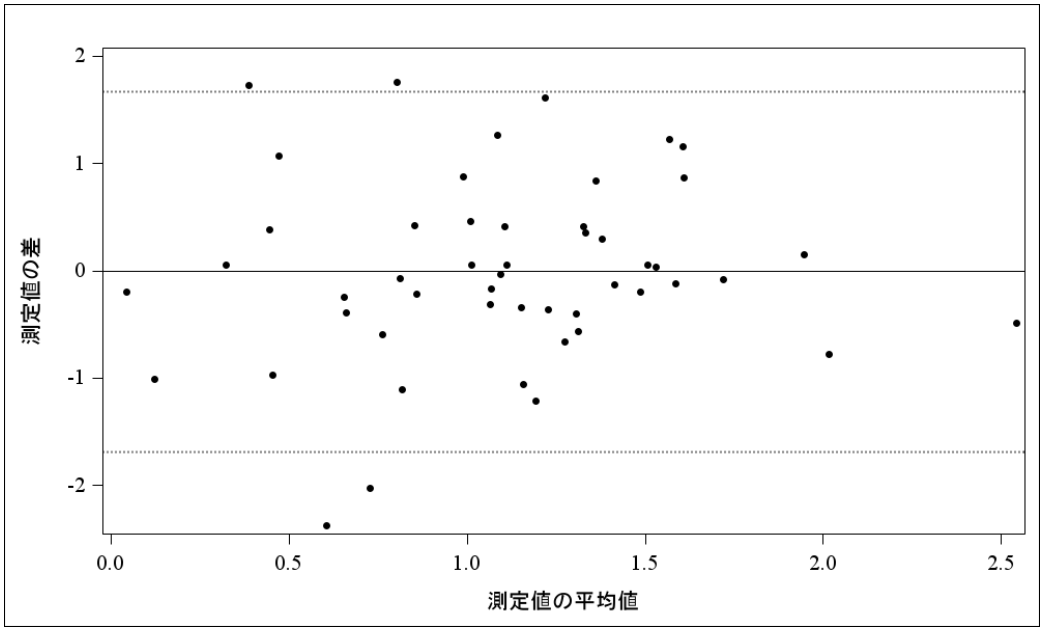


図 3. Bland-Altman Plot に対する LoA（点線）

ここで 1 つ留意すべき点がある。LoA は、測定値の差の分散が測定値の平均値に依存せず一様である仮定の下で、前述の解釈を与えることができた。しかしながら、対数正規分布に従う変数では、測定値の差の分散が測定値の平均に依存して増加する、比例誤差の傾向を示すことが知られている。そのような変数に対して従来法に基づく LoA を適用することで、測定値の差のばらつきが許容されるかを正しく評価できない可能性がある。

そこで本稿では、まず対数正規分布の基本的事項について述べ、次いで Euser らによって提案された対数正規分布に従う変数に対する LoA の計算方法（以下、Euser らの方法）を紹介し、最後にそれを SAS で計算・描画する方法を説明する。

2. 対数正規分布に従う変数の基本的事項

2.1. 対数正規分布に従う変数の性質

正規分布$N(\mu, \sigma^2)$に従う変数Zを考える。この時、

$$X = \exp(Z)$$

は対数正規分布に従うことが知られている。ここで対数正規分布の確率密度関数は次式で与えられる。

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2\right] \quad (x>0)$$

また対数正規分布に従う変数Xの期待値$E[X]$及び分散$V[X]$は

$$E[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

$$V[X] = \{\exp(\sigma^2)-1\}\exp(2\mu+\sigma^2) = \{\exp(\sigma^2)-1\}E[X]^2$$

である。

$E[X]$について、元となる正規分布のパラメータμ及びσ^2の両方に依存しており、μだけでなくσ^2の増加に伴い$E[X]$も増加する。分散$V[X]$について、期待値の二乗である$E[X]^2$が含まれているため、$E[X]$同様にμ及びσ^2の影響を受ける。特にσ^2が$V[X]$に与える影響が大きい。

前述の内容を視覚的にも確認するために、正規分布のパラメータμ, σ^2と$E[X]$と$V[X]$の関係性に図 4.に示した。σ^2を固定した条件下（$\sigma^2 = 0.25$）でμを増加させた場合（図 4.左）、μを固定した条件下（$\mu = 1$）でσ^2を増加させた場合（図 4.右）のいずれの場合においても、$E[X]$及び$V[X]$は増加傾向にあり、特にσ^2の増加が$V[X]$に与える影響が大きいことが確認できた。

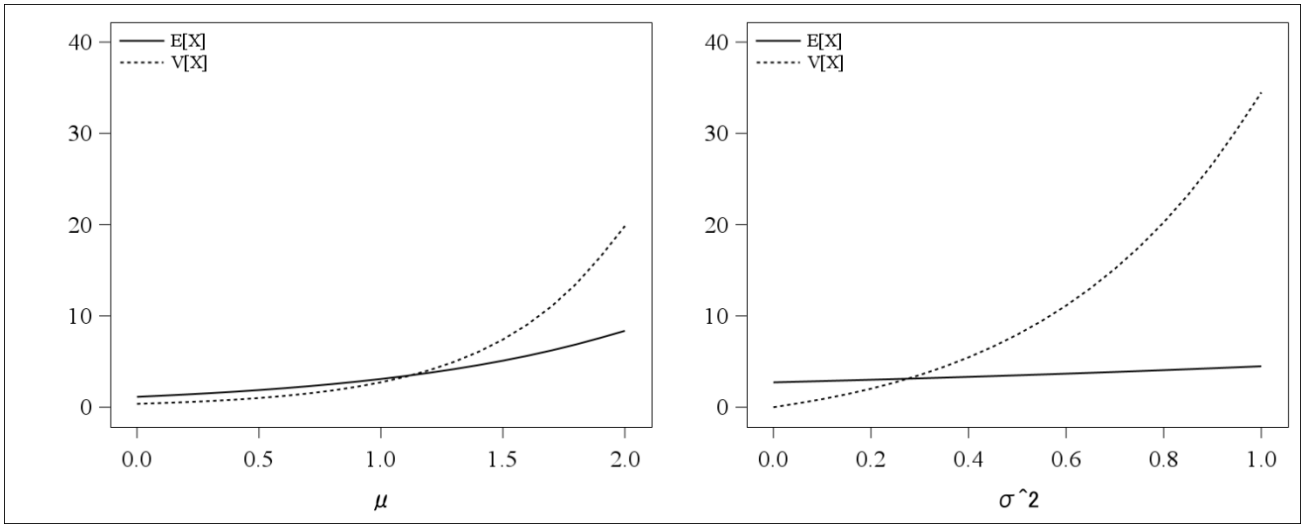


図 4. μ, σ^2と$E[X], V[X]$の関係性

（左：σ^2=0.25 を固定しμを増加させた場合，右：μ=1 を固定しσ^2を増加させた場合）

2.2. 対数正規分布に従う変数の差の性質

正規分布$N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$に従う変数$Z_1, Z_2$を考え、ネイピア数で指数変換した変数を

$$X_1 = \exp(Z_1),\ \ X_2 = \exp(Z_2)$$

とする。これらも 2.1 のXと同様に対数正規分布に従う。

X_1, X_2を各評価方法に基づく測定値、その差を$X_1 - X_2$としたときの期待値$E[X_1 - X_2]$及び分散$V[X_1 - X_2]$は

$$E[X_1 - X_2] = E[X_1] - E[X_2] = \exp\left(\mu_1 + \frac{1}{2}\sigma_1^2\right) - \exp\left(\mu_2 + \frac{1}{2}\sigma_2^2\right)$$

$$V[X_1 - X_2] = V[X_1] + V[X_2] - 2Cov[X_1, X_2] = \{\exp(\sigma_1^2) - 1\}E[X_1]^2 + \{\exp(\sigma_2^2) - 1\}\mathrm{E}[X_2]^2 - 2Cov[X_1, X_2]$$

である。

$V[X_1 - X_2]$について上式より、$E[X_1]^2, E[X_2]^2$の影響を受ける、すなわち$E[X_1], E[X_2]$の増加に比例して$V[X_1 - X_2]$も増加する。図 5.に対数正規分布に従う測定値X_1, X_2の Bland-Altman Plot と従来法に基づく LoA（点線）を示した。

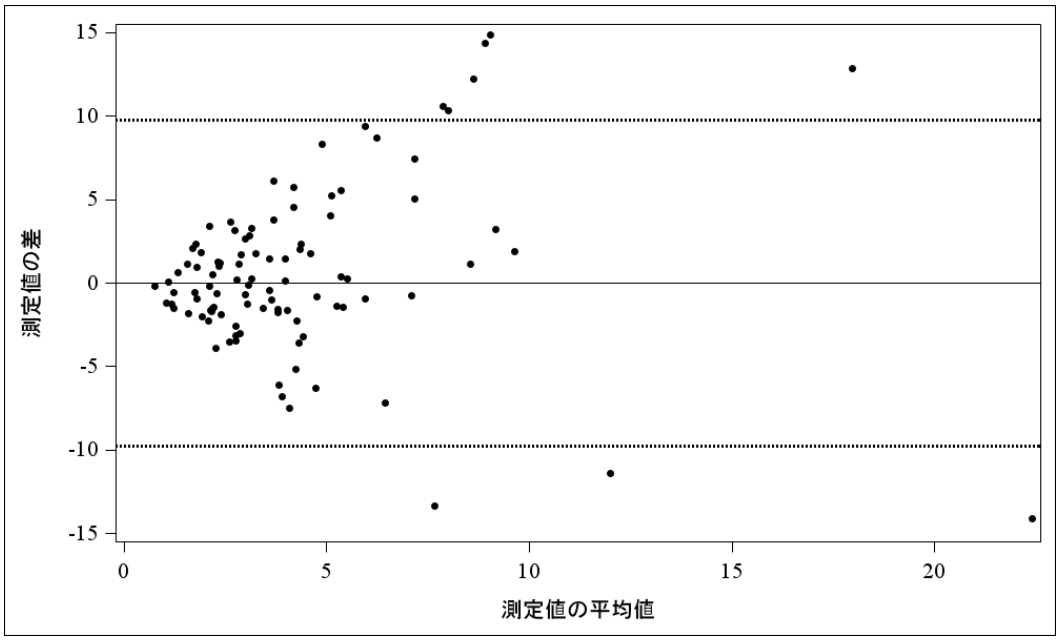


図 5. 対数正規分布に従う測定値の Bland-Altman Plot と従来法に基づく LoA

図 5.においても、測定値の平均値の増加に伴い、測定値の差のばらつきが増加することが視覚的に確認できた。さらに従来法に基づく LoA を描画すると、測定値の平均値が増加するほど LoA 外のプロットが増加する傾向も視覚的に確認できた。

従来法に基づく LoA は、測定値の平均値の大小に関わらず一定であり、前述のような状況で測定値の差のばらつきが、実用上許容できる範囲内にあるかを正しく評価できない可能性がある。3 章では、この問題に対処した Euser らの方法による LoA の計算方法及び SAS による描画方法を説明する。

3. 対数正規分布に従う変数に対する Bland-Altman Plot と LoA の描画方法

3.1. Euser らの方法

対数正規分布に従う測定値は、対数変換することで正規分布に従う。そのため、対数変換した測定値を用いることで、従来法に基づく LoA を計算・描画可能である。しかし、対数変換した測定値の平均、及びその差は、元のスケールの幾何平均及び比に相当する。測定対象の内容によっては幾何平均及び比のままでは解釈が困難な可能性があるため、測定値を対数変換せずに Bland-Altman Plot 及び LoA を描画することを考える。そこで Euser らは、元のスケールを維持したまま LoA を計算・描画する方法を提案した。[3]

Euser らの方法の説明に際して、2.2 で定義したZ_1, Z_2, X_1及びX_2を用いる。正規分布に従う 2 つの測定値の差$Z_1 - Z_2$の標準偏差をSD_{diff}とし、以下のようにおく。

$$a = 1.96 \times SD_{\text{diff}}$$

Euser らの方法では、測定値の差$X_1 - X_2$に対する LoA を以下の式で計算できる。ここで$\bar{X}$は、測定値X_1とX_2の平均値である。

$$X_1 - X_2 \pm 2\bar{X}\frac{(\exp(a) - 1)}{(\exp(a) + 1)}$$

Euser らの方法に基づく LoA の導出過程については、Appendix に詳細を示した。従来法と Euser らの方法の相違点は、計算式に測定値の平均値$\bar{X}$が含まれる点である。従来法ではSD_{diff}に対して定数 1.96 を乗じることで LoA を求めるため、LoA は測定値の平均値に依存せず一定であった。対して Euser らの方法では$\bar{X}$を含めることにより、Bland–Altman Plot の横軸である測定値の平均値に応じて差の分散が増加する状況を考慮した LoA を計算できる。

3.2. SAS による LoA の描画

まず SAS コード 1.により、対数正規分布に従う測定値を想定したサンプルデータを作成した。サンプルサイズは 10,000 とし、各評価方法で得られる測定値はいずれも同一の確率分布とパラメータに従い、相関はないものと仮定した。

SAS コード 1. サンプルデータ作成

```
/* サンプルサイズ、パラメータ設定 */
%let mu = 1;
%let sigma = 0.75;
%let n = 10000;

/* サンプルデータの作成 */
data BA_Plot_Data;
  call streaminit(123);
  do i = 1 to &n;
    /* 正規分布に従う乱数生成 */
    methodA = rand("Normal", &mu, &sigma);
    methodB = rand("Normal", &mu, &sigma);

    /*対数正規分布に従うよう指数変換*/
    exp_methodA = exp(methodA);
    exp_methodB = exp(methodB);
```

```
    /* 正規分布に従う2つの測定値の差*/
    normal_diff = methodA - methodB;

    /* 対数正規分布に従う2つの測定値の差の平均*/
    exp_diff = exp_methodA - exp_methodB;
    exp_mean = (exp_methodA + exp_methodB) / 2;

    /*測定値の差が0の参照線を引く用の変数*/
    ref = 0;
    output;
  end;
run;
```

次いで SAS コード 2.により、正規分布及び対数正規分布に従う 2 つの測定値の差それぞれに対して標準偏差を計算した。以降、正規分布に従う測定値の差の標準偏差を$SD_{\text{normal_diff}}$、対数正規分布に従う測定値の差の標準偏差を$SD_{\text{exp_diff}}$と表すこととする。

SAS コード 2. $SD_{\text{normal_diff}}$及び$SD_{\text{exp_diff}}$の計算

```
/*正規分布に従う2つの測定値の差のSD*/
proc summary data = BA_Plot_Data noprint;
  var normal_diff;
  output out = normal_summary(drop = _:) std =std_normal_diff;
run;

/*対数正規分布に従う2つの測定値の差のSD*/
proc summary data = BA_Plot_Data noprint;
  var exp_diff;
  output out = exp_summary(drop = _:) std =std_exp_diff;
run;
```

次いで SAS コード 3.により、従来法及び Euser らの方法に基づく LoA を計算してサンプルデータに持たせた。1 つ目のデータステップにて、Euser らの方法に基づく LoA の計算式のうち、$(\exp(a)-1)/(\exp(a)+1)$の項を計算し、coef というマクロ変数に持たせた。2 つ目のデータステップで、従来法に基づく LoA を計算した。最後に 3 つ目のデータステップで、従来法及び Euser らの方法に基づく LoA の下限、及び上限をそれぞれ計算してサンプルデータに持たせた。

SAS コード 3. 従来法及び Euser らの方法に基づく LoA の計算

```
data _NULL_;   set normal_summary;
   *Euserらの方法に基づくLoAにおけるa(=1.96×SD_diff);
   normal_a = 1.96 * std_normal_diff;

   /*(exp(a)-1) / (exp(a)+1)の計算*/
   Coef = (exp(normal_a)-1) / (exp(normal_a)+1);
   call symputx("Coef", Coef);
run;

data _NULL_;   set exp_summary;
/*従来法に基づくLoAの計算*/
   exp_a = 1.96 * std_exp_diff;
   call symputx("exp_a", exp_a);
run;

data BA_Plot_Data;   set BA_Plot_Data;
   Normal_LoA_Lower = -&exp_a.; *従来法に基づくLoAの下限;
   Normal_LoA_Upper =  &exp_a.; *従来法に基づくLoAの上限;

   Euser_LoA_Lower = -2 * exp_mean * &Coef. ; *Euserらの方法に基づくLoAの下限;
   Euser_LoA_Upper =  2 * exp_mean * &Coef. ; *Euserらの方法に基づくLoAの上限;
run;
```

最後に、SAS コード 4.により Bland-Altman Plot 及び従来法、Euser らの方法に基づく LoA を描画した。従来法に基づく LoA は、測定値の平均値に依らず一定であるため、REFLINE 文を用いて描画した。一方で Euser らの方法に基づく LoA は、測定値の平均値に依存するため REG 文を用いて描画した。

SAS コード 4. Bland-Altman Plot と LoA の描画

```
proc sgplot data=BA_Plot_Data noautolegend;
   /*軸ラベルの設定*/
   yaxis label="測定値の差" valueattrs = (size = 14) labelattrs = (size = 14) ;
   xaxis label="測定値の平均値" valueattrs = (size = 14) labelattrs = (size = 14) ;;

   /*Bland-Altman Plotの描画*/
   scatter x=exp_mean y=exp_diff /
   markerattrs=(symbol=circlefilled color=black);
```

```
*差=0の参照線を描画;
refline ref / axis=y lineattrs=(color=black pattern=solid);

/*従来法に基づくLoAの描画*/
refline Normal_LoA_Lower / axis=y lineattrs=(color=black thickness = 2 pattern=dot);*下限;
refline Normal_LoA_Upper / axis=y lineattrs=(color=black thickness = 2 pattern=dot);*上限;

/*Euserらの方法に基づくLoAの描画*/
reg y = Euser_LoA_Lower x = exp_mean /
        lineattrs = (color = black thickness = 2 pattern=dash) nomarkers; *下限;
reg y = Euser_LoA_Upper x = exp_mean /
        lineattrs = (color = black thickness = 2 pattern=dash) nomarkers; *上限;
run;
```

図 6 に SAS コード 4.の実行結果を示した。従来法による LoA は点線、Euser らの方法による LoA は破線で描画した。従来法に基づく LoA は測定値の平均値に依存せず一定であるため、対数正規分布に従う測定値に適用すると、平均値の増加に伴い LoA 外のプロットが増加することが確認された。一方、Euser らの方法に基づく LoA は、測定値の平均値の増加に応じて幅が拡大するため、比例誤差を有する測定値に対して適用しても、平均値に依存せず一定の割合でプロットが LoA 内に収まることが視覚的に確認できた。

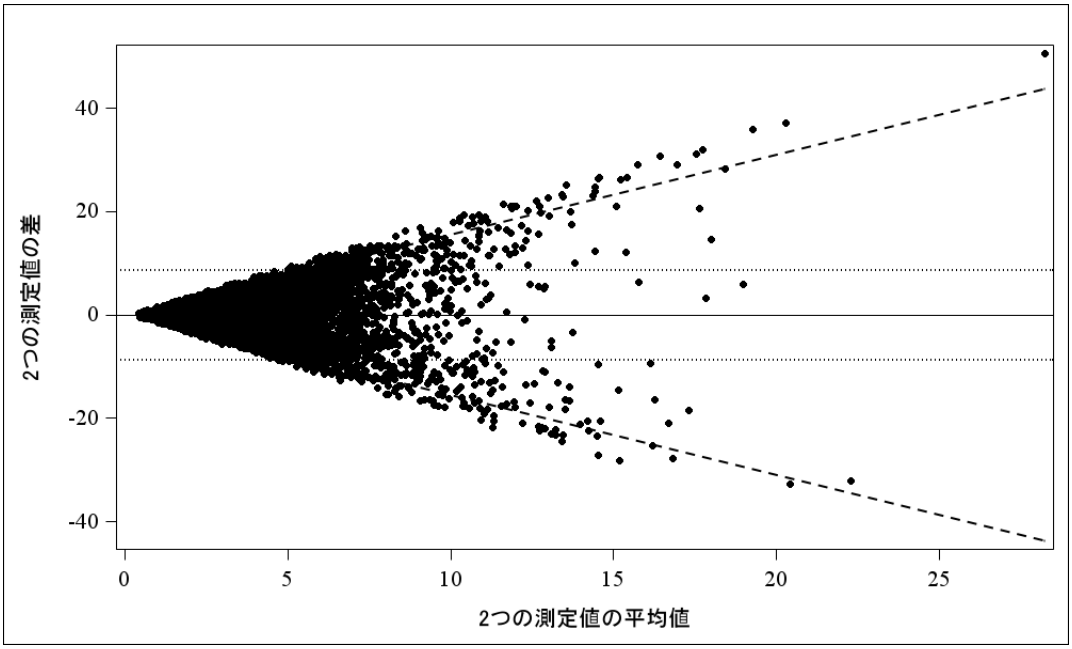


図 6. 従来法（点線）及び Euser らの方法（破線）に基づく LoA の描画

さらに、図 6 で示した結果について、視覚的な確認に加えて数値的な検証も実施した。作成に用いたサンプルデータを基に、LoA 内に含まれるプロット数について、測定値の平均値に基づく四分位を用いて 4 つのカテゴリ（Min ≤ < Q1、Q1 ≤ < Median、Median ≤ < Q3、Q3 ≤ ≤ Max）に分割した上で頻度集計を実施した。

また、データ全体についても同様の集計を実施した。カテゴリ作成の基準となる四分位値、最小値、最大値を表 1.に、集計結果を表 2.に示した。

表 1. カテゴリ作成の基準値

最小値 (Min)	第一四分位 (Q1)	中央値 (Median)	第三四分位 (Q3)	最大値 (Max)
0.432	2.125	3.087	4.494	28.231

表 2. LoA 内に含まれたプロットの頻度集計

LoA の計算方法	測定値の平均のカテゴリ				合計 (N=10000)
	Min ≤ < Q1 (N=2500)	Q1 ≤ < Median (N=2500)	Median ≤ < Q3 (N=2500)	Q3 ≤ ≤ Max (N=2500)	
従来法	2500 (100.0%)	2500 (100.0%)	2500 (100.0%)	1947 (77.9%)	9447 (94.5%)
Euser らの方法	2472 (98.9%)	2435 (97.4%)	2374 (95.0%)	2212 (88.5%)	9493 (94.9%)

表 2.の集計結果より、全体ではいずれの方法においても LoA 内に 95%のプロットが含まれていた。一方、測定値の平均値のカテゴリ別に確認すると、従来法に基づく LoA では、Median ≤ < Q3 までは全てのデータが LoA 内に含まれていたが、Q3 ≤ ≤ Max では LoA 内に含まれるプロットの割合が急減した。Euser らの方法に基づく LoA では、Min < Q3 において安定して 95%以上のプロットが LoA 内に含まれており、Q3 ≤ ≤ Max では従来法と同様に 95%未満となったものの、従来法と比較して LoA 内に含まれるプロットの割合は高い結果であった。

4. まとめ

本稿では、まず Bland-Altman Plot 及び従来の LoA 計算法の概要を紹介し、次いで従来の LoA 計算法の問題点について言及し、最後に Euser らの LoA 計算法とその SAS による実装方法を紹介した。

Bland-Altman Plot では目的に応じて、2 つの評価方法により得られた測定値の一致性、系統・比例誤差の有無、測定値の差のばらつきに関する検討が可能である。測定値が従う確率分布の性質次第で、従来法に基づく LoA により、各評価方法で得られた測定値の関係について誤った解釈を与える可能性がある。

Bland-Altman Plot 及び LoA を用いる際は、対象の評価項目が従う確率分布等のデータ特性について、先行研究の調査や専門家と事前に十分協議したうえで、どの計算・描画方法を選択するかを慎重に取り決める必要がある。

5. 参考文献

[1] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476): 307-10.

[2] Kottner J, Audigé L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64:96–106.

[3] Anne M. Euser, Friedo W. Dekkera, Saskia le Cessie. A practical approach to Bland-Altman plots and variationcoefficients for log transformed variables. Journal of Clinical Epidemiology 61 (2008).

連絡先:
〒112-0002 東京都文京区小石川 1-4-1 住友不動産後楽園ビル
エイツーヘルスケア株式会社　データサイエンスソリューション部
E-mail: tokuda-y@a2healthcare.com

Appendix. 対数正規分布に従う測定値の差に対する LoA の構成方法

1. 前提条件の設定

それぞれ正規分布$N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$に従う変数$Z_1, Z_2$を考える。それぞれにネイピア数で指数変換した変数を

$$X_1 = \exp(Z_1), X_2 = \exp(Z_2)$$

と定義する。このときX_1, X_2は対数正規分布に従う。

また正規分布に従う変数の差$Z_1 - Z_2$の標準偏差をSD_{diff}とし、

$$a = 1.96 \times SD_{diff}$$

とおく。

2. 正規分布に従う変数に対する LoA の計算

正規分布に従う 2 つの変数の差$Z_1 - Z_2$に対して、Bland-Altman Plot で描画される従来法に基づく LoA は

$$-a \le Z_1 - Z_2 \le a \tag{1}$$

で与えられる。

式(1)の各項にネイピア数による指数変換を適用すると、

$$\exp(-a) \le \exp(Z_1 - Z_2) \le \exp(a) \tag{2}$$

を得る。

ここで(2)式の中央の項$\exp(Z_1 - Z_2)$は

$$\frac{\exp(Z_1)}{\exp(Z_2)} = \frac{X_1}{X_2}$$

と表せることから、式(2)は

$$\exp(-\mathrm{a}) \le \frac{X_1}{X_2} \le \exp(a) \tag{3}$$

となる。

3. 比率表記と式変形

ここで

$$\frac{X_1}{X_2} = r$$

とおくと、X_1, X_2の関係式は

$$X_1 = rX_2, \ X_2 = \frac{X_1}{r} \tag{4}$$

と表される。更にX_1, X_2の平均を

$$\bar{X} = \frac{X_1 + X_2}{2} \tag{5}$$

と定義し、式(4)を式(5)に代入すると、

$$X_1 = 2\bar{X}\left(\frac{r}{r+1}\right), \ X_2 = 2\bar{X}\left(\frac{1}{r+1}\right) \tag{6}$$

を得る。従ってX_1とX_2の差は

$$X_1 - X_2 = 2\bar{X}\left(\frac{r-1}{r+1}\right) \tag{7}$$

と表せる。

4. 比(X_1/X_2)の LoA から差$(X_1 - X_2)$の LoA への変換

式(3)より、rの LoA は

$$r \in [\exp(-a), \exp(a)] \tag{8}$$

である。

ここで$a \geq 0$及び$\exp(\cdot)$が単調増加することより、 rの LoA の下限及び上限を式(7)にそれぞれ代入することで、対数正規分布に従う 2 つの変数の差$(X_1 - X_2)$に対応する LoA の下限及び上限が得られる。

<下限の場合>

$r = \exp(-a)$を式(7)に代入すると

$$(X_1 - X_2)_{LoA\ Lower} = 2\bar{X}\left(\frac{\exp(a)+1}{\exp(a)-1}\right) = -2\bar{X}\left(\frac{\exp(a)-1}{\exp(a)+1}\right) \tag{9}$$

を得る。

<上限の場合>

同様に$r = \exp(a)$を式(7)に代入すると

$$(X_1 - X_2)_{LoA\ Upper} = 2\bar{X}\left(\frac{\exp(a)-1}{\exp(a)+1}\right) \tag{10}$$

を得る。

5. 式の整理

式(9)及び(10)より、対数正規分布に従う確率変数X_1, X_2の差の LoA は

$$X_1 - X_2 \pm 2\bar{X}\frac{(\exp(a)-1)}{(\exp(a)+1)} \tag{11}$$

となり、これは 3.1 で示した Euser らの方法に基づく LoA の記載に一致する。

NLPプロシージャを用いた入れ子型ロジットモデルにおけるパラメータ推定に関する一考察

○坂巻英一[1]

([1]国立大学法人宇都宮大学)

A Study on Parameter Estimation in Nested Logit Models Using NLP Procedures

Yoshikazu, Sakamaki
School of Data Science and Management, Utsunomiya University

要旨

消費者の選択行動に関するモデル化の分野では多項ロジットモデルを始めとした様々な確率モデルが先行研究において研究されてきた．このうち入れ子型ロジットモデルは選択肢の類似性をモデル内で考慮することが可能であることから，消費者の選択集合把握をはじめ様々な分野で活用が進められている．一方で入れ子型ロジットモデルの尤度関数は多峰性であることがかねてより指摘されており，パラメータの推定値が局所最適解に収束してしまうといった理由により一意に定まらないことがモデルを扱いにくくする大きな要因の一つになっていた．こうした課題を改善するために本研究では入れ子型ロジットモデルの尤度関数に損失関数の概念を導入することによりパラメータ推定を安定的に行うことを目的としたパラメータ推定法のい改善を提案する．併せて提案手法を実データに適用し NLP プロシージャを用いてパラメータを推定することにより，本研究における提案モデルが入れ子型ロジットモデルにおけるパラメータ推定を行う上で有効であることを示す．

キーワード：Nested Logit Model, Multimodal Likelihood Function, NLP procedure

1 先行研究紹介

消費者の選択行動に関するモデル化において古くから使用されている手法のひとつに多項ロジットモデルが挙げられる。多項ロジットモデルはニューラルネットワークにおける出力層にも利用されており機械学習において重要な役割を果たしている．一方で、赤バス・青バスの問題にみられるようにモデル内で扱われる選択肢が互いに独立である、という仮定が置かれているため、実務において扱いにくさが指摘されてきた[1]。こうした中、Ben-Akiva らはモデル内で扱われる選択肢の類似性をモデル内で許容することを可能にした入れ子型ロジットモデルを提案している [1]. 入れ子型ロジットモデルは交通手段の選択をはじめ消費者の選択集合のモデル化に関する研究において盛んに利用されている．選択集合とは消費者が商品やサービスを購買する際に，最初に購買の候補とする選択肢群のことであり，英語では Evoked Set, Consideration Set 或いは Choice Set と訳されている [2]. 消費者が商品やサービスを選択

する際に考慮する選択肢の数は製品カテゴリーによって異なるもののその数は多くはなく，一般に 2 個から 5 個程度であるといわれている．そのため，技術の標準化が進み差別化が難しくなる今日の消費者市場において自社のブランドを消費者の選択集合に入れることができるか否か，が強いブランドを作り上げる上で重要な要素のひとつとなっている [3][4].

先行研究において行われてきた選択集合を考慮した消費者選択行動のモデル化の多くは Manski が示す段階的ブランド選択モデルを基礎として行われてきた [5]. Manski は

i　選択肢番号(i=1,2,…,I)

k　消費者番号(k=1,2,…,K)

p_{ik}　消費者 k が選択肢 i を選択する確率

$p_k(i|C)$　消費者 k が選択集合 C の中から選択肢 i を選択する確率

$p_k(C|G)$　消費者 k が全ての可能な選択集合の組み合せ G の中から特定の選択集合 C を形成する確率

$G(i)$　G のうち選択肢 i を含む全ての集合

とした時，消費者 k が選択肢 i を選択する確率 p_{ik} を

$$p_{ik} = \sum_{C \in G(i)} p_k(i|C) p_k(C|G) \tag{1}$$

により定式化している.

先行研究を概観すると多くの研究において$p_k(i|C)$は多項ロジットモデルを用いた定式化が行われており選択集合 C が形成される確率$p_k(C|G)$を如何にして定式化するか，が主たる研究課題とされてきた.

先行研究の多くは (2)式により$p_k(C|G)$を定式化している.

$$p_k(C|G) = \frac{\exp\{\alpha W_C + \lambda_C \log[\sum_{j \in C} \exp(V_{jk})]\}}{\sum_G \exp\{\alpha W_C + \lambda_C \log[\sum_{j \in C} \exp(V_{jk})]\}} \tag{2}$$

ただし，

i　選択肢番号(i=1,2,…,I)

k　消費者番号(k=1,2,…,K)

V_{ik}　消費者 k の選択肢 i に対する確定的効用

λ_C　選択集合 C により形成される log_sum 項$\log[\sum_{i \in C} \exp(V_{ik})]$に対するパラメータ

W_{Ch}　log_sum 項$\log[\sum_{i \in C} \exp(V_{ik})]$自体が持つ h 番目の属性の属性値 (h=1,2,…,H)

α_h　W_{Ch} に対するパラメータ (h=1,2,…,H)

である．また，

l　個人属性番号(l=1,2,…,L)

m　選択肢属性番号(m=1,2,…,M)

x_{ikm}　消費者 k の選択肢 i に対する m 番目の属性値

β_m　m 番目の属性値に対する推定パラメータ (m=1,2,…,M)

z_{kl}　消費者 k が持つ l 番目の個人属性の属性値

γ_l　l 番目の個人属性の属性値に対する推定パラメータ (l=1,2,…,L)

とする．その上で，消費者 k が選択肢 i に対して持つ確定的効用を

$$V_{ik} = \sum_{m=1}^{M} \beta_m x_{ikm} + \sum_{l=1}^{L} \gamma_l z_{kl} \tag{3}$$

により定式化している.

また$p_k\ (i|C)$については多項ロジットモデルを用いて(4)式により定式化している.

$$p_k\ (i|C) = \frac{exp(V_{ik})}{\sum_{i \in C} exp(V_{ik})} \tag{4}$$

(2)式及び(4)式を基に消費者 k が選択肢 i を選択する確率を

$$p_{ik} = \sum_{C \in G(i)} p_k(i|C) p_k(C|G) \tag{5}$$

により定式化することになる.

ここで y_{ik} を消費者 k が選択肢 i を選択した時 1, そうでない時 0 をとる 2 値変数とした場合 y_{ik} とそれに対応するp_{ik} が与えられた基で尤度は

$$\begin{aligned} L &= \prod_i \prod_k {p_{ik}}^{y_{ik}} \\ &= y_{ik} \sum_i \sum_k \log\left(p_{ik}\right) \end{aligned} \tag{6}$$

と定式化できる. 先行研究では最尤推定法を用い, (6)式の尤度を最大にするようなモデルパラメータを推定することによりパラメータ推定が行われてきた.

2　本稿における提案手法

入れ子型ロジットモデルの尤度関数は多峰性の形状をしていることが知られている. そのためパラメータの最適解を探索するためには, 乱数を用いて初期値を多数回発生させた上で繰り返し計算を行い, 最適解を決定する必要がある. そのため計算に膨大な時間を要する上に, パラメータ推定結果が不安定であることから, 実務的な利用が進まなかったという背景がある.

本研究ではこうした課題を改善することを目的として lasso 回帰において広く使用されている損失関数

$$\delta_2 \sum \left|\theta_j\right| \tag{7}$$

を尤度関数に組み込むことを試みる. 以降、(7)式を尤度関数に組み込むことを L1 正則化と呼ぶこととする。更に本研究では Yuan ら[6]の group lasso を拡張した上で, 以下の手順に基づき, 入れ子型ロジットモデルにおける尤度関数の改善を試みる.

[手順1] データ分割

分析に先立ち, データセットを学習用データ (train data)と検証用データ (test data)に分割する. その上で学習用データ (train data)を用いて[手順2]から[手順6]を実施する.

[手順2]　選択肢に対する選択確率の定式化

次に第 2 節で示す入れ子型ロジットモデルを用いて消費者の選択肢に対する選択確率を定式化する.

[手順3]　損失関数の導入

Group Lasso の考え方に倣い L2 正則化における損失関数を(8)式のように拡張する.

$$\delta_1 \sum_{g=1}^{G} \sqrt{\sum_{j=1}^{J_g} {\theta_{gj}}^2} \tag{8}$$

ここで

g　　グループ番号 (g =1,2,…,G)

j　　グループ g に含まれる変数の変数番号 (j =1,2,…,J_g)

θ_{gj}　　グループ g 内における j 番目の変数の値

である. 以降、(8)式を尤度関数に組み込むことを L2 正則化と呼ぶこととする。

[手順4] 尤度関数の定式化

(6)式により定式化される尤度関数から(7)式及び(8)式を引くことにより尤度関数を(9)式のように拡張する.

$$L = y_{ik} \sum_{i} \sum_{k} \log\left(p_{ik}\right) - \delta_1 \sum_{g=1}^{G} \sqrt{\sum_{j=1}^{J_g} {\theta_{gj}}^2} - \delta_2 \sum \left|\theta_j\right| \tag{9}$$

[手順5] パラメータ推定

乱数発生させた初期値を基に(9)式に対して最尤推定法によるパラメータ推定を行う.

[手順6] モデル適合度の算出

[手順5]において推定されたパラメータを検証用データ (test data)に適用し, 対数尤度, AIC 統計量及び予測の的中率を算出する.

3　実データを用いた実証実験

3.1 データ概要

本研究ではデータサイエンス学習者向けに Google によって運営されているプラットフォーム Kaggle 上で公開されているデータを使用した. データの概要は以下の通りである.

データの出典元　Kaggle

データセット名　Great American Coffee Taste Test

URL　https://www.kaggle.com/datasets/datalab351/great-american-coffee-taste-test

本データは米国で実施されたコーヒーの味覚調査によって収集されたデータである. 調査実施者はコーヒーカプセルに封入された 4 種類のコーヒーを 5,000 人の被験者に配布する. 被験者は受け取ったコーヒーカプセルを溶かした上で試飲し, アンケート調査に回答する方法がとられている. 調査は被験者にコーヒーの種類を伏せた形で実施している.

アンケート調査には 4 種類のコーヒーに対する評価に加え, 4 種類のコーヒーのうち最も好きな風味, 被験者の個人属性を回答する形で実施された.

データ収集時期　2023 年 10 月

本研究では 5,000 人の被検者が試飲した 4 種類のコーヒーに対する評価データを説明変数, 被検者が最も好むコーヒーを選択結果としたモデルを構築することにより検証を実施した.

評価対象となった 4 種類のコーヒーの風味は表 1 の通りである.

表1　評価対象となった 4 種類のコーヒーの風味

Package No	風　　味
A	浅煎り、ウォッシュド加工
B	ミディアムロースト
C	ダークロースト
D	ライトロースト、ナチュラルプロセス（発酵、ファンキー/フルーティー）

3.2 基礎統計量

アンケート調査により収集されたデータの中に，調査対象となった 4 種類のコーヒーに対して，苦味及び酸味という 2 つの視点から被検者が 5 段階で評価したデータが含まれている．4 種類のコーヒーに対する評価の平均値を表 2，ポジショニングマップを図 1 に示す．

表 2　4 種類のコーヒーに対する評価の平均値

変数	苦味	酸味
COFFEE_A	2.140	3.638
COFFEE_B	3.012	2.222
COFFEE_C	3.073	2.365
COFFEE_D	2.162	3.859

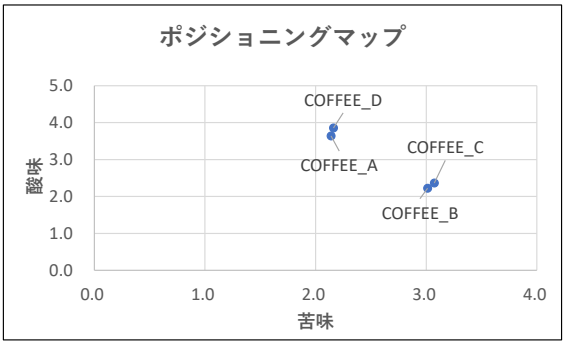


図 1　4 種類のコーヒーのポジショニングマップ

分析を行った結果，COFFEE A と COFFEE D 及び COFFEE B と COFFEE C は苦味及び酸味の観点から相互に類似性が高いと考えられる．

分析に使用したデータには 4 種類のコーヒーの中から最も好きなコーヒーを選択してもらった結果が記録されている．表 3 にそれぞれの選択肢を最も好きなコーヒーとして選択した被検者数をまとめて示す．

表 3　最も好きなコーヒーとして選択した被検者の数

コーヒーの種類	最も好きと回答した人数	構成比率
Coffee A	818	21.7%
Coffee B	783	20.8%
Coffee C	784	20.8%
Coffee D	1385	36.7%
合　　計	3770	100.0%

本研究で使用したデータにはコーヒーに対する評価に加え，個人属性に関する情報が含まれている．

3.3　　本研究における提案モデルを用いたデータ分析

これらのデータを基に[手順1]から[手順 6]の流れで分析を実施した.

[手順1]　データ分割

始めに 3.1 項で示したデータに対して未回答等の理由により欠損値を含むデータを除いた上で, 残ったデータ 3,770 件をランダムに 6 : 4 の割合で分割し, 6 割を「学習用データ」4 割を「検証用データ」として使用した.

本研究において「学習用データ」「検証用データ」として使用した被検者の人数は以下の通りである.

学習用データ	2,262 人
検証用データ	1,508 人
計	3,770 人

これらのデータを基に「学習用データ」を用いてパラメータ推定を実施した上で, 推定されたパラメータを「検証用データ」に適用する方法によりモデルの有効性に関する検証を実施した.

[手順2]　選択肢に対する選択確率の定式化

本研究では以下の変数を分析モデルに使用した.

x_{ik1} 被検者 k が感じるコーヒーi の苦味 (5 段階で評価)

(k=A, B, C, D)

x_{ik2} 被検者 k が感じるコーヒーi の酸味 (5 段階で評価)

(k=A, B, C, D)

z_{k1} 自分が飲んでいるコーヒーの生産地を知っているか

(Yes / No / 欠損値等)

ただし欠損値を基準として Yes / No に対しダミー変数を割り付ける.

z_{k2} 性別 (男性 / 女性 / 欠損値等)

ただし欠損値を基準として男性 / 女性 に対しダミー変数を割り付ける.

W_C 選択集合 C を構成する選択肢数

y_{ik} 被検者 i の選択肢 k に対する選択結果

4 種類のコーヒーのうち最も好きな風味を持つコーヒーに対してフラグ 1 を付与

また以下のパラメータをモデル内で使用する.

β_1　x_{ik1}に対する推定パラメータ

β_2　x_{ik2}に対する推定パラメータ

γ_{11}　z_{k1}に対する推定パラメータ(回答=YES)

γ_{12}　z_{k1}に対する推定パラメータ(回答=NO)

γ_{21}　z_{k2}に対する推定パラメータ(回答=男性)

γ_{22}　z_{k2}に対する推定パラメータ(回答=女性)

α　W_Cに対するパラメータ

λ_C　選択集合 C により形成される log_sum 項に対するパラメータ

これらの変数及びパラメータを基に(3)式により選択肢に対する効用を定式化した上で(4)式及び(5)式を基にした入れ子型ロジットモデルを用いて選択肢に対する選択確率を定式化する.

[手順3]　損失関数の定式化

(7)式及び(8)式により損失関数を定式化する. なお本研究では以下のようにパラメータ群をグループ化した上でパラメータの推定を行うこととする.

グループ 1　　β_1 β_2

グループ 2　　γ_{11} $\gamma_{12}\gamma_{21}$ γ_{22}

グループ 3　　α

グループ 4　　λ_C

$^{\forall}$C∈{A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, ABCD}

この結果，L2 正則化による損失関数は

$$\delta_1\left(\sqrt{\sum_{j=1}^{2}{\beta_j}^2}+\sqrt{\sum_{j=1}^{2}\sum_{l=1}^{2}\gamma_{jl}^2}+\sqrt{\alpha^2}+\sqrt{\sum_{j\in C}{\lambda_j}^2}\right) \tag{10}$$

により定式化されることになる．

また L1 正則化による損失関数の定式化は

$$\delta_2\left(\sum_{j=1}^{2}|\beta_j| \quad +\sum_{j=1}^{2}\sum_{l=1}^{2}|\gamma_{jl}|+|\alpha| \quad +\sum_{j\in C}^{2}|\lambda_j| \quad\right) \tag{11}$$

によって定式化する．

[手順4]　尤度関数の定式化

(10)式及び(11)式を(6)式に組み込むことにより尤度関数を(12)式により定式化する．

$$\begin{aligned} L=&\prod_i\prod_k\prod_t {p_{ik}^t}^{y_{ik}^t}\\ &-\delta_1\left(\sqrt{\sum_{j=1}^{2}{\beta_j}^2}+\sqrt{\sum_{j=1}^{2}\sum_{l=1}^{2}\gamma_{jl}^2}+\sqrt{\alpha^2}+\sqrt{\sum_{j\in C}^{2}{\lambda_j}^2}\right)\\ &-\delta_2\left(\sum_{j=1}^{2}|\beta_j| \quad +\sum_{j=1}^{2}\sum_{l=1}^{2}|\gamma_{jl}|+|\alpha| \quad +\sum_{j\in C}^{2}|\lambda_j| \quad\right)\Bigg\} \end{aligned} \tag{12}$$

[手順5]　正則化パラメータ

(12)式により定式化される尤度関数における正則化パラメータを

δ_1: 0.2

δ_2: 1.0

に設定する．

[手順6]　パラメータ推定

[手順5]において設定された正則化パラメータを(12)式で示す尤度関数に設定した上で「学習用データ」を基にパラメータ推定を実施する．推定は初期値を乱数発生させた上で繰り返し計算を実施する方法により行い，本研究では繰り返し計算を 100 回実施した．その上で，最大の対数尤度を与えるパラメータの組み合わせを本研究における提案モデルにおける最適解とみなすこととした．また本研究では分析環境として SAS システムに搭載されている NLP プロシージャを使用している．探索アルゴリズムには準ニュートン法を使用し，繰り返し計算回数を 5,000 回，勾配が 1×10-6 以下であることを停止条件として設定した．

パラメータの推定結果を表 5 にまとめて示す．

表 5　パラメータ推定結果(本研究における提案モデル)

本研究における提案モデル

変数	N	最適解	平均	標準偏差	最小値	最大値
β1 (苦味)	100	-0.1570	-0.2706	0.1085	-0.5846	-0.1183
β2 (酸味)	100	-0.0370	-0.0789	0.0465	-0.2545	-0.0270
α (選択集合を構成する選択肢数)	100	0.2007	0.2971	1.9926	-12.2221	7.1161
そのコーヒーの生産地を知っている (はい)	100	1.0012	2.5810	3.0114	-7.0842	9.9485
そのコーヒーの生産地を知っている (いいえ)	100	-0.6049	-0.8751	5.9340	-11.5087	28.0158
性別 (男性)	100	0.5561	0.3641	2.6917	-7.6232	6.1058
性別 (女性)	100	-0.0400	-0.0554	2.5487	-5.7556	15.0581
λ_A	100	2.4487	1.5953	0.9468	-0.1658	6.3361
λ_B	100	-0.7168	-0.3880	1.0966	-1.8931	8.0373
λ_C	100	-0.1131	-0.1507	0.9729	-1.5502	6.9209
λ_D	100	6.0882	2.9181	2.2617	-0.6895	7.5071
λ_{AB}	100	2.6875	1.2241	1.1885	-1.7744	5.7130
λ_{AC}	100	2.7704	1.4028	1.1544	-1.3279	3.9842
λ_{AD}	100	3.9733	1.9268	1.6940	-1.5163	4.7801
λ_{BC}	100	3.7084	1.9300	1.5429	-2.3099	5.0291
λ_{BD}	100	1.2162	0.2780	0.9914	-2.1156	6.7943
λ_{CD}	100	2.4976	0.7210	1.3112	-1.0333	4.2186
λ_{ABC}	100	-3.4374	-0.4284	2.0643	-3.4379	3.5152
λ_{ABD}	100	-0.0149	0.3734	0.9046	-2.0696	4.4237
λ_{ACD}	100	-0.2763	0.0128	1.1044	-2.3781	5.7299
λ_{BCD}	100	2.9226	1.7607	1.4824	-4.7657	9.2285
λ_{ABCD}	100	-2.3298	-0.7606	1.9553	-4.0785	9.6791
対数尤度	-1,956.8					
AIC	3,936					
予測の的中率	38.99%					

4　結論

本稿では NLP procedure を用いて格付推計モデルを構築する方法について紹介を行った。本研究では消費者選択行動を記述するモデルのひとつとして古くから利用されてきた入れ子型ロジットモデルにおけるパラメータ推定アルゴリズムの改善に関する研究を実施した．多項ロジットモデルは伝統的に利用されている多項ロジットモデルと比較して選択肢間の類似性を考慮することが可能であることから，企業経営のみならず交通工学における交通手段の選択行動をモデル化する場面等で広く利用されてきた．一般にパラメータ推定は尤度関数を設定した上で最尤推定法を用いて最適解を探索する，という方法が取られることが多い反面，尤度関数が多峰性の形状を持つことが先行研究において指摘されており，パラメータを一意に決定することができず推定結果が不安定になることから，実務において利用されるケースはあまり報告されて来なかった，というのが現状である．

本研究では入れ子型ロジットモデルが持つこうした課題を改善することを目的として，ridge 回帰や lasso モデルで用いられているパラメータの正則化による損失関数を尤度関数に組み込むことにより入れ子型ロジットモデルにおけるパラメータ推定を安定的に実施するとしたパラメータ推定アルゴリズムの改善方法を提案した．

本研究における提案モデル並びに比較対象モデルを実データに適用しパラメータ推定を実施した結果，対数尤度，AIC 統計量並びに予測の的中率に関しては双方のモデル間で大きな差異は確認されなかったものの，本研究において提案するアルゴリズムではパラメータ推定値のばらつきを従来のアルゴリズムと比較して小さく抑えることが可能となることを確認することができた．

損失関数を考慮しない従来の尤度関数に基づくパラメータ推定アルゴリズムでは繰り返し計算が収束せず発散してしまうケースが起こり得る，パラメータ推定結果が外れ値に近い局所最適解に陥ること

からばらつきが大きくなる，という課題が生じていたが，本研究における提案モデルによりこうした課題の影響を抑制することが可能であることが分かった．

また，パラメータ推定値と対数尤度の関係をグラフ化した結果，入れ子型ロジットモデルの尤度関数が従来考えられてきた峰型の形状ではなく尾根型に近い形状をしている可能性があることが分かった．そのため尤度を最大化することを可能にするパラメータの組み合わせが無数に存在する可能性があった．そこで正則化による制約条件を課すことによりパラメータ推定を安定的に実施できる可能性があることが実証実験の結果から分かった．

米国市場におけるコーヒーに対する消費者の選択行動に関して，選択集合を構成する選択肢の数に対するパラメータ α がプラスの符号をとることから，選択集合が増えるに従い消費者の効用が上昇する可能性があることが分かった．また，女性よりも男性の方がコーヒーに対して高い効用を持つこと，コーヒーの生産地に関する知識を持ち合わせている消費者の方がそうではない消費者と比較して高い効用を持つ可能性があること，等が分析を実施した結果から明らかになった．

以上

参考文献

[1] Ben-Akiva M. E. and Steven R. L.: Discrete Choice Analysis: Theory and Application to Travel Demand, The Mit Press (1985).

[2] Shocker, A. D., Ben-Akiva, M., Boccara, B. and Nedungadi, P.: Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions, Marketing Letters, Vol. 2, No. 3, pp.181–197 (1991).

[3] Hauser J. R. and Wernerfelt B.: An Evaluation Cost Model of Consideration Sets, Journal of Consumer Research, Vol. 16, pp. 393-408 (1990).

[4] Brisoux J. E. and Cheron E. J.: Brand Categorization and Product Involvement, Advances in Consumer Research, Proceedings of the Association for Consumer Research Annual Conference, pp. 101-109 (1990).

[5] Manski, C.: The Structure of Random Utility Models, Theory and Decision, Vol. 8, pp. 229-254 (1977).

[6] Yuan, M. and Lin, Y.: Model Selection and Estimation in Regression with Grouped Variables, Journal of the Royal Statistical Society. Series B (statistical Methodology), Vol. 68, No. 1, pp.49-67 (2006).



SAS周辺で使用される文字コードの取り扱いについて

○兵頭 照将

（イーピーエス株式会社）

Handling of Character Encodings in Relation to SAS

Terumasa Hyodo

EPS Corporation

要旨

電子データを取り扱う場合に、どうしても避けて通れないのが文字コードである。近年では、多くのプログラム言語が Unicode の特に UTF-8 を採用しており、事実上の標準となっている。しかし、過去資産の取り扱いも含めまだ SAS においては Unicode の取り扱いに苦労している。また、SAS と他のツール間での文字コードの違いから、トラブルになる事もある。

SAS プログラムやそれを利用する周辺のツール（Windows や Linux、また Office 製品等）での文字コードの取り扱いについて、過去事例また今後発生しうる事象について説明する。

キーワード：文字コード、Unicode、NLS Support、多言語対応

文字コードについての概説

文字コードは、簡易な表現を用いると、我々が日常的に使用している文字のうちの一部を表とし、それを特定のルールに基づき電子データにする方法となる。

この電子データにする対象の文字の一覧（及びそれに記された識別番号）を、「符号化文字集合」と呼び、その「符号化文字集合」を電子データとして変換する方式を「文字コード」と以下呼ぶ。

Unicode では「文字コード」の部分を「文字符号化形式」と「シリアライズの方式」（バイトの並べ方）の二段階に分けて呼ぶ。

以下、日常の業務でしばしば遭遇する「文字コード」の名称を列記する。

- Shift_JIS
- Unicode
  - UTF-8
  - UTF-16 LE
  - UTF-16 BE
- ISO-2022-JP
- EUC-JP
- ASCII
- ISO 8859-1

Shift_JIS は事実上、Windows の日本環境ではテキストファイルの保存形式として使われ、馴染みが一番深い文字コードである。Unicode は日本語だけではなく複数の言語に対応した最も新しく最大の規格であり、Windows では UTF-16 LE、Linux では UTF-8 がよく用いられている。ISO-2022-JP は、通信の領域では未だに用いられており、メールソフトではしばしば使われている（ただし、文字そのものが露出している事が少なく、BASE64 等と合わせて用いられている事が多い為、直接我々が見る事はほぼない)。EUC-JP については UNIX、初期の Linux で用いられていた文字コードであるが、歴史的に存在するが日常的に見かけなくなっている。

旧来の文字コードについて

SBCS(Single Byte Character Set)と DBCS(Double Byte Character Set)

業務上、SBCS については、ASCII7 ビットまでの数値で表される文字コードと、ISO-8859 シリーズの 8 ビットで表される文字コードが含まれる。我々が業務でしばしば海外からの SAS データとして遭遇するのは、ほぼ ISO-8859-1 でなる（別項で触れるが、実際には ISO-8859-1 の亜種である Windows-1252 になる)。

ASCII も、テキストファイルとして扱う分には文字の先頭に 0x0 の 1 ビットを追加し 8 ビットごとに取り扱われる為、これを以下一文字、1 バイトとして表現する。

DBCS については、我々が普段用いている日本語のエンコーディングは Shift_JIS、ISO-2022-JP、EUC-JP となるが、SAS で現在利用出来る環境ではほぼ Shift_JIS となる。Shift_JIS は厳密には 1 バイトと 2 バイトの文字が混在する。1 バイトでは ASCII の 7 ビットの文字及び半角カナ、半角記号（文字集合としては JIS X 0201 で定義）が表される。2 バイトを用いるものとしては、全角文字（文字集合としては JIS X 0208～）となる。

Shift_JIS とその亜種について

Shift_JIS は、半角カナで使われていた箇所以外の箇所、C1 制御文字領域と呼ばれる 0x81-0x9F の箇所及び 0xE0-0xFC の領域に全角文字の 1 バイト目の識別番号をシフトさせる事により、既に Shift_JIS 制定以前にあった「ASCII と半角カナ、半角記号」の JIS X0201 に拡張する形で実装されている。

JIS X 0208 に不足しているが利用頻度の高い文字は個別のベンダーで追加されており、MS-DOS 時代では NEC と IBM では互換のない文字が多数あった。Windows で MS による統合がなされ、Windows-31J となる（別名 CP932、MS932)。その為、Shift_JIS にはない文字が多数含まれる。

1990 年代から 2000 年代にかけて、JIS X 0208 への漢字等の追加（また Windows-31J で用いられている文字の取り込みもある）より、JIS X 0213:2004 が誕生した。この最新版の Shift_JIS-2004 については、Mac や Java では対応しているものの、Windows では Windows-31J では追加の文字には対応していない(2007 年の Windows Vista 以降に Unicode で対応する形をとる)。

ISO-8859-1 とその亜種について

ISO-8859-1 は欧米でよく使われる文字コードであり、アメリカと西ヨーロッパの各国の文字が格納されている。その最新版は ISO-8859-15 であるが、普及する前に Unicode に置き換わり実際にはあまり使われていない。

こちらも Windows-31J と同様に Windows による独自の拡張があり、その拡張版は Windows-1252、通称 Wlatin-1 と呼ばれる。拡張の部分には記号等が多く用いられている。開始と終了が区別される一重引用符及

び二重引用符（‘’、“”）や短剣符（†、‡）が存在すれば Shift_JIS との区別が可能であるが、異なる領域が半角カナである場合にはその文字が格納されている上変換可能でエラー等の警告も発生しない為、文字種の検出や自動変換等には困難が生じる。

Unicode について

初期の構想では 16 ビット、2 バイト（Unicode の文脈では、バイトではなくオクテットで表現される事が正式であるが、ここでは用語としてのバイトをそのまま用いる）で、世界中の文字を全て表す仕組みであり、主要メーカーが参加したユニコードコンソーシアムによって作られた。のちに、別で進められていた ISO/IEC 10646 を完全に置き換える形で Unicode が採用され国際標準規約となる。
しかし UTF-16 は全ての文字を 2 バイトで表現するよう開発されていた為、そのままでは過去のプログラムを大幅に変更する必要があった。過去遺産として ASCII 等で動いていたアプリを Unicode に容易に対応させる為に都合のよい形式として UTF-8 は後に生まれている。

Unicode では、Windows-31J は全て基本多言語面（以下 BMP）に登録されるが、JIS X 0213:2004 に追加されたものについては一部が追加漢字面に登録され、UTF-8 では 4 バイトのサイズの文字となる。

Unicode の文字符号化形式と符号化スキーム

Unicode の文字符号化形式としては、以下が一般的である。

- UTF-8
- UTF-16
- UTF-32

UTF-16 と UTF-32 については、符号化スキームとしてビッグエンディアン方式（BE）とリトルエンディアン（LE）方式がある。これは CPU 内のレジスタでのバイト単位のデータの並べ方に由来する。基本はビッグエンディアンであるが、Windows が Intel x86 由来のリトルエンディアンを採用しており、Windows の内部では UTF-16LE となっている。

また、UTF-16 だけであるが、サロゲートペアという、BMP 以外に格納された文字を 4 バイトで表す表現がある。

UTF-16 や UTF-32 では、予めとして取り決めてある場合には、バイトオーダーマーク（以下、BOM）をつけない UTF-16 LE や UTF-16BE として定義される。UTF-16 のみで表現される符号化スキームでは通常 BOM を付与する（ビッグエンディアンであれば 0xEFFF リトルエンディアンであれば 0xFFEF）。標準はぶっぐエンディアンで格納する。UTF-32 でも同様である。

UTF-8 については、符号化形式よりリトルエンディアンやビッグエンディアンという形式の違いが発生しない。しかし、Windows によって、UTF-8 であるという事を認識出来るよう先頭に「EF BB BF」のバイトを付与される事があり、UTF-8 の BOM 付きという形式が生成される事がある。ないものを UTF-8N と表現したりする事がある。基本的には UTF-8 には BOM は不要である。

Unicode の合字、正規化、異体字セレクタ

Unicode は例えば「ガ」（全角）「ｶﾞ」（半角+濁点記号）を同一として比較したり検索したりする為に文字列を変換する事を Unicode 正規化と呼ぶ。

漢字については同じ意味の漢字を日本・中国・韓国で統合するという事が行われている。統合されたものは CJK 統合漢字、それ以外の統合漢字とは字形が異なり既存の Shift_JIS 等の文字コードから互換性を保つための文字については CJK 互換漢字となり別のコードポイントが与えられている。異体字セレクタが整えられ、比較的近年、各種サービスで Unicode 正規化が行われる事が増えた。正規化の際には CJK 統合漢字＋異体字セレクタに変換されるが、様々な漢字で複雑な問題（過剰な統合、未統合）が発生しており、未だ漢字領域については取り扱いには注意が必要である。

SAS では特に異体字セレクタ・正規化は対応していない。その為、異体字セレクタがある場合には一文字増え、同じ文字が一致しない事となる。

異体字セレクタは他にも絵文字等にも存在する。この異体字セレクタを正しく取り扱えるかどうかはツールやフォントに依存する。

周辺ツールが何の文字コードで動いているか

SAS そのものは、Windows SAS の日本語であれば Shift_JIS(Windows-31J)であるし、SAS Studio 等の製品であれば UTF-8 で動作するが、フォルダやファイルの指定等、OS の環境やロケールに依存してトラブルになる事もある。

その為、周辺ツールが何の文字コードで動いているかは認識しておく必要がある。以下、代表的なものを記載する。

- Windows の OS や Office については、内部的には UTF-16LE
- Windows のファイルシステムである NTFS も UTF-16LE、ただし FAT はロケール依存（日本であると Windows-31J）
- Linux 等については、ファイルシステムはロケールに依存する（現在標準的には UTF-8）
- DDE については、ロケールに依存した文字コードで動く
- HTML 等の Web 経由での入力や出力には現代では UTF-8 が基本的に用いられる

SAS における文字コード変換の仕組みと対応について

SAS では Cross-Environment Data Access (CEDA)という仕組みがあり、SAS データセットに作成時のエンコーディング情報が格納されている為、文字切れが発生しない場合においては、何も意識せずにそのまま別の環境で作成されたデータセットを読み込める。

ただし、Shift_JIS 環境で作成されたデータセットに関しては、基本的に漢字等が 2 バイト表現となるが、UTF-8 に変換する場合には通常 3 バイトになる。その為、CVP エンジンを用いて CVPMULTIPLIER オプションや CVPBYTES オプションを指定する事によって、文字列の長さを自動拡張するビューが提供されている。

また、これらの自動変換の仕組みを回避する為の encoding=any オプション、自力で変換する為の kcvt 関数等がある。

SAS での文字コードへの対応事例

Shift_JIS 環境から UTF-8 へのデータ移行

SAS データセットについては、原則そのままとし、SAS プログラムだけ別途ツールにて Shift_JIS から UTF-8 への変換を実施した。

変換成否の検証としては、往復変換での検証を行った。一度 UTF-8 への変換を行った後、再度 Shift_JIS に変換し、大元のファイルと比較し一致する事で確認を行った。変換に際し変換不能な文字が存在した場合には置換文字（U+FFFD、黒いダイヤの中にクエスチョンマークが標準）に置き換わる為、検出可能である。

但し、Shift_JIS と Wlatin-1 が混在した場合には、原理的な判別方法が難しく、目視確認や統計的に行う方法やルールで恣意的にいずれかの文字コードの前提とする形で対応する必要がある。

また、SAS フォーマットやテンプレート等は移行方法としてはそのままは移行できず、SAS データセットあるいはプログラムに移行前のシステムの上で移しておく方法となり、移行前環境がない場合には対応不能となる為注意が必要である。

UTF-8 環境下の DDE での出力プログラム対応

Shift_JIS で動作していた DDE を用いて Excel に出力するプログラムを、UTF-8 環境下で動作させようとした際に、文字化けが発生する。

Windows のシェルはロケールに依存する為、DDE へ出力する内容を Shift_JIS で出力しないと文字化けしてしまう。その為、出力する内容についてはデータが Shift_JIS の範疇である事を考慮した上で、kcvt 関数を用いて Shift_JIS 変換した後に DDE に出力して対応した。

当該手法は暫定対応であり、恒久的な対応を行うには、レイアウトが重要なものに関しては ODS EXCEL 等での出力に変更する等、DDE の利用を避ける必要がある。あるいは、当該箇所のみ SAS 自体を Shift_JIS で動作させる方法も検討する。

Windows のフォルダパス対応の現状

フォルダのパスに関しては、SAS でしばしばトラブルになってきたが、現在は UTF-8 での SAS であれば、フォルダに UTF-8 にしかない文字についても指定可能になっている。（SAS9.4TS1M8 で確認した）

```
1    libname TEST "Z:¥𠮟" ;
NOTE: Libref TEST was successfully assigned as follows:
      Engine:        V9
      Physical Name: Z:¥𠮟
```

S-JIS の SAS であると、当然ながら Shift_JIS には当該文字はない為エラーとなる。回避策としては、当該フォルダをネットワークドライブとして割り当てる事によって対応が可能である。

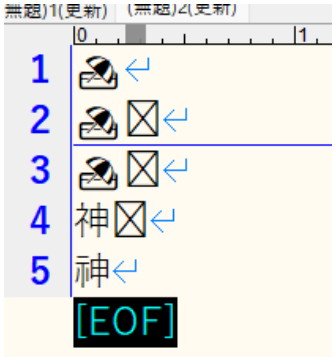
Unicode の異体字セレクタ対応

Excel は異体字セレクタに対応しているが、フォントが該当する異体字セレクタに対応していないパターンがある。絵文字では問題ないが、MS 明朝や IPA P 明朝では対応していない異体字セレクタがあり異なる字形

で出力される。

	A	B	C	D	E
1	游ゴシック	ＭＳ　明朝	Noto Sans JP	IPA P明朝	
2					パラソルの絵文字のみ
3					パラソルの絵文字＋テキスト表現強制
4					パラソルの絵文字＋イラスト表現強制
5	神	神	神	神	CJK統合漢字＋異体字セレクタ
6	神	神	神	神	CJK互換漢字

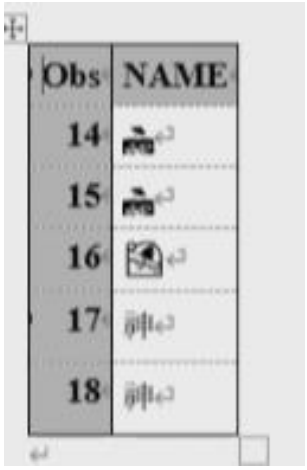
サクラエディタでは、そもそも異体字セレクタに対応していない



SAS Universal Viewer では以下の通り（フォントは MS ゴシック）



SAS9.4 で RTF 出力の場合（フォントは游明朝）



PDF でコピー&ペーストで文字が変になってしまう問題

PDF を作成する際、フォントとフォント内文字番号が保存される。その後逆引きで Unicode のマッピング表を作成する方法がとられている場合、その際に同じ字形を用いる別の漢字（康熙部首の文字等）がフォント内に存在した場合に、文字コードを間違って設定する事が起こる。

ツールによっても未対策・対策済みツール、またフォントに依存したりもする為、「コピー&ペーストが容易ではない」という認識で対策する必要がある。

結語

文字コードの概略、並びに周辺で使われる文字コード等を説明させていただいた。また、実例としていくつかの事例を紹介させていただいた。

特に Unicode は、現在のコンピューター利用ではデファクトスタンダードになっているが、全ての文字コードを統一する目的で作られているが故に仕様が複雑な事もあり、新しい問題も発生している。今後 Unicode の利用シーンが増えていくと、使用可能なフォントの選定等含め、検討する必要も出てくるかと考えている。

当該資料がその際の一助になれば幸いである。

参考文献

一部のみ抜粋

SAS® 9.4 各国語サポート(NLS): リファレンスガイド 第五版

https://documentation.sas.com/doc/ja/pgmsascdc/9.4_3.5/nlsref/titlepage.htm

とほほの文字コード入門

https://www.tohoho-web.com/ex/charset.html

Unicode 正規化

https://ja.wikipedia.org/wiki/Unicode%E6%AD%A3%E8%A6%8F%E5%8C%96

Twitter で旧字が化ける? CJK 互換漢字という罠

https://nixeneko.hatenablog.com/entry/2019/03/16/214654

PDF をコピー&ペーストするとなぜ“文字化け”が起きてしまうのか　変換テーブル“ToUnicode CMap”が原因だった | ログミーBusiness

https://logmi.jp/main/technology/324366

SASプログラムの可能性を広げるFILENAME PIPEオプション - Linuxコマンド連携による拡張性の向上とSASログでの実行管理

○三木　悠吾

（グラクソ・スミスクライン株式会社）

Filename Pipe Option Expands the Possibilities of SAS Programs
- Improve Scalability through Linux Command and Execution Management with SAS logs

Yugo Miki
GlaxoSmithKline K.K.

要旨

本稿では、SAS プログラム内から Linux コマンドをシームレスに実行する filename ステートメントの PIPE オプションに焦点を当てる。現在の SAS 製品の多くは cloud 環境へ移行しており、Windows で行えていた X コマンドを使用できない環境が多くなってきている。PIPE オプションは SAS 9.2 から filename ステートメントに実装されており、Linux SAS において、Linux コマンドを SAS から実行することが可能となると同時に、Linux の実行ログや標準エラー出力を直接 SAS ログに記録することができるという特性について解説し、具体的な活用事例として、ls や find コマンドを用いたプログラムの拡張性向上の例と、プロセス包括的なログによる実行管理を紹介する。

キーワード：filename、PIPE、Linux SAS、cloud 環境

1. はじめに

1.1. 現代の分析環境と SAS

現代のデータ分析環境は、単一のシステムで完結していた時代から、多様なツールや OS が相互に連携するエコシステムへと変貌している。この変化の中で、SAS もまた進化を続けている。従来は PC SAS および SAS Server といったオンプレミス環境での利用が主流であったが、近年はクラウド化が急速に進展している。Web ブラウザベースの統合開発環境である SAS Studio、クラウドネイティブなアーキテクチャで設計された分析プラットフォーム SAS Viya、製薬業界向けに GxP 要件を考慮したクラウドソリューションである LSAF (Life Science Analytics Framework) などがその代表例である。これらのクラウド基盤の多くは Linux で実装されている。Linux が選ばれる理由は、経済性、性能、柔軟性の 3 つの点において他の OS を圧倒しているからである。

1.2. 環境依存性のある SAS の機能

クラウド化された SAS Studio やクラウド上の Linux サーバーで稼働する SAS 環境では、従来の Windows デスクトップ版 SAS とは動作の仕組みが根本的に異なる。そのため、多くの SAS プログラマーに馴染み深い一部のコマンドや機能は使用できなくなっている。その代表的な例が、対話型セッションを管理する DM（Display Manager）と、OS コマンドを実行する X コマンドである。Windows 版の SAS では、プログラム記述用の「エディタ」、実行ログを表示する「ログ」、結果を出力する「アウトプット」といったウィンドウ群をまとめて Display Manager と呼び、これが SAS の標準的なユーザーインターフェースであった。しかし、クラウド環境の SAS はサーバー上でプロセスとして実行されており、ユーザーは Web ブラウザを通じて SAS Studio などのインターフェースから操作する形となる。SAS の本体はユーザーの手元の PC ではなく、データセンターにある Linux サーバーで稼働している。このサーバーには、Windows のようなグラフィカルなデスクトップ画面は存在しないため、Display Manager を起動することはできない。一方、X コマンドは SAS セッションから一時的に抜けて OS のコマンドを実行する機能である。クラウド環境の SAS は Linux サーバー上で動作しているため、X コマンドで実行されるのは Linux のコマンドである。しかし、Display Manager と同様に、コマンドを実行するための対話的なターミナルウィンドウはサーバー上には存在しない。そのため、X コマンドはエラーになるか、バックグラウンドで実行されたとしても、Windows 版で期待された動作にはならない。結果として、DM（Display Manager）や OS コマンドを実行する X コマンドは、以前はログ管理やプログラム開発時のツールとして利用されることが多かったが、最新のクラウド型 SAS システムでは使用できなくなっている。

2. FILENAME PIPE オプションの基本

2.1. FILENAME PIPE オプションとは

PIPE オプションは、filename statement のオプションの一つであり、SAS プログラムと OS（オペレーティングシステム）のコマンドを直接つなぐための機能である。このパイプを通じて、主に以下の 2 つのことが可能になる。

- OS コマンドの実行結果を、ファイルのように SAS で読み込む。
- SAS 上のマクロ変数などのデータを、OS コマンドへ直接入力として渡す。

したがって、Linux OS の場合、ls（ファイル一覧）や grep（文字列検索）、zip（圧縮）といった Linux の強力なコマンドを直接 SAS のワークフローに組み込み、プログラムの能力を大幅に拡張できる。本稿では、cloud 環境下で最もよく使われる OS である Linux について

2.2. 基本構文と動作

```
filename myfile PIPE "echo Hello World!";
data _null_;
    infile myfile;
    input;
    put _infile_;
```

```
run;
```

上記のコードの動作を以下に解説する。

filename myfile PIPE "echo Hello World!":
filename statement だが PIPE オプションにより myfile はディスク上の物理的なファイルではなく、OS コマンド「echo Hello World!」であることを指定している。Filename statement のみでは、OS コマンドは実行されない。

infile myfile;
データステップにおけるデータソースとして、先ほど定義した myfile を指定している。この時点で「echo Hello World!」コマンドが実行される。

put _infile_ ;
自動変数_infile_の内容を SAS ログに出力する。実際にはで指定された OS のコマンドの実行結果が出力される。なお、infile ステートメントも PIPE オプションに対応しているため、以下のようなコードでも同じ結果を得られる。

```
data _null_;
    infile "echo Hello World!" PIPE;
    input;
    put _infile_;
run;
```

2.3. 使用可能な Linux command

Filename statement の PIPE オプションで利用可能な Linux コマンドは、基本的にコマンドラインで完結し、標準入出力（キーボード入力や画面出力）を介して動作する非対話的なコマンドであれば、ほぼすべて使用できる。ただし、GUI（グラフィカルな画面出力）を必要とするコマンドや、cd や top のように対話的な操作を前提とするコマンドは不向きである。SAS のワークフローで特に役立つのは、以下のようなカテゴリのコマンドである。

カテゴリ	コマンド	主な役割と利用シーン
ファイル・ディレクトリ操作	ls	ファイル一覧の表示。特定の拡張子や更新日時のファイルリストを動的に取得する。
	find	高度なファイル検索。複雑な条件（ファイル名、更新日、サイズ等）でファイルを検索する。
テキスト処理・データ整形	grep	文字列検索。ログファイルから特定のエラーメッセージが含まれる行だけを抽出する。

カテゴリ	コマンド	主な役割と利用シーン
	sed	テキスト置換・編集。不要なヘッダー行を削除したり、特定の文字列を一括で置換したりする。
圧縮・アーカイブ	zip, unzip	.zip 形式の圧縮・解凍。外部とやり取りする標準的な圧縮ファイルを扱う。

3. PIPE を利用したプログラム

3.1. 外部ファイルフォルダの内容確認

臨床試験の解析フローにおいて、「特定のディレクトリに存在する特定のファイルを読み込んでデータを作成する」といったハンドリングが頻繁に発生する。このようなプログラムを実行する際、処理時点でそのディレクトリに「どのファイルが、どのような状態で存在したか」という証跡をログに記録しておくことは、データベースロック時のデータの検証や品質管理の観点から非常に重要である。

以下のコードを使用することで、標準的な ls コマンドをパイプから実行し、SAS ログへディレクトリ内のファイル一覧を出力できる。

```
filename myfile PIPE " ls -l /path/to/your/directory/";
data _null_;
    infile myfile;
    input;
    put _infile_;
run;
```

```
<SAS ログ>
103    filename myfile PIPE " ls -l /path/to/your/directory/";
104    data _null_;
105        infile myfile;
106        input;
107        put _infile_;
108    run;
NOTE: The infile MYFILE is:
          Pipe command="ls -l /path/to/your/directory/"
total 4
-rwxrwxrwx 1 root domino 977 May 26 08:45 123456_dev_sig.xlsx
-rwxrwxrwx 1 root domino 977 May 26 08:45 123456_diary_20250630.accdb
-rwxrwxrwx 1 root domino 977 May 26 08:45 aesi_med271_20250808.csv
-rwxrwxrwx 1 root domino 977 May 26 08:45 cmsi_whodd_20250808.csv
NOTE: 5 records were read from the infile MYFILE.
```

```
      The minimum record length was 7.
      The maximum record length was 69.
NOTE: DATA statement used (Total process time):
      real time          0.03 seconds
      cpu time           0.00 seconds
```

3.2. Zip/Unzip を SAS プログラムから実行する

データ作成および解析のワークフローにおいて、ファイルの圧縮（zip）および解凍（unzip）は頻繁に発生する作業である。SAS プログラム内からこれらの処理を直接実行できることは、手作業の排除にとどまらず、データ処理プロセス全体の自動化、堅牢性、効率性の向上という大きなメリットをもたらす。

典型的な業務フローして、外部から zip 形式で圧縮されたデータを受け取り、SAS で加工した後、成果物を再び zip 形式で圧縮して納品するフローを考える。

従来の手法（手作業）:

1. ファイルをサーバーにアップロードする。
2. ターミナルソフトなどでサーバーにログインし、手動で unzip コマンドを実行してファイルを解凍する。
3. SAS プログラムを実行して、解凍されたファイルを処理する。
4. 出力された成果物ファイルを、手動で zip コマンドを使って圧縮する。
5. 圧縮ファイルをダウンロードして送信する。

このプロセスは手間がかかる上に、解凍忘れや圧縮対象の間違いといった人為的ミスを誘発する。

SAS によるシンプル化:

1. ファイルをサーバーにアップロードする。
2. SAS プログラムが FILENAME PIPE 経由で unzip コマンドを実行し、データファイルを解凍する。
3. そのまま後続の DATA ステップや PROC ステップでファイルを処理する。
4. 処理完了後、FILENAME PIPE 経由で zip コマンドを実行し、成果物ファイルを圧縮する。
5. 圧縮ファイルをダウンロードして送信する。

サーバー内のすべての処理を 1 つのプログラムとして記述することで、実行するだけで済み、ミスが発生せず、ログに zip や unzip 処理の実施記録を残すことが可能である。

このように、ワークフローを SAS に集約してシンプル化することで、作業効率が飛躍的に向上し、プロセスの信頼性も高まる。また、SAS で扱うデータ、特に CSV やテキスト形式の生データはサイズが大きくなりがちである。これらの大容量ファイルをネットワーク経由でサーバーに転送する際に、zip で圧縮することで転送時間を短縮できることもメリットである。

3.3. ファイルの Hash 値の計算を行う

データ処理のワークフローにおいて、ファイルが破損していないか、あるいは改ざんされていないかを確認する「完全性（Integrity）検証」は非常に重要である。この検証には、ファイル全体から一意なハッシュ値（チェックサム）を計算し、期待値と比較する手法が一般的に用いられる。

SAS には、ハッシュ値を計算するための標準関数として MD5 関数や SHA256 関数が用意されている。これらの関数は非常に強力であるが、設計上、SAS の変数に対してハッシュ値を計算するものであり、ファイル全体を直接対象とすることは想定されていない。ファイル全体のハッシュ値を SAS 標準関数で計算しようとすると、以下のような問題が発生する。

- ファイル全体を 1 つの SAS 変数に読み込む必要がある。
- SAS の文字変数の長さには最大 32,767 バイトという制限がある。

この制約により、SAS の標準関数だけでファイル自体のハッシュ値を計算し、その完全性を検証することは極めて困難である。

この問題は、PIPE オプションを用いて、ファイルハッシュ値を計算するために最適化された Linux コマンドを呼び出すことで、簡単かつ確実に解決できる。Linux には標準で以下の強力なコマンドが用意されている。

md5sum: MD5 ハッシュ値を計算するコマンド
sha256sum: より強力な SHA-256 ハッシュ値を計算するコマンド

これらのコマンドはファイルサイズに制限なく、高速にハッシュ値を計算できる。以下のコードを PIPE オプションから実行することで、/your/data/sdtm/フォルダに存在するすべての SAS xport ファイルのハッシュ値を取得し、「半角スペース」を「,」へ変換し、txt ファイルへ書き込む。この txt ファイルの拡張子を csv とすることで、Excel などのツールで計算されたハッシュ値を確認することが可能である。

```
filename myfile PIPE " md5sum /your/data/sdtm/*.xpt | sed 's/ /,/' >/your/data/hash
/hashlist.csv";
data _null_;
    infile myfile;
run;
```

3.4. vmember で認識しないファイルの metadata を作成する

VCOLUMN や VTABLE といった SAS の関数やメタデータビューは、ライブラリに割り当てられた SAS データセットやカタログなどの情報を取得するためのものである。しかし、今注目を集めている JSON ファイルは SAS が直接「テーブル」として認識する形式ではないため、JSON ファイルが多数置かれたディレクトリを SAS ライブラリとして割り当てても、その中身をこれらの関数で一覧表示することはできない。

これに対応するには以下の 2 つの方法が挙げられる。

SAS 標準関数（DOPEN, DREAD）を使う方法

この方法は、OS に依存しない「SAS ネイティブ」な方法である。ディレクトリを開き、その中のメンバー（ファイルやサブディレクトリ）の名前を 1 つずつ読み込んで処理する。長所としては、SAS の関数だけで完結しているため、OS（Windows, Linux）を問わず動作する高い移植性がある。一方で短所としては、以下に述べる PIPE 方式に比べてコードの記述量がやや多くなる点である。

PIPE から ls コマンドを使う方法

この方法は、OS のネイティブなコマンドの実行結果を SAS で直接読み込むものである。Linux 環境で作業している場合に特に強力で、簡潔に記述できる。以下に具体例を示す。

```
filename myfile PIPE “ls -1 /your/datasets /*.json”;
data datalist;
   infile myfile;
   length fullname $100. memname $20.;
   input fullname $; *full filepath;
   memname = scan(fullname, -2, “/.”);
run;
```

DATALIST　Total rows: 22　Columns: 2 of 2　Rows 1 to 22

Enter expression

	fullname	memname
1	/domino/datasets/local/DOMINO/sdtm/ae.json	ae
2	/domino/datasets/local/DOMINO/sdtm/cm.json	cm
3	/domino/datasets/local/DOMINO/sdtm/dm.json	dm
4	/domino/datasets/local/DOMINO/sdtm/ds.json	ds
5	/domino/datasets/local/DOMINO/sdtm/ex.json	ex

Linux の ls コマンドにワイルドカード（*）を付けて「.json」の拡張子を持つファイルを一覧表示させ、その実行結果を input ステートメントから読み込んでいる。読み込んだ値にはファイル名以外にもパスと拡張子が含まれているため、scan 関数でファイル名のみ取り出すことでデータリストを作成できる。

4. 考察：導入効果とリスク

PIPE オプションは、SAS の能力を OS レベルで拡張する強力な機能である。その利用には当然ながらメリットとデメリットが存在する。

PIPE オプションを利用する最大のメリットは、機能の拡張性である。OS の専門ツールを活用することで、データ処理や Web アクセス、バージョン管理などの高度な機能を追加することが可能である。記述の簡潔性も大きな利点であり、OS コマンドを使うことで SAS 標準関数よりも簡単で直感的なコード記述が可

能になる。また、OS コマンドはその OS 上で最適化されており、大規模データ処理などで高速に動作することもメリットである。

PIPE オプションの最大のデメリットは、SAS プログラムが OS に依存してしまう点である。「一度書けば、どこでも動く」という SAS の大きな利点の一つが損なわれるため、様々な環境で利用される汎用的なプログラムの開発には不向きである。また、セキュリティリスクも存在する。OS コマンドを直接実行できるため、悪意のあるコードが実行されるリスクを考慮する必要がある。例えば、外部から取得したプログラムに「rm -rf /」のような危険なコマンドが書き込まれている場合、実行されてしまう可能性がある。さらに、エラーのハンドリングが難しい点も課題である。SAS 標準関数のエラーは SAS のエラーコードで管理されるが、PIPE で実行した OS コマンドのエラーは OS の流儀でハンドリングする必要がある。プログラマーには SAS だけでなく、OS におけるエラー処理の知識も求められる。

	メリット（利点）	デメリット（注意点）
PIPE 利用	▪ OS による機能拡張性 ▪ 可読性の高いプログラムによる生産性向上 ▪ パフォーマンスの改善	▪ OS への依存的なコードによる移植性の低下 ▪ セキュリティリスク ▪ OS 側のエラー処理の必要性
SAS Native	▪ 高い移植性 ▪ 安全性 ▪ 一貫したエラー処理	▪ 機能の限界 ▪ 冗長なコード

PIPE オプションのセキュリティリスクは、SAS OnDemand や SAS Viya などクラウドの仕組みによって大きく軽減される。SAS Viya のような最新のクラウドソリューションは、多くの場合コンテナ技術を基盤としている。ユーザーが実行する SAS のプログラムは、隔離され、標準化されたコンテナ環境内で動作するため、影響範囲は限定的である。また、クラウドソリューションでは、誰が何をできるかという権限が専門の管理者によって一元的に管理される。そのため、ユーザーが本来アクセスすべきでないディレクトリに対して OS コマンドを実行できてしまうリスクはほぼないと考えられる。

以上をまとめると、プログラムコードに対する要求が OS を超えて移植性や安全性が最優先される場合は、SAS の標準関数を基本とすべきである。それ以外の場合、特にクラウドソリューション型の SAS 利用においては PIPE オプションによる OS 機能の利用が有効になる場合が多い。

5. おわりに

本稿では PIPE オプションという強力な機能が標準的に利用可能になったことについて述べたが、SAS の製品がクラウドへとその主戦場を移した今、SAS プログラマーの役割とスキルセットも変化を始めていることになる。SAS が OS と連携できるということは、GitHub のような Web API とも連携が可能であることを意味する。現在、AI をはじめ多くの Web サービス、API が乱立しており、それぞれユニークなサービスを提供している。SAS だけでなくどのサービスをどのような環境からどうやって使うかを判断できる幅広いリテラシー能力を持てば、SAS プログラマーはれからもデータ活用の最前線で価値を創出し続けることができ

るだろう。変化を恐れず最新の技術を吸収し続けることで、SAS プログラマーは今後ともデータサイエンスの最前線で活躍することができるだろう。

6. 参考文献

1. **SAS Institute Inc. (2025). *SAS 9.4 Companion for UNIX Environments, Sixth Edition*. Cary, NC: SAS Institute Inc.**

ベイズ流臨床試験デザインの頻度論的較正：

事前情報と中間解析の影響

○三島遼

（京都府立医科大学附属病院）

Bayesian Clinical Trial Designs with Frequentist Calibration: Impact of Prior Information and Interim Analyses

Ryo Mishima

University Hospital, Kyoto Prefectural University of Medicine

要旨

ベイズ流臨床試験デザインにおいて第一種の過誤確率および検出力を評価する SAS プログラムを紹介する。さらに、事前情報の利用および中間解析が与える影響を調べたシミュレーション実験を報告する。

キーワード：ベイズ流臨床試験デザイン、SAS マクロ、第一種の過誤確率、検出力、中間解析

1. はじめに

近年、医薬品・医療機器の開発において、ベイズ流臨床試験デザインが注目を集めている。このデザインは、事後確率（posterior probability）あるいはそれと予測確率（predictive probability）の２つを組み合わせて設計・実施される（Saville et al., 2014; Lee, 2024）。事後確率は、観測データを用いて事前分布を更新して得られる事後分布から算出される、特定の仮説が真である確率である。一方、予測確率は、途中時点までのデータに基づいて予測された、試験完了時に成功基準を満たす確率であり、試験途中での意思決定（早期中止または継続）に有用である。

従来の頻度流と比較したベイズ流の主な利点は以下の 3 つである。第一に、関連する事前情報（先行研究、専門家の意見、類似薬の効果など）を試験データの解析に明示的に組み込める点である（Berry, 2006）。第二に、試験の進行中に蓄積されるデータを用いて事後分布を逐次的に更新し、中間解析や段階的な意思決定を統一的な枠組みで行える点である（Saville et al., 2014）。第三に、「エビデンスの程度」を確率として直接表現できるため、意思決定者にとって理解しやすい結果を提供できる点である（Kruschke & Liddell, 2018）。

一方で、ベイズ流臨床試験デザインを承認申請に用いる場合、「頻度論的に較正（frequency-calibrated）」されているという観点が重要視されている。これは、伝統的な頻度論的アプローチと整合性のあるベイズ流デザインを意味する。ベイズ統計は試験成功を支持する事前情報を解析に含めることができ、また、時点を決めず柔軟に事後分布を更新することが可能だが、その一方で頻度論と同様に第一種の過誤確率を適切に制御することが求められる（Berry, 2006; Lee, 2024）。米国 FDA が 2010 年に公表した医療機器のベイズ流臨床試

験に関するガイダンスでは、計画している試験デザインの動作特性を様々なシナリオの下で事前評価することが明確に推奨されている（FDA, 2010）。特に中間解析やサンプルサイズ再設計を含む適応的試験デザインでは、解析的な方法でサンプルサイズを厳密に導出することが困難であるため、多くの場合シミュレーションによる評価が不可欠である（Saville et al., 2014; Teramukai et al., 2023）。

本稿では、事後確率および予測確率に基づく中間解析を考慮したベイズ流臨床試験デザインのシミュレーションを行う SAS マクロプログラムを提案する。このプログラムは、単群および二群比較の状況下で、二項モデルと正規モデルを扱う計 4 つのモジュールから構成される。さらに、様々な試験設定（サンプルサイズ、事前分布の選択、中間解析の実施タイミング、意思決定の閾値など）における動作特性の評価結果も報告する。本研究の特色は、特に二群比較において事後確率と予測確率の両方に基づく中間解析を組み込んだシミュレーションを可能にした点にある。これにより、実用的なベイズ流臨床試験デザインの実装と評価に貢献することが期待される。

2. 事後確率および予測確率に基づくベイズ流臨床試験デザイン

2.1 解析事前分布

解析事前分布（analysis prior distribution）とは、パラメータ θ に関する事前情報を確率分布として表現したものである（Teramukai, 2023）。この分布は、試験データから得られる尤度関数と組み合わせることで、ベイズの定理に従って事後分布を導出するために用いられる。解析事前分布は、無情報事前分布と情報のある事前分布の 2 種類に大別される（Lee, 2024）。無情報事前分布は、想定されるパラメータ空間 Θ を広くカバーし、事前情報の影響を最小限に抑えるように設計される。この場合、事後分布は主としてデータに基づいて形成され、客観的な推論が可能となる。一方、情報のある事前分布は、先行研究、専門家の意見、類似薬の効果など、関連する既存の知見を定量的に組み込むために用いられる。臨床的に妥当な事前情報を適切に取り入れることで、パラメータの推定精度が向上し、従来の頻度論的アプローチと比較して必要サンプルサイズを削減することが可能となる。この特性は、希少疾患や小児を対象とした臨床試験など、試験参加者のリクルートが困難な状況において特に重要な利点となる（AMED, 2021; Rognli, 2023）。

2.2 デザイン事前分布

デザイン事前分布（design prior distribution）は、臨床試験のパラメータに関する事前の期待を確率分布として定量的に表現したものである（Teramukai, 2023）。これは試験計画段階においてアウトカムを予測するために使用される。デザイン事前分布は先行研究の結果や専門家の知見に基づいて設定され、やや楽観的な見通しを反映するように構成される。

2.3 事後分布の導出

ベイズ流アプローチでは、試験計画時に設定した事前分布と、試験から得られたデータの尤度関数を組み合わせることで、ベイズの定理により事後分布を導出する。ベイズ流手法の特徴は、データ収集の進行に合わせて事後分布を逐次的に更新できる点にある。この特性により、中間解析に基づく早期中止や適応的なデザイン変更など、臨床試験における逐次的な意思決定プロセスに特に適している（Saville et al., 2014）。

尤度関数と事前分布が共役関係にある場合、例えば二項分布に対するベータ分布や正規分布に対する正規

-逆ガンマ分布などでは、事後分布を解析的に導出することが可能である（Teramukai, 2023）。しかし、モデルが複雑な場合は解析的な導出は困難であり、マルコフ連鎖モンテカルロ法（MCMC）などのサンプリング手法を用いて事後分布を近似的に得ることが一般的である（Saville et al., 2014; Kruschke, 2018）。

○二項モデルにおける共役事前分布

二値アウトカムを扱う場合、成功確率θに対するベータ分布 $Beta\,(\alpha,\beta)$ を事前分布とし、二項分布 $Bin\,(n,\theta)$ を尤度関数とすると、観測データ（ n 回試行中 y 回成功）に基づく事後分布は $Beta\,(\alpha+y,\beta+n-y)$ となる。

○正規モデルにおける共役事前分布

正規分布に従う連続値アウトカムを扱う場合、平均パラメータ μ に対する正規分布と、尤度関数として正規分布を組み合わせることが一般的である。具体的には、μ に対する正規分布 $N(\mu_a,\sigma^2/n_a)$ を事前分布とし、正規分布 $N(\mu,\sigma^2/n)$ を尤度関数とする場合、観測データ（標本平均 $\bar{y}$）に基づく μ の事後分布は $N((n\bar{y}+n_a\mu_a)/(n+n_a),\ \sigma^2/(n+n_a))$ となる。

2.4 事後確率および予測確率

以下では、パラメータ θ に関する仮説 $\theta > \theta_0$ を評価する試験を考える。これは頻度論的アプローチにおける片側検定で、閾値 θ_0 を定め、帰無仮説 $\theta \le \theta_0$ と対立仮説 $\theta > \theta_0$ を設定することに相当する。なお、$\theta < \theta_0$ を評価する場合も同様の議論が成り立つ。

2.4.1 事後確率

事後分布に基づく研究仮説 $\theta > \theta_0$の確率を事後確率という。事後確率が予め設定した閾値 λ 以上であれば、試験目的を達成したとみなす判断基準が広く採用されている（FDA, 2010; AMED, 2021）。

2.4.2 予測確率

中間解析は、臨床試験の完了前に予め定められた時点でデータを評価し、早期意思決定を可能にする手法である（Giovagnoli, 2021）。被験治療が明らかに有効、無効、あるいは有害である場合に試験を早期終了できる。これは群逐次デザインと呼ばれ、最大サンプルサイズを固定した上で、特定条件下での早期終了の可能性を組み込む。予測確率は、中間解析時点までに得られたデータに基づいて算出される、試験終了時に事後確率が閾値λを超える確率として定義される(Saville et al., 2014)。中間解析時点より後のデータを予測している点において、実際に収集されたデータのみに依拠している事後確率とは異なる。臨床試験の中間解析における意思決定では、予測確率が閾値 γ_L を下回れば無効中止、γ_U を上回れば有効中止という判断基準を設定できる。

2.4.3 ベイズ流検出力

ベイズ流臨床試験の事前評価では、デザイン事前分布 p^D を用いることで、事後確率と予測確率に基づいて $\theta > \theta_0$ と判断する確率を算出できる。中間解析を伴わない場合、データ $\boldsymbol{y}_N$ を用いて算出される事後確率を $T(\boldsymbol{y}_N)$ と表すと、

$$\beta_{p^D} = \mathbb{P}[T(\boldsymbol{y}_N) > \lambda]$$

と記述でき、この確率はベイズ流検出力と呼ばれる（Psioda, 2019; Teramukai, 2023）。

本稿では、中間解析を組み込み、事後確率に基づく有効性の判定および予測確率に基づく無益性の判定を行うことを考える。したがって、各段階で試験目標の達成を判断する確率をすべて合算したものを検出力と呼ぶ（Giovagnoli, 2021; Lee, 2024）。例えば、中間解析を 1 回実施する場合、

$$\beta_{p^D} = \beta_{p^D}^1 + \beta_{p^D}^2$$

と分解できる。ここで、

・ $\beta_{p^D}^1$：中間解析時に事後確率の基準を達成する確率

$$\beta_{p^D}^1 = \mathbb{P}[T(\boldsymbol{y}_{N_1}) > \lambda]$$

・ $\beta_{p^D}^2$：最終解析時に事後確率の基準を達成する確率

$$\beta_{p^D}^2 = \mathbb{P}[\, T(\boldsymbol{y}_{N_1}) <= \lambda \ \text{ and } \ T(\boldsymbol{y}_{N_1}, \boldsymbol{y}_{N_2}) > \lambda \,]$$

であり、p^Dはデザイン事前分布、$\boldsymbol{y}_{N_1}$は中間解析までに得られるデータ、$\boldsymbol{y}_{N_2}$は中間解析から最終解析の間に得られるデータ、$T(\boldsymbol{y}_*)$ は $\boldsymbol{y}_*$ を用いて算出する事後確率、λ は事後確率の閾値を表す。本稿では閾値 λ を中間解析および最終解析で共通としているが、中間解析ではより厳しくするといった運用も可能である。無益性については、$\boldsymbol{y}_{N_1}$ が得られた元で $\boldsymbol{y}_{N_2}$ を予測して算出する予測確率 $\tilde{T}(\boldsymbol{y}_{N_2} \mid \boldsymbol{y}_{N_1})$ と無益性の閾値 γ_L に対して、

$$\tilde{T}(\boldsymbol{y}_{N_2} \mid \boldsymbol{y}_{N_1}) < \gamma_L$$

が成り立つときに中止を判断する。

頻度流の仮説検定と同様に、パラメータ θ として帰無仮説の閾値 θ_0 を設定すると誤った結論を導く可能性を評価することができる。デザイン事前分布 p^D ではなく帰無仮説の閾値 θ_0 を与えた場合に上述の計算によって得られる指標 β_{θ_0}を、ここでは第一種の過誤確率と定義する（Lee, 2024）。

サンプルサイズ設計において、このベイズ流検出力がある水準（例えば 80%）以上となるサンプルサイズを決定する方法は「検出力規準」と呼ばれる (Teramukai, 2023)。

2.5 ベイズ流臨床試験デザインの"頻度論的較正"

ベイズ流臨床試験では、サンプルサイズを事前に固定する必要はなく、得られたデータに基づいて事後分布を逐次的に更新できる。しかし、この柔軟性は誤った判断のリスクを高める可能性があり、規制当局への承認申請では、従来の頻度論的アプローチと同様に第一種の過誤確率の適切な制御が求められる（Berry, 2006; Giovagnoli, 2021）。この要件を満たすように事後確率の閾値 λ や予測確率の閾値 γ_U, γ_L を調整しなければならないが、通常、様々なシナリオ下でのシミュレーションを必要とし、特に中間解析を含む複雑なデザインでは計算負荷が大きい（Saville, 2014; Lee, 2024）。

3. SAS を用いたシミュレーション

本研究では、ベイズ流臨床試験デザインの評価に利用可能な SAS マクロ関数を開発した。読者は https://github.com/rio-mishmash/BayesSSD から SAS ファイルを入手することができる。先に挙げた、共役な事前分布を用いる状況を想定しており、群の数（単群または二群）とアウトカムの種類（二値または正規分布に従う連続値）に応じて 4 つのモジュールで構成されている。SAS の POWER procedure を参考にして引数を命名した。なお、共役事前分布を用いない場合については、SAS の MCMC procedure を活用したサンプリング法による分布の近似を活用できる。この方法は過去の SAS ユーザー総会で紹介されている（張, 2021 他）。

本節では、開発した SAS マクロ関数の提案およびそれらを用いて実施した第一種の過誤確率と検出力を評価したシミュレーションの結果を報告する。一般的なチュートリアルでは単群二値アウトカムの例が多く取り上げられるが（Matano et al., 2016; Teramukai, 2023 など）、本研究では二群比較や連続値アウトカムの場合についても検討した。これらの評価を通じて、ベイズ流試験デザインの頻度論的特性の概観を確認するとと

もに、提案する SAS プログラムあるいはその改変が試験計画のためのシミュレーションに利用可能であることを示す。なお、シミュレーション回数は 1000 回、分布からのサンプリング回数は 1000 回であり、その他のパラメータはシミュレーション結果をまとめた表に記載する。

3.1 1 標本 2 項モデル

事後確率および予測確率に基づく手続きという点で 4 つのプログラムは同様の構造をしている。本稿では、単群二値アウトカムの場合を例に、SAS プログラムの構成を説明する。

○事後確率を計算する関数 f_posterior_prob

この関数は事後分布に基づいて閾値 nullproportion に対する確率を計算する。データ収集前（ntotal=0）では事前分布のパラメータをそのまま使用し、データ収集後（ntotal>0）では事前分布と尤度から得られる事後分布のパラメータを計算する。データ収集前における事後確率を算出しているのは、選択した解析事前分布が研究仮説を支持する程度を確認できるようにするためである。

```
* posterior probability;
if ntotal = 0 then do;
	a = prior_a;
	b = prior_b;
end;
else do;
	a = prior_a + y_sum;
	b = prior_b + ntotal - y_sum;
end;
```

次に、ベータ分布の累積分布関数（cdf）を用いて上側または下側の事後確率を計算する。共役な事前分布を利用しない場合、この計算をモンテカルロ法などで代替することになる。

```
if sides = "U" then do;
	posterior_prob  = 1 - cdf("beta", nullproportion + margin, a, b);
end;
if sides = "L" then do;
	posterior_prob  =   cdf("beta", nullproportion + margin, a, b);
end;
```

○予測確率を計算する関数 f_predictive_prob

この関数は現在までのデータから得られた事後分布を用いて、試験完了時の成功確率を予測する。具体的には、中間解析時点（ncurrent）から最終サンプルサイズ（ntotal, npergroup）までに得られる未観測データについて、(1) 中間解析時点の事後分布から確率値をサンプリングし、(2)サンプリングした確率値に基づいて残りの試験期間で得られるであろう予測アウトカムをシミュレーションし、(3)観測データと予測データを合わせた最終データで事後確率を計算して試験成功の判定を行う、という一連の過程を繰り返し実施し、その平均値として予測確率を算出する。

```
* predictive response;
if (ntotal-ncurrent) > 0 then do;
      * sampling from the posterior distribution;
      p_sample = J(nrow, &nmc., .);
      call randgen(p_sample, "beta", a, b);
      * predictive response;
      y_sum_pred = rand("binomial", p_sample, ntotal-ncurrent);
      posterior_prob = f_posterior_prob(ntotal, prior_a, prior_b,
                                        y_sum_obs + y_sum_pred,
                                        nullproportion, lambda, sides, margin);
end;
else do;
      posterior_prob = f_posterior_prob(ntotal, prior_a, prior_b,
                                        y_sum_obs,
                                        nullproportion, lambda, sides, margin);
end;
predictive_prob = (posterior_prob >= lambda)[,:]; *rowMeans;
```

○試験目標達成の判断

事後確率および予測確率は設定された解析時点(interim)ごとに計算される。解析時点の初期設定は interim={0 1}で、これはデータ収集前と全データ収集後の 2 時点で計算する設定である。例えば interim={0 0.7 1}と設定すれば、最大サンプルサイズの 70%が集まった時点で中間解析を行う試験デザインを評価できる。各解析時点で、事後確率が閾値 λ を超えれば試験目標達成(promising)と判断する。一方、予測確率が γ_L を下回った場合は無効中止(futility)とする。

```
do j = 1 to ncol(&interim.);
      ...
     if 1 < j < ncol(&interim.) then do;
          * stop early with promising result;
          promising[,j] = (predictive_prob[,j] > &gamma_U. );
          * stop early with futility    result;
          futility[,j]    = (predictive_prob[,j] < &gamma_L.);
     end;
     if j = ncol(&interim.) then do;
          * termination with promising result;
          promising[,j] = (posterior_prob[,j]    > &lambda. );
     end;
end;
```

○シミュレーション結果

無情報事前分布として一様分布（$Beta\,(1,1)$）を採用した場合、閾値$\lambda = 0.90$における第一種の過誤確率はおよそ 10%となった。この傾向はサンプルサイズに依存していない。一方、検出力はサンプルサイズの増加に伴って上昇した。検出力 80%を達成するためには 100 例程度が必要であるとわかる。

最大サンプルサイズの 70%時点で中間解析を実施し、予測確率の閾値$\gamma_L = 0.10$を用いた場合を検討した。第一種の過誤確率を 10%に制御するには、λを 0.90 よりも大きく設定する必要がある。例えば$\lambda = 0.92$とした場合、第一種の過誤確率と検出力はいずれも中間解析を行わない場合と同程度に維持される。一方で、中間解析によって早期終了が可能となり、期待サンプルサイズ（ESS）は減少する。

表 1 第一種の過誤確率および検出力（1 標本 2 項モデル（$\theta_0 = 0.3$））

	Null Hypothesis					Alternative Hypothesis (Design Prior)				
$\boldsymbol{\theta}$	θ_0					$\theta_0 + 0.1$				
Analysis Prior	Non-Informative $Beta(1,1)$			Optimistic $Beta(3,2)$	Pessimistic $Beta(1,4)$	Non-Informative $Beta(1,1)$			Optimistic $Beta(3,2)$	Pessimistic $Beta(1,4)$
Interim Analysis*	Final only	Interim + Final		Final only		Final only	Interim + Final		Final only	
$\boldsymbol{\lambda}$	0.90	0.92		0.90		0.90	0.92		0.90	
$\boldsymbol{\gamma_L}$	-	0.10		-		-	0.10		-	
N (total)	**type I error**	**type I error**	**ESS**	**type I error**	**type I error**	**power**	**power**	**ESS**	**power**	**power**
20	0.130	0.148	15.6	0.230	0.054	0.388	0.429	16.9	0.562	0.244
40	0.125	0.124	30.0	0.214	0.075	0.561	0.568	32.3	0.671	0.448
60	0.100	0.127	44.2	0.166	0.062	0.686	0.662	47.4	0.768	0.569
80	0.086	0.115	57.8	0.128	0.053	0.689	0.717	62.3	0.776	0.597
100	0.128	0.079	71.1	0.177	0.084	0.827	0.747	76.9	0.882	0.767
150	0.096	0.102	106.0	0.122	0.068	0.914	0.875	110.6	0.932	0.886
200	0.109	0.070	141.0	0.140	0.088	0.961	0.950	146.0	0.971	0.946
300	0.082	0.082	211.0	0.122	0.082	0.988	0.978	213.4	0.995	0.988
500	0.098	0.105	351.0	0.126	0.086	0.999	0.997	351.6	0.999	0.999

λ: Promising threshold of posterior probability, γ_L: Futility threshold of predictive probability, ESS: Expected Sample Size

* Interim analysis at 0.7 × total sample size

事前分布の選択も試験デザインの動作特性に影響を及ぼす。治療効果に対して楽観的な事前分布（$Beta\,(3,2)$, 期待値 0.6）を設定した場合、第一種の過誤確率と検出力がともに上昇する傾向が認められた。反対に、悲観的な事前分布（$Beta\,(1,4)$, 期待値 0.20）を採用した場合は、これらの指標が低下することが確認された。しかしながら、サンプルサイズが増加するにつれてこの影響は徐々に減少し、最終的には事前分布の選択に関わらず、無情報事前分布を採用した場合と同様の性能に近づいた。

3.2 1 標本正規モデル

表 2 に、正規分布に従う連続値アウトカムを対象とした単群試験デザインのシミュレーション結果を示す。このシミュレーションでは、目標値 $\theta_0 = 0$ に対して試験群の平均値が+1 となる状況を想定している。事前分布として、無情報事前分布 $N(0{,}100^2)$、楽観的事前分布 $N(+2, 2^2)$、悲観的事前分布 $N(-1, 2^2)$ をそれぞれ設定した。結果に関して、中間解析の影響および事前分布の選択による影響は、先に述べた単群二値アウトカムの場合と同様の傾向であった。

表 2　第一種の過誤確率および検出力（１標本正規モデル（$\theta_0 = 0$））

	Null Hypothesis					Alternative Hypothesis (Design Prior)				
$\boldsymbol{\theta}$	θ_0					$\theta_0 + 1$				
Analysis Prior	Non-Informative $N(0, 100^2)$			Optimistic $N(+2, 2^2)$	Pessimistic $N(-1, 2^2)$	Non-Informative $N(0, 100^2)$			Optimistic $N(+2, 2^2)$	Pessimistic $N(-1, 2^2)$
Interim Analysis*	Final only	Interim + Final		Final only		Final only	Interim + Final		Final only	
$\boldsymbol{\lambda}$	0.90	0.92		0.90		0.90	0.92		0.90	
$\boldsymbol{\gamma_L}$	-	0.10		-		-	0.10		-	
N (total)	**type I error**	**type I error**	**ESS**	**type I error**	**type I error**	**power**	**power**	**ESS**	**power**	**power**
20	0.094	0.100	15.7	0.172	0.038	0.352	0.343	16.6	0.503	0.192
40	0.097	0.105	30.8	0.149	0.055	0.497	0.480	33.3	0.617	0.386
60	0.111	0.093	46.4	0.159	0.079	0.621	0.592	50.1	0.705	0.530
80	0.102	0.112	60.4	0.145	0.068	0.699	0.679	64.7	0.757	0.629
100	0.090	0.110	76.8	0.142	0.062	0.756	0.724	81.3	0.813	0.698
150	0.107	0.117	114.1	0.143	0.085	0.874	0.895	115.7	0.908	0.847
200	0.112	0.104	156.8	0.141	0.091	0.936	0.926	151.2	0.959	0.922
300	0.107	0.094	227.2	0.130	0.094	0.981	0.982	217.9	0.986	0.974
500	0.096	0.086	375.7	0.120	0.088	1.000	1.000	352.3	1.000	1.000

λ: Promising threshold of posterior probability, γ_L: Futility threshold of predictive probability, ESS: Expected Sample Size

* Interim analysis at 0.7 × total sample size

3.3　２標本２項モデル

二値アウトカムの二群比較におけるシミュレーション結果を表 3 に示す。二群を扱う本節および次節において、縦軸の N は一群あたりのサンプルサイズを示していることに留意されたい。

最大サンプルサイズの 50%と 70%時点で計 2 回の中間解析を実施し、予測確率の閾値 $\gamma_L = 0.10$ を用いた場合を検討した。このとき、第一種の過誤確率のインフレを抑えるため、事後確率の閾値を$\lambda = 0.94$とし

表 3　第一種の過誤確率および検出力（２標本２項モデル（$\theta_0 = 0.3$ ））

	Null Hypothesis					Alternative Hypothesis (Design Prior)				
$\boldsymbol{\theta}$	θ_0					$\theta_0 + 0.1$				
Analysis Prior	Non-Informative $Beta(1, 1)$			Optimistic $Beta(3, 2)$	Pessimistic $Beta(1, 4)$	Non-Informative $Beta(1, 1)$			Optimistic $Beta(3, 2)$	Pessimistic $Beta(1, 4)$
Interim Analyses*	Final only	Interim + Final		Final only		Final only	Interim + Final		Final only	
$\boldsymbol{\lambda}$	0.90	0.94		0.90		0.90	0.94		0.90	
$\boldsymbol{\gamma_L}$	-	0.10		-		-	0.10		-	
N (arm)	**type I error**	**type I error**	**ESS**	**type I error**	**type I error**	**power**	**power**	**ESS**	**power**	**power**
20	0.110	0.114	13.8	0.159	0.054	0.278	0.279	15.4	0.346	0.160
40	0.087	0.096	25.9	0.118	0.062	0.357	0.344	31.0	0.451	0.263
60	0.105	0.100	38.4	0.138	0.077	0.491	0.431	45.3	0.538	0.388
80	0.096	0.103	51.1	0.138	0.077	0.532	0.472	62.2	0.577	0.447
100	0.098	0.099	65.0	0.121	0.075	0.586	0.602	75.9	0.631	0.535
150	0.102	0.091	94.6	0.122	0.082	0.709	0.665	111.5	0.748	0.667
200	0.090	0.105	128.0	0.114	0.076	0.777	0.771	142.3	0.807	0.750
300	0.087	0.105	193.3	0.105	0.073	0.894	0.896	193.2	0.905	0.876
500	0.115	0.080	322.0	0.128	0.088	0.969	0.971	284.5	0.987	0.950

λ: Promising threshold of posterior probability, γ_L: Futility threshold of predictive probability, ESS: Expected Sample Size

* Interim analyses at 0.5 × total sample size and 0.7 × total sample size

た。今回の設定では、同じ検出力を達成するためには、中間解析を行う方が多くのサンプルサイズを必要とする傾向が見られた。一方で、期待サンプルサイズは中間解析を行った方が小さくなる場合があった。

3.4 ２標本正規モデル

表 4 に正規分布に従う連続値アウトカムを扱った二群比較試験デザインのシミュレーション結果を示す。このシミュレーションでは、デザイン事前分布として退化分布ではなく正規分布 $N(+1, 0.1^2)$ を設定した。計画段階では治療効果の想定に不確実性が伴うため、その不確実性を反映したサンプルサイズ設計が重要である。開発した SAS マクロ関数では、パラメータとして幅のある分布を設定することが可能である。

表 4　第一種の過誤確率および検出力（２標本正規モデル（$\theta_0 = 0$ ））

	Null Hypothesis					Alternative Hypothesis (Design Prior)				
$\boldsymbol{\theta}$	θ_0					$N(\theta_0 + 1, 0.1^2)$				
Analysis Prior	Non-Informative $N(0, 100^2)$			Optimistic $N(+2, 2^2)$	Pessimistic $N(-1, 2^2)$	Non-Informative $N(0, 100^2)$			Optimistic $N(+2, 2^2)$	Pessimistic $N(-1, 2^2)$
Interim Analyses*	Final only	Interim + Final		Final only		Final only	Interim + Final		Final only	
$\boldsymbol{\lambda}$	0.90	0.94		0.90		0.90	0.94		0.90	
$\boldsymbol{\gamma_L}$	-	0.10		-		-	0.10		-	
N (arm)	**type I error**	**type I error**	**ESS**	**type I error**	**type I error**	**power**	**power**	**ESS**	**power**	**power**
20	0.102	0.090	13.5	0.156	0.068	0.287	0.214	14.9	0.353	0.183
40	0.089	0.080	24.5	0.130	0.069	0.299	0.319	27.5	0.361	0.231
60	0.103	0.079	36.6	0.148	0.078	0.423	0.347	43.2	0.478	0.351
80	0.092	0.078	45.6	0.123	0.070	0.577	0.455	56.5	0.624	0.497
100	0.113	0.107	65.0	0.140	0.098	0.583	0.481	75.4	0.620	0.529
150	0.089	0.105	89.2	0.108	0.081	0.582	0.677	106.0	0.631	0.539
200	0.125	0.087	116.5	0.140	0.110	0.785	0.748	134.0	0.810	0.748
300	0.103	0.086	180.5	0.115	0.095	0.871	0.855	200.2	0.894	0.859
500	0.099	0.079	295.6	0.109	0.093	0.895	0.889	307.5	0.906	0.885

λ: Promising threshold of posterior probability, γ_L: Futility threshold of predictive probability, ESS: Expected Sample Size

* Interim analyses at 0.5 × total sample size and 0.7 × total sample size

4. まとめ

本稿では、単群・二群比較における二値・連続値アウトカムに対応したベイズ流臨床試験デザイン評価用 SAS マクロを提案した。このマクロは事後確率と予測確率に基づくシミュレーションを可能とし、複数の中間解析時点や閾値パラメータの調整による試験性能への影響を詳細に検討できる柔軟性を備えている。シミュレーションにより、事前分布の選択や中間解析導入が第一種の過誤確率や検出力に与える影響を定量的に評価でき、最適なデザイン構成の特定が可能となる。

謝辞

本論文の草稿に対して貴重なご助言をいただきました京都府立医科大学生物統計学教室の手良向聡教授に心より感謝申し上げます。

引用文献

AMED 医薬品等規制調和・評価研究事業. (2021, July 16). 希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針. 日本医療研究開発機構.

Berry, D. A. (2006). Bayesian clinical trials. Nature Reviews Drug Discovery, 5(1), 27-36.

Cao, H., Yao, C., & Yuan, Y. (2023). Bayesian approach for design and analysis of medical device trials in the era of modern clinical studies. Medical Review, 3(5), 408-424.

Food and Drug Administration. (2010, February 5). Guidance for the use of Bayesian statistics in medical device clinical trials. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Division of Biostatistics, Office of Surveillance and Biometrics, Center for Biologics Evaluation and Research.

Giovagnoli, A. (2021). The Bayesian design of adaptive clinical trials. International journal of environmental research and public health, 18(2), 530.

Heuts, S., Kawczynski, M. J., Velders, B. J. J., Brophy, J. M., Hickey, G. L., & Kowalewski, M. (2025). Statistical primer: An introduction into the principles of Bayesian statistical analyses in clinical trials. European Journal of Cardio-Thoracic Surgery, 67(4), ezaf139.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review, 25(1), 178-206.

Lee, S. Y. (2024). Using Bayesian statistics in confirmatory clinical trials in the regulatory setting: A tutorial review. BMC Medical Research Methodology, 24(1), 110.

Matano, F., & Sambucini, V. (2016). Accounting for uncertainty in the historical response rate of the standard treatment in single-arm two-stage designs based on Bayesian power functions. Pharmaceutical Statistics, 15(6), 517-530.

Psioda, M. A., & Ibrahim, J. G. (2019). Bayesian clinical trial design using historical data that inform the treatment effect. Biostatistics, 20(3), 400-415

Rognli, E. W., Zahl-Olsen, R., Rekdal, S. S., Hoffart, A., & Bertelsen, T. B. (2023). Editorial perspective: Bayesian statistical methods are useful for researchers in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*, *64*(2), 339-342.

Saville, B. R., Connor, J. T., Ayers, G. D., & Alvarez, J. (2014). The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. Clinical Trials, 11(4), 485-493.

Teramukai, S. (2023). Sample size determination in Bayesian clinical trials: Inferential performance-based approach using two priors. Japanese Journal of Biometrics, 44(1), 35-51.

張方紅. (2021). SAS による単群製造販売後調査におけるベイズ流解析. SAS ユーザー総会論文集, 2.

線形モデルのプロシジャ選択を考える

○大山暁史

（イーピーエス株式会社）

Considering SAS Procedure Selection for Linear Models

Akifumi Oyama
EPS Corporation

要旨

SAS では多数の線形モデル関連のプロシジャが実装されているが、どのように使い分けるべきだろうか。本稿では回帰分析、t 検定、分散分析およびロジスティック回帰に焦点を当て、各解析に用いることができるプロシジャとそれらを使い分けるための判断材料となる情報を提示する。

キーワード：Procedure Selection, Linear Model

1. はじめに

臨床試験の解析プログラムの作成時には、参考試験のプログラムを流用することが多いと思われる。しかし、参考試験で用いられていたという理由でプロシジャを用いることは問題ないだろうか。線形モデルのプロシジャの推定法の違いに言及した文献は多少あるものの、SAS 9.4 での線形モデルのプロシジャ選択に関する文献は少ないため、本稿では回帰分析、t 検定、分散分析およびロジスティック回帰について、各プロシジャの概要とその使い分けについての 1 つの考えを提示する。

* 本稿では ods graphics を on にしている環境を想定する。
* SAS を動かしながら本稿を読めるよう、簡単なダミーデータ作成プログラムを別途アップロードするので、適宜参照されたい。

2. 回帰分析

REG プロシジャ、HPREG プロシジャ、GLM プロシジャおよび GLMSELECT プロシジャを用いて回帰分析を実施することができる。

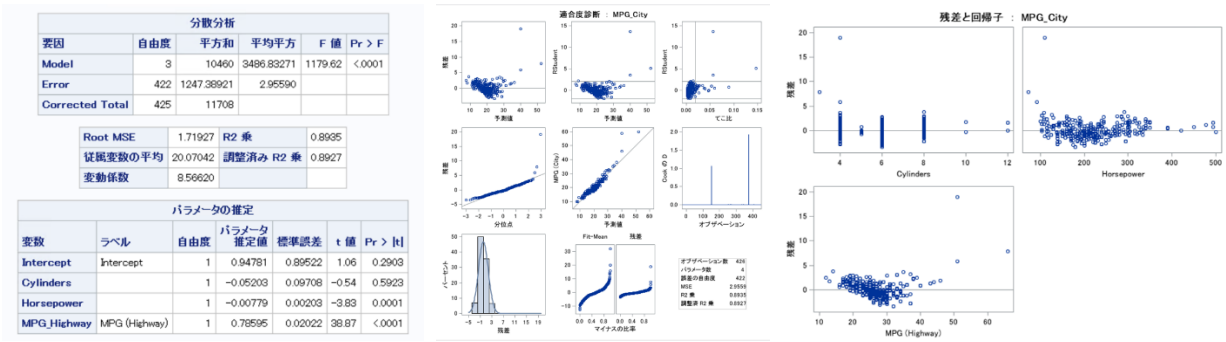
2.1 コードの書き方と出力

線形モデル関連のプロシジャは基本的に model 文にモデル式（応答変数 = 説明変数 1 説明変数 2 …）を指定する必要がある。各プロシジャの具体的な記載例と出力を以下に示す。

REG プロシジャ

パラメータ推定値に加え、回帰診断パネルや残差の散布図がデフォルトで出力されるため、正規性などを確認できる。

```
proc reg data=sashelp.cars;
    model MPG_City = Cylinders Horsepower MPG_Highway;
run; quit;
```



分散分析

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	3	10460	3486.83271	1179.62	<.0001
Error	422	1247.38921	2.95590		
Corrected Total	425	11708			

Root MSE	1.71927	R2 乗	0.8935
従属変数の平均	20.07042	調整済み R2 乗	0.8927
変動係数	8.56620		

パラメータの推定

変数	ラベル	自由度	パラメータ推定値	標準誤差	t 値	Pr > |t|
Intercept	Intercept	1	0.94781	0.89522	1.06	0.2903
Cylinders		1	−0.05203	0.09708	−0.54	0.5923
Horsepower		1	−0.00779	0.00203	−3.83	0.0001
MPG_Highway	MPG (Highway)	1	0.78595	0.02022	38.87	<.0001

HPREG プロシジャ

分散並列処理をすることができる High Performance プロシジャである HPREG プロシジャは、AIC などの評価指標も出力する。

```
proc hpreg data=sashelp.cars;
    model MPG_City = Cylinders Horsepower MPG_Highway;
run; quit;
```

分散分析

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	3	10460	3486.83271	1179.62	<.0001
Error	422	1247.38921	2.95590		
Corrected Total	425	11708			

Root MSE	1.71927
R2 乗	0.89346
調整済み R2 乗	0.89270
AIC	893.68105
AICC	893.82391
SBC	481.89881
ASE	2.92814

パラメータの推定

パラメータ	自由度	推定値	標準誤差	t 値	Pr > |t|
Intercept	1	0.947810	0.895221	1.06	0.2903
Cylinders	1	−0.052033	0.097080	−0.54	0.5923
Horsepower	1	−0.007789	0.002032	−3.83	0.0001
MPG_Highway	1	0.785950	0.020222	38.87	<.0001

GLM プロシジャ

パラメータ推定値に加え、タイプI平方和やタイプIII平方和なども出力する。REG プロシジャとは異なり、回帰診断パネルや残差の散布図がデフォルトでは出力されないが、それぞれ PLOTS=DIAGNOSTICS や PLOTS=RESIDUALS を指定することで、出力することができる[1]。

```
proc glm data=sashelp.cars;
    model MPG_City = Cylinders Horsepower MPG_Highway;
run; quit;
```

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	3	10460.49812	3486.83271	1179.62	<.0001
Error	422	1247.38921	2.95590		
Corrected Total	425	11707.88732			

R2 乗	変動係数	Root MSE	MPG_City の平均
0.893457	8.566201	1.719273	20.07042

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
Cylinders	1	5484.039660	5484.039660	1855.29	<.0001
Horsepower	1	511.164449	511.164449	172.93	<.0001
MPG_Highway	1	4465.294008	4465.294008	1510.64	<.0001

要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
Cylinders	1	0.849149	0.849149	0.29	0.5923
Horsepower	1	43.442841	43.442841	14.70	0.0001
MPG_Highway	1	4465.294008	4465.294008	1510.64	<.0001

パラメータ	推定値	標準誤差	t 値	Pr > |t|
Intercept	0.9478098796	0.89522078	1.06	0.2903
Cylinders	-.0520326988	0.09707987	-0.54	0.5923
Horsepower	-.0077890235	0.00203174	-3.83	0.0001
MPG_Highway	0.7859503189	0.02022157	38.87	<.0001

GLMSELECT プロシジャ

デフォルトで stepwise 法の変数選択を実施するため、selection=none を指定する必要があるが、以下のコードで回帰分析を実施できる。HPREG プロシジャとほぼ同じ内容を出力する。

```
proc glmselect data=sashelp.cars;
    model MPG_City = Cylinders Horsepower MPG_Highway /selection=none;
run;
```

2.2 複数のモデル指定

REG プロシジャは model 文を複数指定することができるため、複数のモデルを同時に解析することができる。

2.3 交互作用の考慮

GLM プロシジャ、HPREG プロシジャおよび GLMSELECT プロシジャは、model 文の右辺に Cylinders * Horsepower のように交互作用項を指定することができる。REG プロシジャの model 文では交互作用項を指定することができないため、REG プロシジャで交互作用を考慮したい場合には、2 変数の積を格納した変数を前処理で作成しておく必要がある。

2.4 カテゴリー変数の指定

GLM プロシジャ、HPREG プロシジャおよび GLMSELECT プロシジャは、内部的にダミーデータを作成する class 文で指定することで説明変数にカテゴリー変数を含むことができる。一方で REG プロシジャは class 文が実装されていないため、説明変数にカテゴリー変数の情報を含めたい場合には、GLMMOD プロシジャか

データステップ 0 または 1 の数値型に変換したダミー変数を作成しなければならない[2] [3]。（GLMMOD プロシジャの使用方法については参考文献[2]を参照されたい。）

2.5 多重共線性の指標確認

REG プロシジャや HPREG プロシジャは tol オプションや vif オプションを指定すると、トレランスや VIF（分散拡大係数）を出力することができる。なお、トレランスが 0.1 以下、VIF が 10 以上だと、多重共線性（説明変数間の相関関係）があると考えられる[4]。

2.6 変数選択・正則化

REG プロシジャ

以下のように model 文にオプションを記載すると変数選択を実施可能である[5]。

```
model MPG_City = Cylinders Horsepower MPG_Highway / selection= STEPWISE slentry=0.5 slstay=0.5;
```

selection オプションには FORWARD, BACKWARD, STEPWISE, MAXR, MINR, RSQUARE,ADJRSQ, CP のような変数選択法を指定できる。また、以下のように記載すると Ridge 回帰を実施できる[6]。

```
proc reg data=sashelp.cars outest=xxx ridge=0.001 to 0.02 by 0.001;
```

GLMSELECT プロシジャ

REG プロシジャと書き方が異なるが、model 文にオプションを記載すると変数選択を実施可能である[5]。

```
model MPG_City = Horsepower_cat Cylinders MPG_Highway Cylinders*MPG_Highway / selection=
        STEPWISE(select=SL slentry=0.5 slstay=0.5);
```

selection オプションには FORWARD, BACKWARD, STEPWISE, FORWARDSWAP, LAR, LASSO を指定できる[6]。（例えば LASSO を指定すると LASSO 回帰を実施できる。）

HPREG プロシジャ

selection 文を指定することで変数選択を実施できる。method オプションには、FORWARD, BACKWARD, STEPWISE, LAR, LASSO, ELASTICNET, GROUPLASSO の指定が可能である[6]。

```
selection method= STEPWISE (select=sl sle=0.5 sls=0.5);
```

2.7 小括

回帰分析について、各プロシジャの差異を示した。シチュエーションと適しているプロシジャの例示を以下に示す。

表 1 回帰分析のシチュエーションと適しているプロシジャの例

シチュエーション	適しているプロシジャ
回帰診断のグラフを確認したい場合	REG
説明変数にカテゴリー変数や交互作用項を指定したい場合	HPREG, GLM, GLMSELECT
カテゴリー変数や交互作用項を含めて変数選択したい場合	HPREG, GLMSELECT
Ridge 回帰を実施したい場合	REG
Lasso 回帰を実施したい場合	GLMSELECT

多重共線性の確認をしたい場合	REG, HPREG
タイプI, タイプIII平方和を確認したい場合	GLM
AIC などの統計的評価指標も出力したい場合	HPREG, GLMSELECT
複数のモデルを 1 回でまとめて実行したい場合	REG

3. 対応のない 2 標本 t 検定

TTEST プロシジャ、GLM プロシジャおよび MIXED プロシジャを用いて、対応のない 2 標本 t 検定を実施することができる。

3.1 コードの書き方

TTEST プロシジャ

Student の t 検定(Pooled)と Welch の t 検定(Satterthwaite)の両方の結果を出力する。また、等分散性の検定結果・データの分布・Q-Q プロットを出力する[7]。

```
proc ttest data= testdata;
    class ARM;
    var AVAL;
run;
```

GLM プロシジャ

以下のコードでは Student の t 検定を実施することができる。t 検定の結果に加え、データ分布や群間差なども出力することができる[8]。

```
proc glm data= testdata;
    class ARM;
    model AVAL=ARM;
    lsmeans ARM /tdiff = control("対照群")cl;
run; quit;
```

以下のコードではデータ分布や Welch の t 検定の結果などを出力する[8]。

```
proc glm data= testdata;
    class ARM ;
    model AVAL=ARM;
    means ARM /welch;
run; quit;
```

MIXED プロシジャ

以下のコードでは Student の t 検定の結果に加え、共分散パラメータの推定、適合度統計量、固定効果の Type III検定等を出力する[8]。

```
proc mixed data= testdata;
    class ARM ;
    model AVAL=ARM ;
    lsmeans ARM / pdiff = control("対照群")
cl ;
run;
```

以下のコードでは Welch の t 検定の結果、反復履歴、共分散パラメータの推定、適合度統計量、帰無モデルの尤度比検定、固定効果の Type III検定を出力する。群ごとに推定した共分散パラメータを用いて t 検定の結果を得ることができる[9]。

```
proc mixed data=testdata;
  class ARM ;
  model AVAL=ARM /ddfm=satterth;
  repeated /group=ARM ;/* repeated 文は誤差項の分散共分散行列を指定する。group=オプションでは、指定した変数の水準ごとの共分散パラメータを推定できる*/
run;
```

3.2 小括

TTEST プロシジャは書き方がシンプルであり、1 パターンのコードで Students 検定と Welch 検定の両方を実施できるため、基本的には TTEST プロシジャを使うことが良いと考える。なお、SAS Version6 以前の TTEST プロシジャによる Welch 検定は p 値を求める時に線形近似を用いていた[9]ため、当時の結果を再現したい場合には留意が必要である。

4. 対応のある 2 標本 t 検定

MEANS プロシジャ、UNIVARIATE プロシジャ、TTEST プロシジャおよび MIXED プロシジャを用いて、対応のある 2 標本 t 検定を実施することができる。

4.1 コードの書き方

MEANS プロシジャ

2 群の差を格納した変数に対する 1 標本 t 検定として、t 値を算出することができる。

```
proc means data=testdata_paired noprint;
    var CHG; /* 2 群の差を格納した変数*/
    output out=outdata_means
    T=T PROBT=PROBT;
run;
```

UNIVARIATE プロシジャ

MEANS プロシジャと同様の記載方法で算出可能[7]。

```
proc univariate data=testdata_paired noprint;
    var CHG;
    output out=outdata_univariate
    T=T PROBT=PROBT;
run;
```

TTEST プロシジャ

paired オプションに第 1 引数*第 2 引数を指定すると、第 1 引数-第 2 引数の差を検定する。

差の分布、全症例の推移図、一致度のグラフおよび差の Q-Q プロットを出力するので視覚的に理解しやすい。

```
proc ttest data=testdata_paired;
    paired POST*PRE; /*この場合は POST-PRE を検定する*/
run;
```

また、paired オプションを用いずに、2 群の差を格納した変数に対する 1 標本 t 検定として t 値を算出することもできる。

```
proc ttest data=testdata_paired;
    var CHG;
run;
```

MIXED プロシジャ

対応のあるデータを縦積みにし、被験者を変量効果に指定することで、対応のある t 検定を実施することができる[10]。

```
data v_stacked;
    set testdata_paired;
    AVAL=PRE;
    PAIRNUM=2; /*2 つの対応する値を区別する変数。POST-PRE の差に興味があるため POST に 1,PRE に 2 を設定*/
    output;
    AVAL=POST;
    PAIRNUM =1;
    output;
run;
proc mixed method=type3;
    class SUBJID PAIRNUM;
    model AVAL= PAIRNUM /ddfm=kenwardroger;
    random SUBJID; /*  個人差を考慮するため、変量効果に SUBJID を設定  */
    lsmeans PAIRNUM /tdiff cl;    run;
```

4.2 小括

MEANS プロシジャ、UNIVARIATE プロシジャおよび TTEST プロシジャはいずれもコードがシンプルである。データを視覚的に理解したい場合は TTEST プロシジャの paired オプションを用いた書き方をすると良いだろう。いずれのプロシジャを使う場合も検定の方向（差の向き）が適切であるかについては留意する必要がある。

5. 二元配置分散分析

分散分析用のプロシジャとして ANOVA プロシジャが実装されている。しかし、アンバランスド（要因の各水準に割付けられたデータの数が等しくない）データに対して ANOVA プロシジャでは適切に解析することができない。例えば、以下のようなコードを実行すると、アンバランスドデータのため、GLM プロシジャで解析する方が妥当な結果を出力できる可能性がある旨の Warning（WARNING: PROC ANOVA has determined that the number of observations in each cell is not equal. PROC GLM may be more appropriate.）が出力される。

```
proc anova data=exp;
        class A B;
        model Y=A B A*B;
run; quit;
```

GLM プロシジャや MIXED プロシジャは ANOVA プロシジャと同様の書き方で実行可能である[6] [11]。いずれもデータがバランスドか否かに関わらずに解析できるため、二元配置分散分析は GLM プロシジャや MIXED プロシジャを用いることが良いと考えられる。

6. ロジスティック回帰

応答変数が 2 値のロジスティック回帰について、以下の 7 プロシジャを検討した[6] [12]。

表 2 ロジスティック回帰を実施できるプロシジャ例とその概要

プロシジャ名	概要
LOGISTIC	ロジスティック回帰向けのプロシジャ
HPLOGISTIC	ロジスティック回帰向けの HP プロシジャ
CATMOD	分割表データを扱うために設計された、プロシジャ
GENMOD	一般化線形モデル向けのプロシジャ
GLIMMIX	一般化線形混合モデル向けのプロシジャ
NLMIXED	非線形混合モデル向けのプロシジャ
PROBIT	プロビット回帰分析向けのプロシジャ

6.1 コードの書き方と出力

各プロシジャのコード例と出力を以下に示す。今回紹介するコードの書き方では、いずれも最尤法で推定される。なお、いずれのプロシジャを使用する場合も応答変数の基準が意図通りであることを確認する必要があるため、基準を指定できるプロシジャについては、応答変数に event オプションを指定している。

LOGISTIC プロシジャ

各パラメータの最尤推定値に加え、応答変数の基準、適合度統計量、推定オッズ比、予測と観測データの関連性なども出力する[12]。

```
proc logistic data = heart;
    model status_N (event='0') = Smoking Cholesterol; /*  応答変数の値が 0 の時に対するモデルを推定する  */
run;
```



反応プロファイル

順番	status_N	度数の合計
1	0	3128
2	1	1921

モデルの確率基準は status_N=0 です。

モデル収束状態

収束基準(GCONV=1E-8)は満たされました。

モデルの適合度統計量

基準	切片のみ	切片と共変量
AIC	6710.045	6545.063
SC	6716.572	6564.644
-2 Log L	6708.045	6539.063

包括的帰無仮説: BETA=0 の検定

検定	カイ 2 乗値	自由度	Pr > ChiSq
尤度比	168.9618	2	<.0001
スコア	168.2704	2	<.0001
Wald	160.8889	2	<.0001

最尤推定値の分析

パラメータ	自由度	推定値	標準誤差	Wald カイ 2 乗値	Pr > ChiSq
Intercept	1	2.3238	0.1593	212.6801	<.0001
Smoking	1	-0.0166	0.00242	47.0975	<.0001
Cholesterol	1	-0.00732	0.000668	120.0829	<.0001

オッズ比の推定

効果	点推定	95% Wald 信頼限界	
Smoking	0.984	0.979	0.988
Cholesterol	0.993	0.991	0.994

予測確率と観測データの応答との関連性

一致の割合	60.6	Somers の D	0.215
不一致の割合	39.1	ガンマ	0.216
タイの割合	0.3	Tau-a	0.102
組	6008888	c	0.608

HPLOGISTIC プロシジャ

各パラメータの最尤推定値に加え、反復履歴や適合度統計量なども出力する[6]。

```
proc hplogistic data = heart;
    model status_N (event='0') = Smoking Cholesterol ; /*  応答変数の値が 0 の時に対するモデルを推定する*/
run;
```



反応プロファイル

順番	status_N	度数の合計
1	0	3128
2	1	1921

確率 status_N='0' を基準にモデル化しています。

反復履歴

反復	評価	目的関数	変化	最大勾配
0	4	0.6518286849	.	7.233984
1	2	0.6475783845	0.00425030	0.364024
2	2	0.6475602341	0.00001815	0.001072
3	2	0.6475602339	0.00000000	1.344E-8

収束基準 (GCONV=1E-8) は満たされました。

次元の数

X の列	3
効果数	3
最大効果の列数	1
交差積行列のランク	3
最適化のパラメータ数	3

適合度統計量

-2 対数尤度	6539.06
AIC (小さいほどよい)	6545.06
AICC (小さいほどよい)	6545.07
BIC (小さいほどよい)	6564.64

包括的帰無仮説: BETA=0 の検定

検定	カイ 2 乗値	自由度	Pr > ChiSq
尤度比	168.9618	2	<.0001

パラメータの推定

パラメータ	推定値	標準誤差	自由度	t 値	Pr > |t|
Intercept	2.3240	0.1593	Infty	14.58	<.0001
Smoking	-0.01662	0.002421	Infty	-6.86	<.0001
Cholesterol	-0.00733	0.000668	Infty	-10.96	<.0001

CATMOD プロシジャ

以下のコードで実行可能である。各パラメータの最尤推定値に加え、先頭 40 件までの母集団プロファイルも出力する。連続変数は direct 文で指定しておく必要がある[6] [12]。

```
proc catmod data = heart;
    direct Smoking Cholesterol;
    model status_N = Smoking Cholesterol; /*応答変数に event オプションや ref オプションは実装されていない*/
run; quit; /*  対話型のプロシジャのため、quit が必要  */
```



母集団プロファイル

標本	Smoking	Cholesterol	標本サイズ
1	0	117	1
2	0	121	1
3	0	124	1
4	0	125	1
5	0	128	2
6	0	129	1
7	0	134	3
8	0	138	3
9	0	140	2
10	0	142	4

反応プロファイル

応答	status_N
1	0
2	1

最尤法分析

最大尤度の計算は収束しました。

最尤分散分析

要因	自由度	カイ 2 乗	Pr > ChiSq
Intercept	1	212.72	<.0001
Smoking	1	47.10	<.0001
Cholesterol	1	120.11	<.0001
Likelihood Ratio	1E3	1540.82	<.0001

最尤推定の分析

パラメータ	推定値	標準誤差	カイ 2 乗	Pr > ChiSq
Intercept	2.3240	0.1593	212.72	<.0001
Smoking	−0.0166	0.00242	47.10	<.0001
Cholesterol	−0.00733	0.000668	120.11	<.0001

GENMOD プロシジャ

各パラメータの最尤推定値に加え、適合度評価の基準も出力する[12]。

```
proc genmod data = heart;
    model status_N (event='0') = Smoking Cholesterol / dist = binomial link = logit;
run;
```



反応プロファイル

順番	status_N	度数の合計
1	0	1921
2	1	3128

PROC GENMODによるモデルの確率基準はstatus_N='0'です。モデルの確率基準をstatus_N='1'に変更するには、PROCステートメントにDESCENDINGオプションを指定してください。



適合度評価の基準

基準	自由度	値	値/自由度
対数尤度		−3269.5316	
完全対数尤度		−3269.5316	
AIC (小さいほどよい)		6545.0632	
AICC (小さいほどよい)		6545.0680	
BIC (小さいほどよい)		6564.6441	

アルゴリズムは収束しました。

最大尤度パラメータ推定値の分析

パラメータ	自由度	推定値	標準誤差	Wald 95% 信頼限界		Wald カイ 2 乗	Pr > ChiSq
Intercept	1	−2.3240	0.1593	−2.6363	−2.0117	212.72	<.0001
Smoking	1	0.0166	0.0024	0.0119	0.0214	47.10	<.0001
Cholesterol	1	0.0073	0.0007	0.0060	0.0086	120.11	<.0001
尺度	0	1.0000	0.0000	1.0000	1.0000		

GLIMMIX プロシジャ

固定効果のみ指定することでロジスティック回帰を実行可能。Solution オプションを指定することでパラメータの推定を出力する[6]。なお、dist オプションに binomial を指定していると、応答変数の基準が設定できない仕様である。

```
proc glimmix data = heart;
    model status_N = Smoking Cholesterol / dist = binomial link = logit solution;
run;
```



次元の数

X の列	3
Z の列	0
サブジェクト(V のブロック)	1
サブジェクト毎の最大 OBS	5049

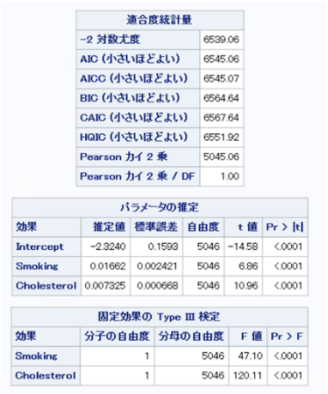
最適化の情報

最適化の手法	Newton-Raphson
最適化のパラメータ数	3
下限があるパラメータ数	0
上限があるパラメータ数	0
固定効果	Not Profiled

反復履歴

反復	再開	評価	目的関数	変化	最大勾配
0	0	4	3274.0725508	.	85.91816
1	0	3	3269.5346116	4.53793916	1.799245
2	0	3	3269.5316207	0.00299088	0.001035
3	0	3	3269.5316207	0.00000000	4.47E-10

収束基準(GCONV=1E-8)は満たされました。



適合度統計量

−2 対数尤度	6539.06
AIC (小さいほどよい)	6545.06
AICC (小さいほどよい)	6545.07
BIC (小さいほどよい)	6564.64
CAIC (小さいほどよい)	6567.64
HQIC (小さいほどよい)	6551.92
Pearson カイ 2 乗	5045.06
Pearson カイ 2 乗 / DF	1.00

パラメータの推定

効果	推定値	標準誤差	自由度	t 値	Pr > |t|
Intercept	−2.3240	0.1593	5046	−14.58	<.0001
Smoking	0.01662	0.002421	5046	6.86	<.0001
Cholesterol	0.007325	0.000668	5046	10.96	<.0001

固定効果の Type III 検定

効果	分子の自由度	分母の自由度	F 値	Pr > F
Smoking	1	5046	47.10	<.0001
Cholesterol	1	5046	120.11	<.0001

NLMIXED プロシジャ

以下のように二項分布とロジスティック関数を組み合わせることでロジスティック回帰を実行できる[12]。
応答変数の基準は設定できない。

```
proc nlmixed data = heart;
    parms beta0 = 0 beta1 = 0 beta2 = 0; /*偏回帰係数の初期値の指定*/
    theta = 1 - logistic(beta0 + beta1 * Smoking + beta2 * Cholesterol);
    model status_N ~ binomial(1, theta); /*status_N ~ binary(theta)でも良い*/
run;
```

また、対数尤度関数を記載する方法もある[12]。

```
proc nlmixed data = heart;
    parms beta0 = 0 beta1 = 0 beta2 = 0;
    theta = 1 - logistic(beta0 + beta1 * Smoking + beta2 * Cholesterol);
    L = status_N * log(theta) + (1 - status_N) * log(1 - theta);
    model status_N ~ general(L);
run;
```

初期パラメータ

beta0	beta1	beta2	負の対数尤度
0	0	0	3499.70011

反復履歴

反復	呼び出し	負の対数尤度	差	最大勾配	傾き
1	9	3400.7359	98.96426	32250.4	-1.445E8
2	13	3396.4663	4.269534	30509.5	-35569.5
3	19	3286.7516	109.7147	31532.8	-123.979
4	21	3270.6320	16.11968	6424.99	-26.7335
5	23	3269.5371	1.094858	256.656	-2.07842
6	26	3269.5316	0.005468	14.3699	-0.01057
7	29	3269.5316	5.128E-6	0.070625	-9.74E-6

NOTE: GCONV convergence criterion satisfied.

適合度統計量

-2 対数尤度	6539.1
AIC (小さいほどよい)	6545.1
AICC (小さいほどよい)	6545.1
BIC (小さいほどよい)	6564.6

パラメータの推定

パラメータ	推定値	標準誤差	自由度	t 値	Pr > |t|	95% 信頼限界		勾配
beta0	2.3240	0.1593	5049	14.59	<.0001	2.0117	2.6363	-0.00047
beta1	-0.01662	0.002421	5049	-6.86	<.0001	-0.02137	-0.01187	-0.00578
beta2	-0.00733	0.000668	5049	-10.96	<.0001	-0.00864	-0.00602	-0.07062

PROBIT プロシジャ

以下のように累積分布関数にロジスティック分布を指定すると実行可能である[6]。

```
proc probit data=heart;
    model status_N(event='0') = Smoking Cholesterol / dist =logistic;
    /* (event= '興味ある水準')  または(ref='参照水準')  で応答変数の基準を指定可能  */
run;
```

反応プロファイル

順番	status_N	度数の合計
1	0	3128
2	1	1921

PROC PROBITは反応プロファイル表の従属変数status_Nの水準の低い順位に対する確率を基準に処理しています。

アルゴリズムは収束しました。

効果に対する Type III 分析

効果	自由度	Wald カイ 2 乗値	Pr > ChiSq
Smoking	1	47.1049	<.0001
Cholesterol	1	120.1083	<.0001

最大尤度パラメータ推定値の分析

パラメータ	自由度	推定値	標準誤差	95% 信頼限界		カイ 2 乗	Pr > ChiSq
Intercept	1	2.3240	0.1593	2.0117	2.6363	212.72	<.0001
Smoking	1	-0.0166	0.0024	-0.0214	-0.0119	47.10	<.0001
Cholesterol	1	-0.0073	0.0007	-0.0086	-0.0060	120.11	<.0001

6.2 カテゴリー変数

応答変数および説明変数へのカテゴリー変数の指定可否を以下にまとめた[6]。

表 3 各プロシジャのカテゴリー変数の指定可否

プロシジャ名	応答変数	説明変数
LOGISTIC	○	○
HPLOGISTIC	○	○
CATMOD	○	○
GENMOD	○	○
GLIMMIX	×	○
NLMIXED	×	△
PROBIT	○	○

説明変数にカテゴリー変数を指定する場合

LOGISTIC プロシジャ、HPLOGISTIC プロシジャ、GENMOD プロシジャ、GLIMMIX プロシジャおよび PROBIT プロシジャは class 文への指定が必要になる。一方で CATMOD プロシジャはカテゴリー変数を想定したプロシジャであるため、model 文にそのまま指定するだけで良い。（カテゴリー変数を想定しているため class 文が実装されていない。逆に連続変数を説明変数に指定したい場合には、前述の通り direct 文に指定する必要がある）NLMIXED プロシジャでは REG プロシジャと同様に、class 文が実装されていないため、前処理でダミー変数を生成することで、説明変数にカテゴリー変数を含むことができる。

前処理なしで説明変数を指定できるプロシジャは、プロシジャ内でダミー変数を生成している。以下に各プロシジャのダミー変数生成法を示す[6] [13]。

表 4 各プロシジャのダミー変数生成法

ダミー変数生成法	LOGISTIC	HPLOGISTIC	CATMOD	GENMOD	GLIMMIX	PROBIT
EFFECT	☆		☆	○		
GLM	○	☆		☆	☆	☆
ORDINAL THERMOMETER	○			○		
POLYNOMIAL	○			○		
REFERENCE	○	○	○	○		
ORTHEFFECT	○			○		
ORTHORDINAL ORTHOTHERM	○			○		
ORTHPOLY	○			○		
ORTHREF	○			○		

☆：デフォルト ○：指定可能

このように、プロシジャごとにダミー変数生成方法が異なることに留意が必要である。（例えば、LOGISTIC プロシジャと GENMOD プロシジャはデフォルトのダミー変数生成法が異なるため、手当なしに実行した場合には推定結果が一致しない。[13]）

6.3 交互作用

本章で紹介した 7 つのプロシジャは全て説明変数に交互作用項を指定することができる。（いずれのプロシジャも説明変数を指定する箇所に、変数 1*変数 2 のように交互作用項を記載すれば良い。）

6.4 複数モデルの指定

CATMOD プロシジャでは複数の model 文を記載できる。ただし、それぞれの model 文を run 文で区切る必要がある。SAS 8.2 より前のバージョンに限り、PROBIT プロシジャでも複数の model 文を指定できる。

6.5 完全分離・準完全分離

完全分離とは、説明変数の値をある境界値で分割することで応答変数が綺麗にグループ分け出来ている状態である。パラメータが有限の値であると完全分離のデータにフィットせず、うまく推定できないことが知られている。（境界値においてのみ 2 値応答変数の両方の値が存在する場合は準完全分離という。）
以下に完全分離や準完全分離の場合の各プロシジャの挙動を示す[6]。

LOGISTIC プロシジャ

完全分離も準完全分離も実行されるが、Warning や Output にその旨が出力される。（準完全分離は検出されないこともある。）準完全分離データへの対応として、以下 2 点を実施することができる。

Firth 法：model 文に firth オプションを指定することで、罰則付き最尤推定値を算出

Exact 法：exact 文を指定することで Exact な最尤推定値を算出

HPLOGISTIC プロシジャ

完全分離も準完全分離も実行されるが、Note や Output にその旨が出力される。

CATMOD プロシジャ

完全分離も準完全分離も実行されるが、出力結果に注釈が付く。

GENMOD プロシジャ

完全分離の場合も準完全分離の場合も実行が通り、推定値も算出される。（完全分離の場合に対数尤度が 0 と出力されること以外は、（準）完全分離であることを結果からは気づくことは難しい。）準完全分離データに対しては exact 文を指定することで、Exact な最尤推定値を算出することができる。

GLIMMIX プロシジャ

完全分離の場合も準完全分離の場合も実行が通り、推定値も算出される。（完全分離の場合に-2 対数尤度が 0 と出力されること以外は、（準）完全分離であることを結果からは気づくことは難しい。）

NLMIXED プロシジャ

完全分離・準完全分離ともに収束しない旨の Warning は出るが、推定値も算出される。（完全分離の場合には-2 対数尤度が 0 と出力される。）いずれの場合もパラメータ推定値の共分散行列や相関行列が出力される。

PROBIT プロシジャ

完全分離のデータ実行時には Warning が出力される。完全分離と準完全分離ともに推定値と合わせてトレランス分布のプロビットモデルやトレランスパラメータの共分散行列の推定値の出力もされる。

6.6 変数選択

LOGISTIC プロシジャは model 文の selection オプションに変数選択法を指定すると、変数選択を実施できる。HPLOGISTIC プロシジャは selection 文の method オプションに変数選択法を指定すると、変数選択を実施できる。それぞれのプロシジャで指定できる変数選択法を以下にまとめる。

表 5 LOGISTIC プロシジャと HPLOGISTIC プロシジャで指摘できる変数選択法

変数選択法	LOGISTIC	HPLOGISTIC
FORWARD	○	○
BACKWARD	○	○
BACKWARD(FAST)	×	○
STEPWISE	○	○
SCORE	○	×

6.7 小括

2 値ロジスティック回帰について、7 つのプロシジャを検討した。シンプルなコードであり、本稿で紹介したような実務上使用する頻度が多い事項について概ね対応できることから、基本的に LOGISTIC プロシジャを使用することで良いと考える。ただし、カテゴリー変数のダミー変数生成法が GLM 法でないことについては留意が必要である。なお、ユーザーガイドには、CATMOD プロシジャは分割表に対してのモデリングを想定しているため、連続変数を扱うことを推奨していない旨の記載がある[6]。

7. まとめ

本稿では回帰分析、t 検定、分散分析およびロジスティック回帰についての各プロシジャの概要とプロシジャ選択の判断材料となり得る情報を示した。解析時のプロシジャ選択を再考いただくきっかけになれば幸いである。

8. 参考文献

[1] SAS/STAT 15.2 User's Guide The GLM Procedure ODS Graphics
https://documentation.sas.com/doc/en/statug/15.2/statug_glm_details71.htm （最終アクセス日 2025/8/31）
[2] 中川雄貴 デザイン行列 SAS ユーザー総会 2024
https://sas-user2024.ywstat.jp/download.html?n=45&key=paper （最終アクセス日 2025/8/31）
[3] REG プロシジャで CLASS ステートメントがなかったり，交互作用の事前処理が必要な点は HPREG で改善されているよという話データステップ 100 万回 SAS 新手一生
https://sas-tumesas.blogspot.com/2025/03/regclasshpreg.html （最終アクセス日 2025/8/31）
[4] 高田浩成 重回帰分析における多重共線性について 第 9 回大阪 SAS 勉強会
https://www.docswell.com/s/6484025/5P9DDQ-2024-03-26-062816 （最終アクセス日 2025/8/31）
[5] 横溝孝明 他 臨床試験データへの GLMSELECT procedure の適用 SAS ユーザー総会 2010
https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups10-a-06.pdf （最終アクセス日 2025/8/31）
[6] SAS/STAT User's Guide
https://documentation.sas.com/doc/en/pgmsascdc/v_064/statug/titlepage.htm （最終アクセス日 2025/8/31）
[7] 臨床評価研究会(ACE) 基礎解析分科会 新版 実用 SAS 生物統計ハンドブック 第 2 版
[8] 杉本真理 t 検定の基礎 (Student の t 検定と Welch の t 検定) 第 9 回大阪 SAS 勉強会
https://www.docswell.com/s/6484025/ZQ8DYN-2024-03-28-051934 （最終アクセス日 2025/8/31）
[9] TTEST プロシジャの Welch 検定が、GLM,MIXED プロシジャのものと異なる。 SAS FAQ
https://www.sas.com/offices/asiapacific/japan/service/technical/faq/list/body/stat027.html （最終アクセス日 2025/8/31）
[10] Jack S. Nyberg The Paired T-Test: Does PROC MIXED Produce the Same Results as PROC TTEST? PharmaSUG 2004
https://www.lexjansen.com/pharmasug/2004/Posters/PO05.pdf （最終アクセス日 2025/8/31）
[11] SAS/STAT 15.2 User's Guide PROC GLM: PROC GLM for Unbalanced ANOVA
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_glm_sect005.htm （最終アクセス日 2025/8/31）
[12] SAS でロジスティック回帰いろいろ ぷる日記
https://prunus1350.hatenablog.com/entry/2015/03/15/225706（最終アクセス日 2025/8/31）
[13] 吉田早織 魚住龍史 線形モデルにおける CLASS ステートメントの機能
https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups14-a-03.pdf （最終アクセス日 2025/8/31）

用量漸増デザインにおける
BHLRM+EXNEXの動作特性の評価

○吉田和正[1,4]，○石田弘輔[2,4]，○相澤愛奈[3,4]

([1]日本ベーリンガーインゲルハイム株式会社，[2]ノバルティス ファーマ株式会社，
[3]ファイザーR&D合同会社，[4]日本製薬工業協会医薬品評価委員会 データサイエンス部会)

Evaluation of the Operating Characteristics of BHLRM + EXNEX in Dose-escalation Designs

Kazumasa Yoshida[1,4], Kosuke Ishida[2,4], Mana Aizawa[3,4]
[1]Nippon Boehringer Ingelheim Co., Ltd., [2]Novartis Pharma K.K., [3]Pfizer R&D Japan G.K.
[4]Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association

要旨

がんの第I相試験は一般的に用量制限毒性(DLT)の発現率を指標とし，最大耐量（MTD）または第II相推奨用量(RP2D)の決定を目的として実施される。近年では従来の3+3デザインなどのrule-basedのデザインに加えて統計モデルにもとづくデザインが利用されている。MTD/RP2Dの推定に使用される統計モデルの一つにBayesian Logistic Regression Model（BLRM）がある。BLRMはDLTの発現率にロジスティック回帰モデルを仮定することで柔軟な用量反応関係を表現できる。BLRMの拡張として，階層を用いて被験者の異質性を考慮するBayesian Hierarchical Logistic Regression Model（BHLRM），さらに交換可能(Exchangeable)な分布と非交換可能(non-exchangeable)な分布を混合した分布を用いるBHLRM+EXNEXがある。これらは，がん種や投与レジメン，剤型の違い，複数の過去試験の存在などによりDLTの傾向が異なる可能性がある場合でも情報を適切に借用し，より広範なデータに基づいた意思決定を可能とする。本発表では，BHLRM+EXNEXを用いた仮想的な第I相用量漸増試験を設定し，Early considerationを踏まえて複数のシナリオの下で実行した動作特性のシミュレーション結果を示す。

キーワード：ベイズ統計，階層モデリング，shrinkage，exchangeability-nonexchangeability，ベイジアンロジスティックモデル，BLRM，BHLRM＋EXNEX，がん第 I 相試験，用量漸増

1. はじめに

がん領域における第 I 相試験は, 治験薬の安全性評価および最大耐用量（MTD）や第 II 相推奨用量（RP2D）の決定を目的として行われる。従来の 3+3 デザインをはじめとした Rule-based 手法は実装が容易である一方で, 統計的な効率性や柔軟性の上で課題がある。このような背景から, 近年では連続再評価法（CRM）, Bayesian Logistic Regression Model（BLRM）などの Model-based 手法，また BOIN，mTPI といった Model-assisted 手法が広く採用されつつある。それぞれの手法において利点と欠点が存在し，試験薬の情報や過去類似試験の実施の有無等によって使い分けることが求められる。

本論文では，特に BLRM について取り上げる。BLRM は用量と毒性（DLT）発現の関係をロジスティック回帰モデルで表現する。ベイズ的枠組みを用いることで，事前分布を設定し，新たなデータ取得に応じて事後分布を更新する適応的設計が可能となる。BLRM では，Escalation with Overdose Control（EWOC）の枠組みで DLT 発現確率が閾値を超えるリスクを制限することで，過量投与の確率を抑える設計が可能となり，患者の安全性を確保することができる。

BLRM の拡張モデルとして BHLRM，BHLRM+EXNEX がある。これらは異なるがん腫や剤型，地域をまたぐデータ統合の必要性を背景に，BLRM を階層化し，さらに BHLRM+EXNEX によって情報共有の程度を調整可能とする構造を導入することができる。

そこで，がん第 I 相試験での適用を想定した仮想的な状況における動作特性について評価を行う。

2. がん第 I 相試験のシミュレーションの留意点

用量漸増デザインの適切性について，統計学的な観点よりプロトコル等に最低限含めるべき内容として，医薬品医療機器総合機構(PMDA)より「30 日調査照会事項チェックリスト（抗悪性腫瘍剤分野）[1]」,「抗悪性腫瘍剤の第 I 相試験計画時の統計学的留意事項（Early Consideration）[2]」が公表されている。それらは用量漸増デザインが被験者の安全性確保の観点で問題ないかを判断するために含めるべきとされているが，記述すべき内容としてシミュレーションによる動作特性の評価が含まれる。記述すべき内容として挙げられている例として以下がある。

1. デザイン仕様の明確化（用量漸増ルール，MTD 定義，終了基準，DLT 評価期間など）
2. 動作特性の評価（各用量での MTD 選択率，平均被験者数，平均 DLT 数，中止割合など）
3. 過毒性シナリオにおける中止の妥当性（最低用量が過量投与の閾値をわずかに超えるケース）

このうち，1 や 3 はシミュレーションの設定に関係することで，2 は出力すべき結果に関係する。

以降の章において PMDA に求められている記述について含めた上で，BHLRM，BHLRM+EXNEX の動作特性について評価を行う。

3. 解析で用いる統計モデル

本章では, 事前データ及び新規試験の観測データにおける異質性を考慮した BLRM の拡張である BHLRM, 及び BHLRM に交換可能(exchangeability)及び非交換可能(Non-exchangeability)な分布を混合した分布を用いる BHLRM+EXNEX を解説する [3,4]。

3.1. BLRM

用量l $(l = 1, \ldots, L)$に対して，その用量の DLT 評価可能被験者数をn_l，DLT 発現例数をY_lとおく。Y_l を以下の確率法則を満たす確率変数とする。

$$Y_l|\pi(d_l) \sim Bin\left(n_l, \pi(d_l)\right)$$

$$logit\left(\pi(d_l)\right) = log(\alpha) + \beta log\left(\frac{d_l}{d_*}\right)$$

ここで，d_*をオッズ比の基準(参照用量)とする。パラメータ $\left(log(\alpha),\ log(\beta)\right)^t$ の事前分布が，以下の正規分布に従うと仮定する。

$$\left(log(\alpha),\ log(\beta)\right)^t \sim N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right)$$

ここで用いられる事前分布のパラメータ$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$は，ユーザーが指定する。なお，弱情報事前分布となるように，分散パラメータ$(\tau_\alpha, \tau_\beta)$は$2^2$など大きな値を指定する。新試験開始前に事前情報となる既存のデータがある場合には，上記の設定の事前分布を既存のデータでベイズ更新して，新試験の事前分布として利用することも出来る。$(\alpha,\ \beta)$は確率変数であり，$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$はユーザーが既存又は新試験のデータでベイズ更新する前の分布を指定するために指定するパラメータである(確率変数でない)点に注意する。

BLRM モデルのパラメータ，確率変数の依存関係を図 1 に示す。なお，確率変数であるノードを黒色で表し，それ以外の固定された値を青色で表している。なお，ユーザーによるd_*の選択は，事前分布のモデルパラメータの解釈及び First-in-human (FIH)試験では試験開始初期の動作特性に影響を与えることを強調するために，ユーザーの入力を表す青色のノードとした。

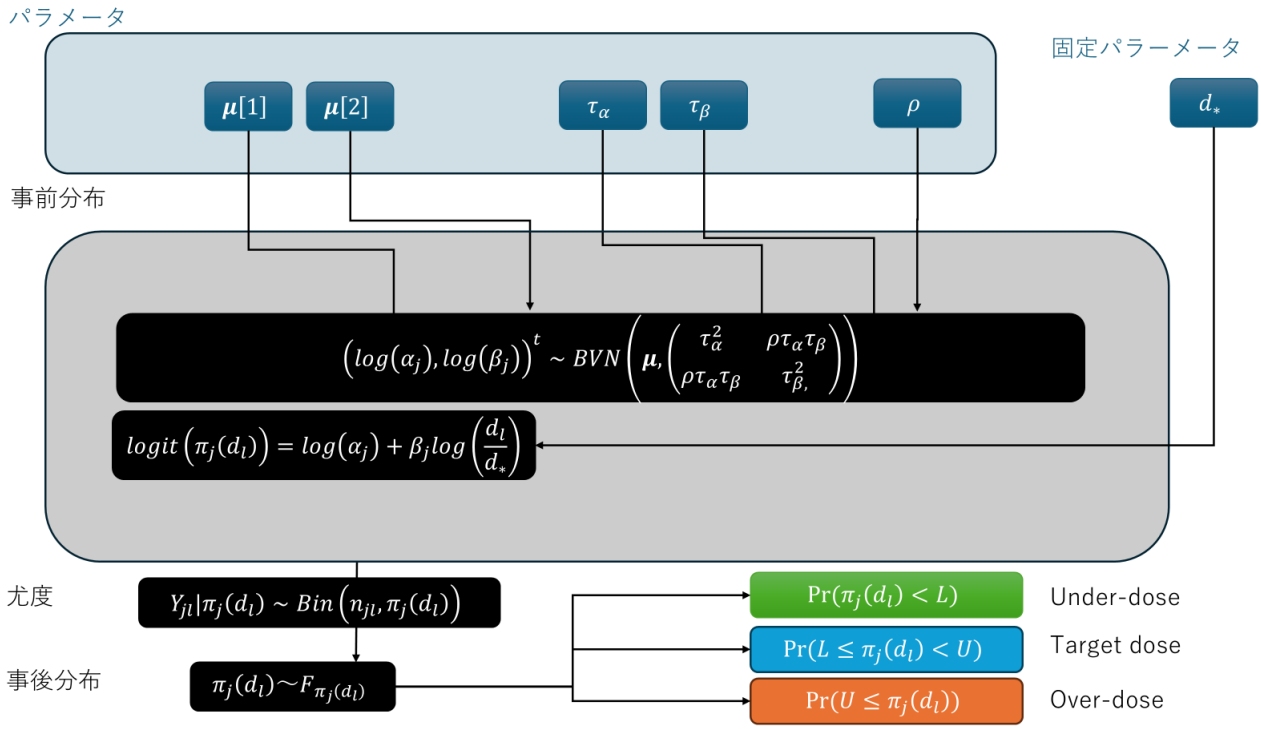


図 1. BLRM のパラメータ，確率変数の依存関係

確率変数α, βの関数である$\pi(d_l)$もまた確率変数であり，ある分布を持つ。事前分布及び事後分布を用いることで, 用量$l\ (l = 1, \dots, L)$に対して$\pi(d_l)$が以下の範囲にある確率を数値化することが可能であり , 順に under-dose の確率，target dose の確率，over-dose の確率と解釈される。

$$\Pr\left(\pi_j(d_l) < L\right)$$
$$\Pr\left(L \le \pi_j(d_l) < U\right)$$
$$\Pr\left(U \le \pi_j(d_l)\right)$$

ここで，$[L, U)$はターゲット毒性確率の範囲を表し，$[0.16, 0.33)$などが設定される。

3.2. BHLRM

既存又は新試験が異なる疾患・投与レジメンの症例を含む場合など，データの異質性を考慮する必要がある場合も考えられる。疾患・レジメンは異なっていても，DLT の発現傾向はある程度類似していると考えられる場合で，かつ相互の情報借用が望まれる場合などに BHLRM が候補となりえる。ここでは，既存又は新試験の症例で，いくつかのグループがあると仮定したモデルを解説する。なお，情報借用が可能だと考える複数の症例のまとまりを Neuenschwander, et.al.(2016)[4] で記される“strata”(層)ではなく，グループと呼んでいる点に注意する。

グループ$j(j = 1, \dots, J)$の用量$l\ (l = 1, \dots, L)$に対して，その用量の DLT 評価可能被験者数をn_{jl}，DLT 発現例数をY_{jl}とおく。Y_{jl} を以下の確率法則を満たす確率変数とする。

$$Y_{jl} | \pi_j(d_l) \sim Bin\left(n_{jl}, \pi_j(d_l)\right)$$
$$logit\left(\pi_j(d_l)\right) = log(\alpha_j) + \beta_j log\left(\frac{d_l}{d_*}\right)$$

ここで，d_*を参照用量とする。パラメータ$\left(log(\alpha_j),\ log(\beta_j)\right)^t$の事前分布が，以下の正規分布に従うと仮定する。

$$\left(log(\alpha_j),\ log(\beta_j)\right)^t | \boldsymbol{\mu},\ \tau_\alpha, \tau_\beta, \rho \sim N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right)$$

このうち，パラメータ$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$は BLRM と異なり確率変数となっている点に注意する。さらに，$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$は以下の超事前分布に従うものとする。

$$\boldsymbol{\mu} \sim N\left(\boldsymbol{m_\mu},\ \boldsymbol{S_\mu}\right)$$
$$log(\tau_\alpha) \sim N(m_{\tau\alpha}, \sigma_{\tau\alpha}^2)$$
$$log\left(\tau_\beta\right) \sim N\left(m_{\tau\beta}, \sigma_{\tau\beta}^2\right)$$
$$\rho \sim LKJ(\eta)$$

ここで用いられるパラメータ$(\boldsymbol{m_\mu},\ \boldsymbol{S_\mu}, m_{\tau\alpha}, \sigma_{\tau\alpha}^2, m_{\tau\beta}, \sigma_{\tau\beta}^2, \eta)$は，ユーザーが指定するパラメータであり，確率変数ではない点に注意する。Lewandowski-Kurowicka-Joe (LKJ)分布は相関係数の事前分布に使われる分布であり，特にη=1 のときは一様分布となる [5]。BLRM と異なり，$(\alpha_j,\ \beta_j)$がグループj 毎に定義されており，$\left(log(\alpha_j),\ log(\beta_j)\right)$はある正規分布からの$j$個目の(条件付きの)サンプリングとしている。その正規分布を特徴づけるパラメータ$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$にも確率を与えることで，事前又は観測データに依存した正規分布を推測させるモデルであることに注意する。なお，$\left(log(\alpha_j),\ log(\beta_j)\right)$は同じ(条件付き)正規分布を仮定しているため，グループごとのパラメータは交換可能であることを仮定している。

グループ変数が異なるレジメンや剤型を表し，d_lと DLT の関係がグループ毎に異なる場合が考えられる。例えば，異なる剤型の投与量(d_lとd_l'と表す)の間に$d_l = f(d_l')$のような変換式が仮定できる場合は，変換を行った後にモデル構築を行うことが推奨される。例えば，BID と QD のようにレジメンが異なるが，1 日当たりの投与量をd_lと置く場合や，異なる剤型の投与量が線形の関係にあると考えられる場合に変換を行う。なお，仮定する変換は超パラメータのばらつきを不必要に大きくしないことを目的としているが，仮定が適切でない場合は，超パラメータの分散の大きさに反映されることとなる。

確率変数の依存関係を図 2 に示す。BLRM と比べて，パラメータ$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$が確率変数となっている点，さらにその超事前分布のパラメータが追加されている点に注意する。

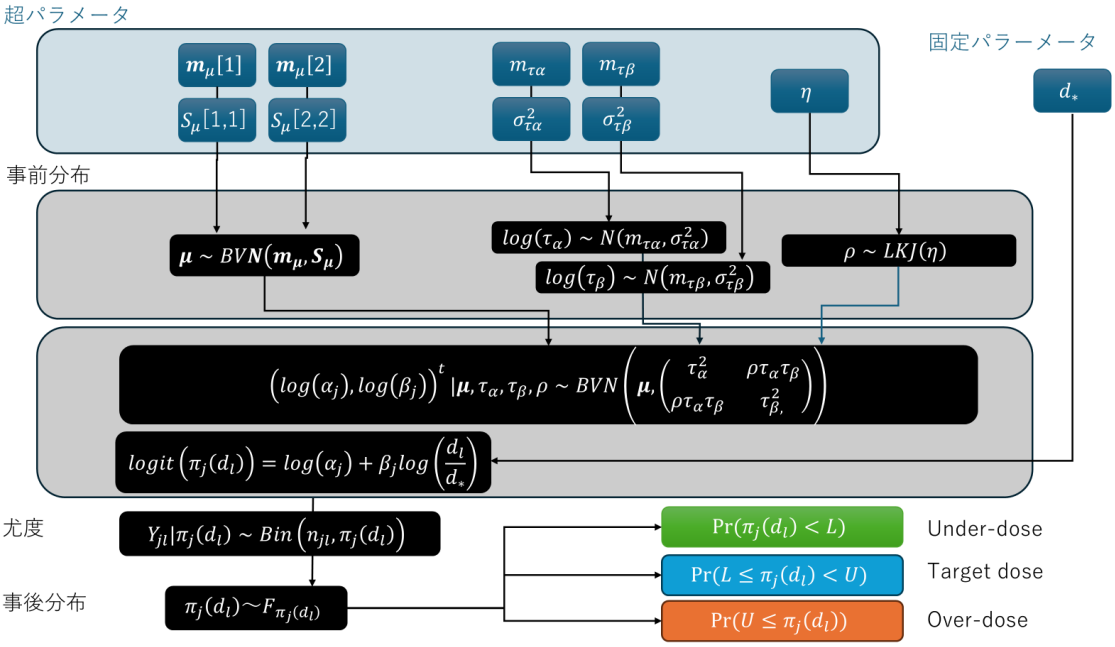


図 2　BHLRM のパラメータ，確率変数の依存関係

3.3. BHLRM + EXNEX

BHLRM では，すべてのグループ間で情報借用を行えるように交換可能性を仮定してモデリングしたが，情報借用の程度を明示的にコントロールしたい場合が考えられる。例えば，グループの一部で DLT の傾向が異なる可能性が否定できない場合などが考えられる。そのような場合に，交換可能な分布と非交換可能な分布の混合分布を利用することも考えられる。BHLRM を拡張して，パラメータの分布にある混合正規分布用いたモデルを BHLRM+EXNEX とここでは呼ぶこととする。臨床試験での使用例として Schöffski (2022)[6] などがあげられる。グループ$j(j = 1, \ldots, J)$の用量l $(l = 1, \ldots, L)$に対して，その用量の DLT 評価可能被験者数をn_{jl}，DLT 発現例数をY_{jl}とおく。Y_{jl} を以下の確率法則を満たす確率変数とする。

$$Y_{jl}|\pi_j(d_l) \sim Bin\left(n_{jl}, \pi_j(d_l)\right)$$

$$logit\left(\pi_j(d_l)\right) = log(\alpha_j) + \beta_j log\left(\frac{d_l}{d_*}\right)$$

ここで，d_*を参照用量とする。パラメータ$\left(log(\alpha_j),\ log(\beta_j)\right)^t$ の事前分布が以下の混合正規分布に従うと仮定する。

$$\left(log(\alpha_j),\ log(\beta_j)\right)^t \mid \boldsymbol{\mu},\ \tau_\alpha, \tau_\beta, \rho \sim p_j N\left(\boldsymbol{\mu}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix}\right) + (1-p_j)N(\boldsymbol{m}_j, \mathbf{S}_j)$$

このうち，パラメータ$(\boldsymbol{\mu}, \tau_\alpha, \tau_\beta, \rho)$は確率変数とし，以下の分布に従うものとする。なお，グループ毎に異なる事前分布のパラメータ$(\boldsymbol{m}_j, \mathbf{S}_j)$及び混合割合を表す$p_j$はユーザーが指定する。

$$\boldsymbol{\mu} \sim N(\boldsymbol{m}_{\boldsymbol{\mu}},\ \boldsymbol{S}_{\boldsymbol{\mu}})$$
$$log(\tau_\alpha) \sim N(m_{\tau\alpha}, \sigma_{\tau\alpha}^2)$$
$$log(\tau_\beta) \sim N(m_{\tau\beta}, \sigma_{\tau\beta}^2)$$
$$\rho \sim LKJ(\eta)$$

ここで用いられるパラメータ$(\boldsymbol{m}_{\boldsymbol{\mu}},\ \boldsymbol{S}_{\boldsymbol{\mu}}, m_{\tau\alpha}, \sigma_{\tau\alpha}^2, m_{\tau\beta}, \sigma_{\tau\beta}^2, \eta)$は，ユーザーが指定する。$p_1 = \cdots = p_J = 1$とするとBHLRM モデルとなり，$p_1 = \cdots = p_J = 0$とすると$J$ 個の個別の BLRM となる。

BHLRM+EXNEX モデルのパラメータ，確率変数の依存関係を図 3 に示す。超パラメータにグループjのパラメータのみが表記してあるが，実際にはそれぞれJ 個指定する必要がある点に注意する。

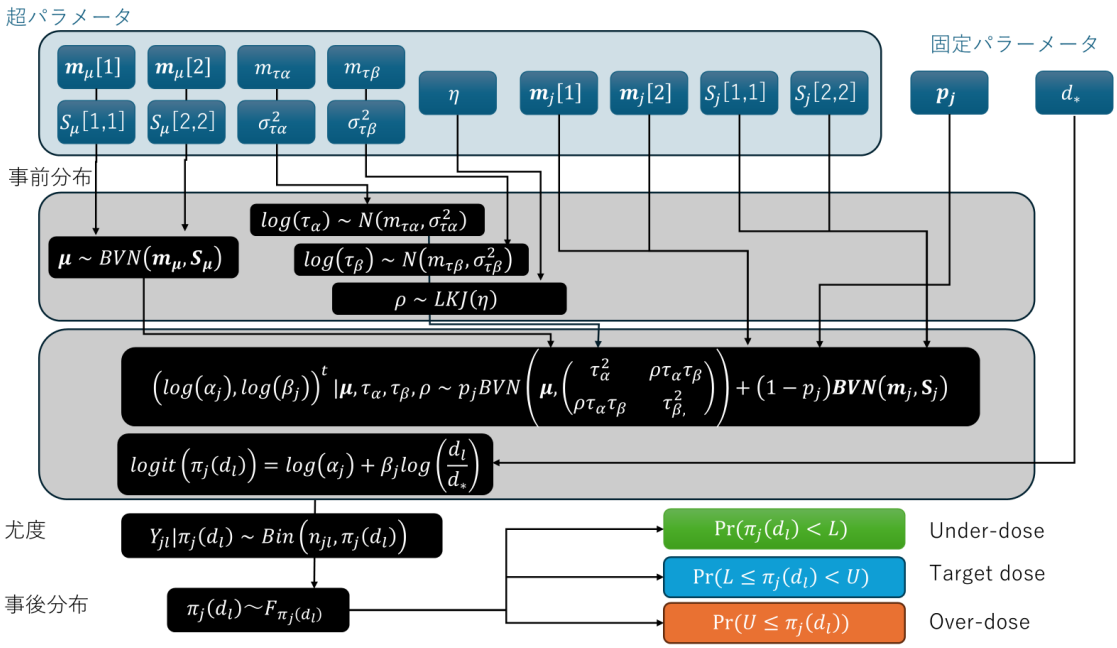


図 3　BHLRM+EXNEX のパラメータ，確率変数の依存関係

4. シミュレーション

BHLRM+EXNEX を用いた用量探索デザインを適用した場合の動作特性を確認するため，仮想的な第 I 相用量漸増試験を設定し，シミュレーションを実施した。

4.1. 仮想的な試験設定

試験目的

薬剤 A の単剤療法（1 日 1 回［QD］又は 1 日 2 回［BID］）における漸増用量の安全性及び忍容性について評価し，MTD を決定する。本シミュレーションでは QD での投与のデータがある状態で，BID での投与に切り替える状況を想定し，BID コホートを追加した場合の検討を実施することとした。

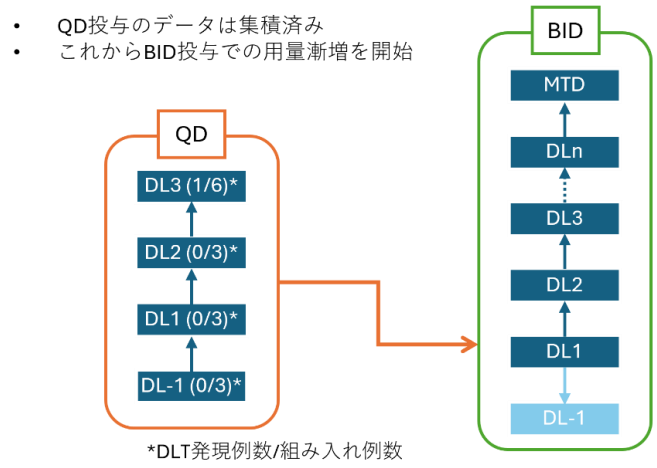


図 4　試験概略図

試験デザイン

- 患者数の設定：

最大で 60 例まで患者を組み入れることとする。コホートサイズは 3～6 例とした。

- 漸増・減量基準：

EWOC 原則（当該用量が OD となる確率が 25%未満である）に基づく BHLRM+EXNEX を用いる。全ての用量レベルについて DLT 発現確率が[0, 0.16)の区間を過少用量［under dose: UD］，[0.16, 0.33)の区間を目標用量［target dose: TD］，[0.33, 1]の区間を過大用量［over dose: OD］）とし，各用量区間に入る確率の事後確率を計算した。新たに組み入れる患者へ投与する用量レベルは，EWOC 原則に基づき決定した。

- 用量レベル：

検討する用量レベルは，以下の表 1 に示す用量とした。

表 1　用量漸増における暫定的な薬剤 A の用量レベル

用量レベル（DL）	用量（mg/day）
DL-1	10 (BID 1 回用量：5 mg)
DL1（開始用量）	20 (BID 1 回用量：10 mg)
DL2	40 (BID 1 回用量：20 mg)
DL3	80 (BID 1 回用量：40 mg)
DL4	120 (BID 1 回用量：60 mg)
DL5	160 (BID 1 回用量：80 mg)

- MTD／中止の基準：

以下の基準の両方を満たす場合には，用量d_lを MTD とみなし試験を中止する。

- 少なくとも 6 例の患者が用量d_lで投与を受けた。

- 用量d_lで以下のいずれかを満たした。
    - 用量d_lが TD である確率が 50%を超えている。
    - 少なくとも 15 例の患者が本試験で投与を受けており，用量d_lが EWOC 基準を満たす最大用量である。

なお, これらの条件を満たさない場合でも, 検討する用量レベルがすべて EWOC 基準を満たさなかった場合, 又は最大被験者数に達した場合には，MTD を決定せずに試験を中止した。

4.2. シミュレーション設定

統計モデル

統計モデルは BHLRM＋EXNEX モデルを用いることとした。用量レベル l (l = -1, 1,…,5)に対応する用量をd_lとし，用量d_lでの薬剤 A の DLT 発現率について，用量-DLT モデルとしてロジスティックモデルを仮定した。なお，用量レベルは 1 日あたりの総量で換算するものとした。参照用量d_*を 80mg/day（DL3）とした。また，グループに該当する j (j=1, 2) は投与スケジュール（1 = QD, 2 = BID）を表すものとした。各パラメータの事前分布については，3.3 項に示した分布に従うものとした。なお，パラメータ$(\boldsymbol{m}_j, \mathbf{S}_j)$は投与スケジュールごとに異なる分布を持つものとした。モデルパラメータの事前分布の設定は，以下のように与えることとした。

表 2　モデルパラメータの事前分布

パラメータ	値	意味
$m_\mu[1]$	$logit(0.3)$	$\mu[1]$の事前平均
$m_\mu[2]$	log (1)	$\mu[2]$の事前平均
$S_\mu[1,1]$	2	$\mu[1]$の標準偏差
$S_\mu[2,2]$	1	$\mu[2]$の標準偏差
$m_{\tau\alpha}$	log (0.25)	τ_αの事前平均
$m_{\tau\beta}$	log (0.125)	τ_βの事前平均
$\sigma_{\tau\alpha}$	log(2) /1.96	τ_αの標準偏差
$\sigma_{\tau\beta}$	log(2) /1.96	τ_βの標準偏差
η	1	相関ρの事前パラメータ（LKJ 分布）
$m_j[1]$	$logit(0.3)$	α_jの非交換可能な部分に対する事前平均
$m_j[2]$	0	β_jの非交換可能な部分に対する事前平均
$S_j[1,1]$	2	α_jの非交換可能な部分に対する標準偏差
$S_j[2,2]$	1	β_jの非交換可能な部分に対する標準偏差

また，BHLRM+EXNEX の特徴として，交換可能事前分布の混合割合p_jを設定することができる点が挙げられる。この値を 1 とした場合は単純な階層ベイズを用いた BHLRM となり，0 とした場合にはグループ間の情報の借り入れを考慮せず独立に各試験を評価する BLRM となる。実際の試験においては，交換可能事前分布の混合割合は臨床的な観点などを総合的に判断して決定される。本稿ではp_jの値を 0.8，1 及び 0 とした場合のシミュレーション結果を示す。なお，プログラムの動作の都合上，0 を設定できなかったた

め，本シミュレーションでは 0.000001 と設定した。

さらに，これまでに得られている QD 投与におけるデータとして，以下を用いた。

表 3　QD 投与のデータ

投与スケジュール	用量（mg/day）	被験者数	DLT 発現例数
QD	10	3	0
QD	20	3	0
QD	40	3	0
QD	80	6	1

表 2 の値を表 3 の情報を用いて更新することにより得られたモデルパラメータの事前分布を，これから実施する BID コホートでのモデルパラメータの事前分布として用いることとした。

シナリオ設定

動作特性を確認するにあたり，以下の 5 つのシナリオについてシミュレーションを実施した。

表 4　真の DLT 発現確率のシナリオ

シナリオ	各用量レベルにおける真の DLT 発現確率					
	10 mg/day	20 mg/day	40 mg/day	80 mg/day	120 mg/day	160 mg/day
1	0.05	0.1	0.2	0.3	0.4	0.6
2	0.33	0.4	0.5	0.6	0.7	0.8
3	0.025	0.05	0.1	0.15	0.2	0.3
4	0.025	0.05	0.075	0.1	0.25	0.5
5	0.1	0.2	0.4	0.5	0.6	0.7

シナリオ 1 は，各用量レベルにおける真の DLT 発現確率が UD，TD ，OD の全ての区間に分散されるように設定したシナリオである。シナリオ 2 は，検討する最低用量でも OD となるような，過毒性の場合のシナリオである。シナリオ 3 は，検討する最大用量でも OD とならない場合のシナリオである。シナリオ 4 は，検討する最高用量のみが OD となる場合のシナリオである。シナリオ 5 は，開始用量より高い用量が OD となる場合のシナリオである。

p_jの値を変えた場合の各用量における DLT 発現確率の事前分布，及び各シナリオにおける用量ごとの DLT 発現確率を以下に示す。

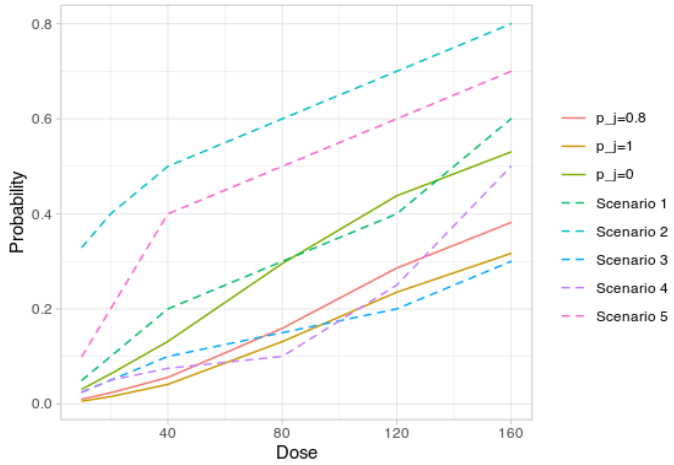


図 5　混合割合p_j別の DLT 発現確率の事前分布（中央値）／シナリオ別の DLT 発現確率

4.2.1.　シミュレーションプログラム

R の OncoBayes2 Package[7] を利用し，BHLRM+EXNEX の動作特性を示すプログラムを作成した。OncoBayes2 Package は，抗がん剤領域におけるベイジアンロジスティック回帰モデルを用いた用量漸増試験に向けて開発されたパッケージであり, オプションとして EXchangeability-NonEXchangeability（EXNEX）パラメータモデルを使用することで，ヒストリカルデータ及び並行して行われる試験のデータなどから柔軟な情報借用を可能としている。詳細は OncoBayes2 Package の vignette[8] を参照されたい。

4.3. シミュレーション結果

4.3.1. 各シナリオの下での動作特性

交換可能事前分布の混合割合p_jを変えた下での，表 4 で設定した 5 つのシナリオに対する BHLRM＋EXNEX を用いたシミュレーション実験の結果を以下に示す。また，結果は 3 つの用量区間及び用量レベルごとに提示している。

表 5　各シナリオに対する BHLRM+EXNEX の動作特性

シナリオ	MTD として選択される割合（%）			中止の割合 (%)		平均被験者数				平均 DLT 発現例数			
	UD	TD	OD	過毒性[a]	最大被験者数[b]	UD	TD	OD	計	UD	TD	OD	計
1（$p_j = 0.8$）	22.9	70.2	1.8	5.1	0.0	7.9	9.7	0.8	18.4	0.7	2.2	0.4	3.3
2（$p_j = 0.8$）	-	-	22.5	77.5	0.0	-	-	9.4	9.4	-	-	3.6	3.6
3（$p_j = 0.8$）	67.4	31.6	-	1.0	0.0	17.5	4.0	-	21.5	1.7	1.0	-	2.6
4（$p_j = 0.8$）	69.9	21.8	7.1	1.2	0.0	17.8	2.7	2.6	23.2	1.3	0.7	1.3	3.3
5（$p_j = 0.8$）	6.4	53.2	19.2	21.2	0.0	2.0	8.8	4.3	15.2	0.2	1.7	1.8	3.7
1（$p_j = 1$）	27.3	65.0	1.9	5.8	0.0	6.9	7.5	0.7	15.0	0.7	1.7	0.3	2.7
2（$p_j = 1$）	-	-	25.0	75.0	0.0	-	-	8.5	8.5	-	-	3.3	3.3
3（$p_j = 1$）	73.3	26.1	-	1.0	0.6	15.2	3.5	-	18.7	1.4	0.9	-	2.3

シナリオ	MTD として選択される割合（%）			中止の割合 (%)		平均被験者数				平均 DLT 発現例数			
	UD	TD	OD	過毒性[a]	最大被験者数[b]	UD	TD	OD	計	UD	TD	OD	計
4（$p_j = 1$）	70.8	19.7	8.4	1.1	0.0	17.9	2.8	2.7	23.4	1.3	0.7	1.4	3.4
5（$p_j = 1$）	5.4	54.5	17.0	23.1	0.0	1.8	8.7	4.4	14.8	0.2	1.7	1.8	3.7
1（$p_j = 0$）	28.8	64.7	1.4	5.1	0.0	7.1	7.4	0.5	14.9	0.7	1.7	0.2	2.6
2（$p_j = 0$）	-	-	25.0	75.0	0.0	-	-	8.8	8.8	-	-	3.4	3.4
3（$p_j = 0$）	74.6	23.8	-	1.4	0.2	15.0	3.2	-	18.2	1.4	0.8	-	2.2
4（$p_j = 0$）	71.0	18.9	8.8	1.3	0.0	17.7	2.7	2.5	22.9	1.2	0.7	1.2	3.2
5（$p_j = 0$）	4.8	57.6	15.3	22.3	0.0	1.7	8.9	4.1	14.7	0.2	1.9	1.7	3.7

a.検討する最低用量でも EWOC 基準が満たされず，中止となった割合。

b.最大被験者数に到達したが MTD が決定できず，中止となった割合。

表 6　用量ごとの各シナリオに対する BHLRM+EXNEX の動作特性（MTD 選択割合）

	MTD として選択される割合（%）					
シナリオ/用量(mg/day)	10	20	40	80	120	160
1（$p_j = 0.8$）	0.9	22.0	55.8	14.4	1.2	0.6
2（$p_j = 0.8$）	7.5	13.4	1.5	0.1	0.0	0.0
3（$p_j = 0.8$）	0.0	5.5	35.3	26.6	10.5	21.1
4（$p_j = 0.8$）	0.2	3.0	22.4	44.3	21.8	7.1
5（$p_j = 0.8$）	6.4	53.2	18.8	0.4	0.0	0.0
1（$p_j = 1$）	1.1	26.2	53.1	11.9	1.6	0.3
2（$p_j = 1$）	5.2	17.0	2.8	0.0	0.0	0.0
3（$p_j = 1$）	0.3	7.0	40.2	24.8	9.2	16.9
4（$p_j = 1$）	0.0	3.9	20.8	46.1	19.7	8.4
5（$p_j = 1$）	5.4	54.5	16.4	0.6	0.0	0.0
1（$p_j = 0$）	1.3	27.5	54.2	10.5	1.2	0.2
2（$p_j = 0$）	7.5	15.7	1.8	0.0	0.0	0.0
3（$p_j = 0$）	0.1	9.2	39.5	25.8	8.0	15.8
4（$p_j = 0$）	0.1	3.4	22.9	44.6	18.9	8.8
5（$p_j = 0$）	4.8	57.6	14.7	0.6	0.0	0.0

表 7　用量ごとの各シナリオに対する BHLRM+EXNEX の動作特性（平均被験者数）

	平均被験者数						
シナリオ/ 用量(mg/day)	10	20	40	80	120	160	計
1（$p_j = 0.8$）	0.6	7.3	7.4	2.3	0.5	0.3	18.4
2（$p_j = 0.8$）	2.0	6.4	0.9	0.04	0.0	0.0	9.4
3（$p_j = 0.8$）	0.2	5.9	7.1	4.4	1.7	2.3	21.5
4（$p_j = 0.8$）	0.2	5.6	6.4	5.5	2.7	2.6	23.2
5（$p_j = 0.8$）	2.0	8.8	4.0	0.4	0.02	0.01	15.2
1（$p_j = 1$）	0.5	6.4	5.6	2.0	0.4	0.3	15.0
2（$p_j = 1$）	1.6	6.0	0.9	0.06	0.0	0.0	8.5
3（$p_j = 1$）	0.2	5.5	5.8	3.8	1.4	2.0	18.7
4（$p_j = 1$）	0.3	5.8	6.3	5.5	2.8	2.7	23.4
5（$p_j = 1$）	1.8	8.7	4.0	0.3	0.01	0.02	14.8
1（$p_j = 0$）	0.6	6.5	5.6	1.8	0.3	0.2	14.9
2（$p_j = 0$）	1.9	6.0	0.8	0.04	0.0	0.0	8.8
3（$p_j = 0$）	0.2	5.4	5.7	3.7	1.4	1.8	18.2
4（$p_j = 0$）	0.2	5.8	6.3	5.3	2.7	2.5	22.9
5（$p_j = 0$）	1.7	8.9	3.8	0.3	0.00	0.01	14.7

表 8　用量ごとの各シナリオに対する BHLRM+EXNEX の動作特性（平均 DLT 発現例数）

	平均 DLT 発現例数						
シナリオ/ 用量(mg/day)	10	20	40	80	120	160	計
1（$p_j = 0.8$）	0.03	0.7	1.5	0.7	0.2	0.2	3.3
2（$p_j = 0.8$）	0.6	2.5	0.4	0.02	0.0	0.0	3.6
3（$p_j = 0.8$）	0.0	0.3	0.7	0.6	0.3	0.6	2.6
4（$p_j = 0.8$）	0.01	0.3	0.5	0.5	0.7	1.3	3.3
5（$p_j = 0.8$）	0.2	1.7	1.6	0.2	0.01	0.01	3.7
1（$p_j = 1$）	0.03	0.6	1.1	0.6	0.2	0.2	2.7
2（$p_j = 1$）	0.6	2.3	0.4	0.04	0.0	0.0	3.3
3（$p_j = 1$）	0.00	0.3	0.6	0.5	0.3	0.6	2.3
4（$p_j = 1$）	0.01	0.3	0.4	0.5	0.7	1.4	3.4
5（$p_j = 1$）	0.2	1.7	1.6	0.2	0.01	0.01	3.7
1（$p_j = 0$）	0.04	0.7	1.2	0.6	0.1	0.1	2.6
2（$p_j = 0$）	0.6	2.4	0.4	0.02	0.0	0.0	3.4
3（$p_j = 0$）	0.01	0.3	0.6	0.6	0.3	0.5	2.2
4（$p_j = 0$）	0.01	0.3	0.4	0.5	0.7	1.2	3.2

	平均 DLT 発現例数						
シナリオ/ 用量(mg/day)	10	20	40	80	120	160	計
5（$p_j = 0$）	0.2	1.9	1.6	0.1	0.00	0.00	3.7

まず，$p_j = 0{,}8$の下での各シナリオに対する動作特性を評価する。シナリオ 1 では，TD が適切に MTD として選択される割合が 70.2%と最も高い結果であり，用量別では 40mg/day（DL2）が MTD として選択される割合が 55.8%と最も高かった。また，OD が MTD として選択される割合は 5.1%と低く，OD に割り付けられた平均被験者数は 0.8 例であり，過大用量への投与は抑制できていると考えられる。検討するすべての用量が OD となるシナリオ 2 では, 検討する最低の用量レベルでも過毒性となり中止となる確率が 77.5%と高く，また平均被験者数は 9.4 例，平均 DLT 発現例数は 3.6 例と少ない被験者数に抑えられていた。すべての用量が OD に満たないシナリオ 3 では, UD が MTD として選択される確率が 67.4%と最も高かったが, これは UD に相当する最高用量における DLT 発現確率が 0.15 と, TD に相当する DLT 発現確率の下限に近かったことが一因と考えられる。また，事前分布に用いた QD のデータにおいて，UD である 80mg/day（DL3）の用量で 6 例中 1 例に DLT の発現があったことも影響を与えている可能性がある。検討する最高用量のみが OD となる場合のシナリオ 4 では，シナリオ 3 と同様に UD が MTD として選択される確率が 69.9%と最も高かった。これは TD となる用量が 120mg/day（DL4）の 1 用量のみであったこと，事前分布に用いた QD のデータにおいて，UD の最高用量である 80mg/day（DL3）の用量で DLT 発現があったことなどが影響を与えてる可能性がある。開始用量より高い用量が OD となる場合のシナリオ 5 では, TD が MTD として選択される確率が 53.2%と最も高く，OD が MTD として選択される確率は 19.2%と低い結果であった。また，平均 DLT 発現例数も 3.7 例と少ない被験者数に抑えられていた。これらの傾向は，p_jの値を変えた下でも同様であった。この結果より，シナリオの設定によっては UD が MTD として選択される確率が高くなる場合もあり，やや保守的な傾向はあるものの，本デザインを用いることで MTD は適切に評価でき，また過大用量への投与についても抑制できると考えられる。

次に，各シナリオの下でp_jの値を変えた場合の動作特性について評価する。今回のシナリオの設定の下では，p_jの値を変えた場合でも，動作特性に大きな違いは見られなかったが，$p_j = 0.8$とした BHLRM+EXNEX モデルを用いた場合の方が，平均被験者数や平均 DLT 発現例数は同程度でありながらも，シナリオ 1，3 及び 4 では TD が適切に MTD として選択される割合が高かった。また，すべての用量が過毒性となるシナリオ 2 では，過毒性として適切に中止となる割合が高かった。一方，シナリオ 5 では，$p_j = 0$とした場合が TD を MTD として選択する割合が最も高く，また OD を MTD として選択する割合は低かった。これは，事前分布に用いたデータにおいて, OD である 40mg/day まで DLT 発現が認められなかったことが影響している可能性がある。実際の試験においては，交換可能事前分布の混合割合はグループ間の異質性や臨床的な観点などを総合的に判断して決定されるべきであろう。

5. まとめ

本稿では，抗がん剤第 I 相試験におけるベイズ流シミュレーションを要するデザインとして，BLRM に注目し，その拡張である BHLRM 及び BHLRM+EXNEX モデルを含めた理論的な背景及びシミュレーションによる動作特性を示した。今後，抗がん剤第 I 相試験を実施する上で，異なるがん種での評価，開発途中での

剤形又は投与レジメンの変更といった複雑な状況が生じる場面が増加することが予想される。そのような状況を考慮する際，本稿で紹介した BLRM 及びその拡張モデルは有用であると考える。抗がん剤第 I 相試験のデザインを検討する際，本稿がその一助となれば幸いである。

6. 参考文献

1. PMDA. 30日調査照会事項チェックリスト（抗悪性腫瘍剤分野），令和6年12月.
2. PMDA. 抗悪性腫瘍剤の第 I 相試験計画時の統計学的留意事項 ― 安全性の観点から（Early Consideration），令和6年12月.
3. Weber, S., Widmer, L., Bean, A. (2023). _OncoBayes2: Bayesian Logistic Regression for Oncology Dose-Escalation Trials_. R package version 0.8-9.
4. Neuenschwander, B., Wandel, S., Roychoudhury, S., and Bailey, S. (2016) Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceut. Statist.*, 15: 123-134.
5. Lewandowski, Daniel; Kurowicka, Dorota; Joe, Harry (2009). "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". Journal of Multivariate Analysis. 100 (9): 1989–2001.
6. Schöffski, P., Tan, D. S., Martín, M., et al. (2022). Phase I/II study of the LAG-3 inhibitor ieramilimab (LAG525)±anti-PD-1 spartalizumab (PDR001) in patients with advanced malignancies. Journal for immunotherapy of cancer, 10(2).
7. Sebastian, W., Lukas, AW., Andrew, B. Package 'OncoBayes2'. https://cran.r-project.org/web/packages/OncoBayes2/index.html
8. Andrew, B., Sebastian, W., Lukas, W., Guiding Oncology Dose-Escalation Trials. https://cran.r-project.org/web/packages/OncoBayes2/vignettes/introduction.html
9. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会（2025）．ベイズ流試験デザインにおけるシミュレーションの利用法.

IMLプロシジャで理解するCox回帰

○堀本隆一

（株式会社エス・シー・エー）

Understanding Cox Regression with the IML Procedure

Ryuichi Horimoto
S.C.A Co., Ltd

要旨

IML プロシジャを用いて、Cox 比例ハザードモデルによるハザード比の推定を行い、その推定過程について解説した。さらに、時間依存性共変量を含むモデルを構築し、時間依存ハザード比の推定も行った。

キーワード：Cox 比例ハザードモデル、時間依存性共変量、IML プロシジャ

1. はじめに

生存時間解析は臨床試験や疫学研究において必要不可欠な統計的手法である。その中でも Cox 比例ハザードモデルは、特定の分布仮定を設けることなく共変量の影響を評価できることから、最も広く利用されているモデルの一つである。実務では SAS の PHREG プロシジャを用いることで容易に Cox 回帰を実行することができるが、その推定過程については十分に理解しないまま使用する場合も少なくはない。

そこで本稿では、Cox 回帰に対する理解を深めることを目的として、SAS の IML プロシジャを用いて Cox 比例ハザードモデルによるハザード比の推定を行い、その推定過程について解説した。具体的には、部分尤度の構築、ニュートン・ラフソン法による最適化及び信頼区間の算出を IML プロシジャで実装し、それらについて解説した。さらに、比例ハザード性が成り立たない状況への拡張として、時間依存性共変量を導入したモデルを取り上げ、時間依存的なハザード比の推定も IML プロシジャで実装した。

2. Cox 比例ハザードモデル

2.1. 概要

Cox 比例ハザードモデルは、生存時間解析において広く利用されている半パラメトリックモデルであり、生存時間 t に対するハザード関数を次のように表す。

$$h(t|X) = h_0(t)\exp(\beta X) \tag{2.1.1}$$

ここで、$h_0(t)$ は基準ハザード（ベースラインハザード）と呼ばれるものであり、時間 t のみに依存する。ベースラインハザードとは、共変量が全て 0 の基準集団の時間 t におけるリスクを表す関数のことである。つまり、共変量の影響を除外した土台としてのリスクであり、その上に共変量の効果（$\exp(\beta X)$）が掛け合わされることでハザードが決まる。一方、$\exp(\beta X)$ は共変量 X の効果を指数関数的に表現しており、時間に依存しない比率として作用する。共変量 X が、薬物投与群の場合は 1、対照群の場合は 0 を取る変数とすると、薬物投与群と対照群のハザード比（Hazard Ratio）は次式で表される。

$$HR = \frac{h(t|X=1)}{h(t|X=0)} = \frac{h_0(t)\exp(\beta \cdot 1)}{h_0(t)\exp(\beta \cdot 0)} = \exp(\beta) \tag{2.1.2}$$

ここで、ベースラインハザード $h_0(t)$ は分母と分子に現れて相殺されるため、ハザード比は時間に依存せずに一定であるという比例ハザード（Proportional Hazards）の性質が導かれる。本モデルの特徴は、ベースラインハザードを特定の関数形に仮定せずに β を推定できる点にある。

2.2. 部分尤度の構築

Cox 比例ハザードモデルでは、部分尤度（Partial Likelihood）を構築し、それを最大化することでハザード比を推定する。部分尤度は、イベントが発生した瞬間における条件付き確率を積み上げることで構築される。

時刻 t_i にイベントが発生した個体 i を考える。ここで重要なのは、イベント発生時における条件付き確率を考える際、基準ハザードが分子と分母に共通して現れるため相殺される点である。例えば、時刻 t_i において個体 i がイベントを起こす確率は次式で表される。

$$\frac{h_0(t_i)\exp(\beta X_i)}{\sum_{j\in R(t_i)} h_0(t_i)\exp(\beta X_j)} \tag{2.2.1}$$

このとき、ベースラインハザード $h_0(t)$ は分子と分母で打ち消され、

$$\frac{\exp(\beta X_i)}{\sum_{j\in R(t_i)} \exp(\beta X_j)} \tag{2.2.2}$$

と簡略化される。したがって、Cox 比例ハザードモデルにおいては、ベースラインハザードを特定の関数形に仮定する必要がなく、共変量の効果 β のみを部分尤度に基づいて推定できる。この性質により、Cox 比例ハザードモデルは半パラメトリックモデルと呼ばれている。ここで、$R(t_i)$ はリスク集合（Risk Set）と呼ばれ、時間 t_i においてまだイベントが発生していない全ての個体の集合を指す。(2.2.2)式により、全てのイベントについての尤度は次式で与えられる。

$$L(\beta) = \prod_{i:\delta_i=0} \frac{\exp(\beta X_i)}{\sum_{j\in R(t_i)} \exp(\beta X_j)} \tag{2.2.3}$$

ここで、δ_i はイベントか打ち切りかを表す指示変数（イベント =0、打ち切り =1）である。また、両辺の対数を取ることで、計算で扱いやすい対数部分尤度が得られる。

$$l(\beta) = \sum_{i:\delta_i=0} \left[\beta X_i - \log\left(\sum_{j\in R(t_i)} \exp(\beta X_j) \right) \right] \tag{2.2.4}$$

この対数部分尤度を最大化することで、回帰係数 β の推定値を得る。

対数部分尤度の構築について IML プロシジャでの実装コードを以下に示す。使用するデータセット ph_data は付録 A.1 のプログラムを用いて生成した。本データセットは trt、id、time、cnsr の 4 変数から成る。trt は薬物投与群の場合は 1、対照群の場合は 0 を取る。id は被験者 ID、time は生存時間を表す。cnsr はイベントの

場合は 0、打ち切りの場合は 1 を取る。また、薬物投与群と対照群のハザード比は時間を通じて一定（0.5）となるように作成した。

```
proc iml;
   /* #2.1. データの読み込み */
   use ph_data;
   read all var{trt time cnsr} into ph_data;
   call sort(ph_data, {2, 3}); /* time、cnsr の昇順にソート */

   trt = ph_data[, 1];
   time = ph_data[, 2];
   cnsr = ph_data[, 3];
   n = nrow(ph_data); /* 行数を取得 */

   /* #2.2. 部分尤度の構築 */
   start negLogLik(beta) global(trt, time, cnsr, n); /* 関数定義の開始 */
      totalLogLik = 0;
      do i = 1 to n;
         if cnsr[i] = 0 then do;
            S0 = 0;
            do j = i to n;
               S0 = S0 + exp(beta * trt[j]);
            end;
            totalLogLik = totalLogLik + beta*trt[i] - log(S0);
         end;
      end;
      return(-totalLogLik); /* 戻り値 */
   finish; /* 関数定義の終了 */
quit;
```

2.3. ニュートン・ラフソン法による最適化

推定量 $\hat{\beta}$ は部分尤度を最大化することで得られる。本節では、PHREG プロシジャと同様に、ニュートン・ラフソン法を用いて部分尤度を最大化する解を求める。部分尤度の対数を $l(\beta)$ とすると、推定量 $\hat{\beta}$ は次の条件を満たす解である。

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = 0 \tag{2.3.1}$$

ニュートン・ラフソン法は、(2.3.1)式のスコア関数 $U(\beta)$ と観測情報行列 $I(\hat{\beta})$ を用いて次の反復式を実行することで推定量 $\hat{\beta}$ を求める。

$$\beta^{(k+1)} = \beta^{(k)} - I\left(\beta^{(k)}\right)^{-1} U\left(\beta^{(k)}\right) \tag{2.3.2}$$

これにより、β の更新を繰り返すことで収束解 $\hat{\beta}$ を得る。また、観測情報行列 $I(\hat{\beta})$ は次式で与えられる。

$$I(\hat{\beta}) = \sum_{i:\delta_i=0} \left\{ \frac{S^{(2)}(\hat{\beta}, t_i)}{S^{(0)}(\hat{\beta}, t_i)} - \left(\frac{S^{(1)}(\hat{\beta}, t_i)}{S^{(0)}(\hat{\beta}, t_i)} \right)^2 \right\} \tag{2.3.3}$$

ここで、$S^{(0)}(\hat{\beta}, t_i)$、$S^{(1)}(\hat{\beta}, t_i)$、$S^{(2)}(\hat{\beta}, t_i)$ は次式で表される。

$$S^{(0)}(\hat{\beta}, t_i) = \sum_{j \in R(t_i)} \exp(\hat{\beta} X_j), \quad S^{(1)}(\hat{\beta}, t_i) = \sum_{j \in R(t_i)} X_j \exp(\hat{\beta} X_j), \quad S^{(2)}(\hat{\beta}, t_i) = \sum_{j \in R(t_i)} X_j^2 \exp(\hat{\beta} X_j). \tag{2.3.4}$$

IML プロシジャでは、call nlpnra ルーチンを用いることでニュートン・ラフソン法による最適化を実行することができる。実装コードを以下に示す。

```
proc iml;
   /* #2.1. データの読み込み*/
   /* #2.2. 部分尤度の構築 */

   /* #2.3. ニュートン・ラフソン法による最適化 */
   beta0 = 0; /* 初期値 */
   optn = {0 2}; /* 最適化オプション */
   call nlpnra(rc, betaHat, "negLogLik", beta0, optn);
   HR = exp(betaHat);
   print betaHat HR;
quit;
```

call nlpnra ルーチンの引数の意味は次のとおりである。第一引数 rc は最適化の終了コードが格納される変数であり、収束状況を確認することができる。第二引数 betaHat は推定されたパラメータが格納される変数である。第三引数"negLogLik"は最適化したい目的関数であり、2.3 節で定義した対数部分尤度を戻り値として返す関数を指定した。第四引数 beta0 はパラメータの初期値を設定する変数であり、0 を設定した。第五引数 optn は最適化の挙動を指定するベクトルであり、optn[1]＝0 で最小化問題を指定し、optn[2]＝2 で収束過程における最初と最後のパラメータを出力することを指定した。

実行結果として、推定値 $\hat{\beta} = -0.591$ が得られた。推定された回帰係数からハザード比は $HR = \exp(\hat{\beta})$ として算出される。したがって、得られたハザード比は $HR = \exp(\hat{\beta}) = 0.554$ であった。これは、データセット ph_data で設定した「薬物投与群と対照群のハザード比は時間を通じて一定（0.5）である」という真値を適切に推定できているといえる。

2.4. 信頼区間の算出

推定量 $\hat{\beta}$ の信頼区間は(2.3.3)式の観測情報行列 $I(\hat{\beta})$ を用いて構築する。推定量の分散は観測情報行列の逆数で与えられる。

$$Var(\hat{\beta}) = \frac{1}{I(\hat{\beta})} \tag{2.4.1}$$

したがって、標準誤差は、$SE(\hat{\beta}) = \sqrt{Var(\hat{\beta})}$ であり、95%信頼区間は、$\hat{\beta} \pm 1.96 \times SE(\hat{\beta})$ として求められる。これをハザード比に変換すれば、$[\exp(\hat{\beta} - 1.96 \times SE), \exp(\hat{\beta} + 1.96 \times SE)]$ が得られる。

信頼区間の算出について IML プロシジャでの実装コードを以下に示す。

```
proc iml;
   /* #2.1. データの読み込み*/
   /* #2.2. 部分尤度の構築 */
   /* #2.3. ニュートン・ラフソン法による最適化 */

   /* #2.4. 信頼区間の算出 */
   infoSum = 0;
   do i = 1 to n;
      if cnsr[i] = 0 then do;
         S0 = 0; S1 = 0; S2 = 0;
         do j = i to n;
            S0 = S0 + exp(betaHat * trt[j]);
            S1 = S1 + trt[j] * exp(betaHat * trt[j]);
            S2 = S2 + (trt[j]**2) * exp(betaHat * trt[j]);
         end;
         if S0 > 0 then infoSum = infoSum + (S2/S0 - (S1/S0)**2);
      end;
   end;

   varBeta = 1 / infoSum;
   SE = sqrt(varBeta);
   LCL_beta = betaHat - 1.96 * SE;
   UCL_beta = betaHat + 1.96 * SE;
   LCL_HR = exp(LCL_beta);
   UCL_HR = exp(UCL_beta);
   print SE LCL_beta UCL_beta LCL_HR UCL_HR;
quit;
```

実行結果として、推定量 $\hat{\beta}$ の標準誤差は 0.088 が得られた。これに基づく 95%信頼区間は、$\hat{\beta} \pm 1.96 \times SE = (-0.763, -0.418)$ となり、これをハザード比に変換すると、$[0.466, 0.658]$ となる。真のハザード比 0.5 が区間内に含まれており、推定の精度と妥当性が確認できる。

2.5. PHREG プロシジャの結果との比較

最後に、PHREG プロシジャによる結果と比較する。PHREG プロシジャでの実装コードを以下に示す。なお、IML プロシジャでの実装と同様に、データセット ph_data を使用した。

```
proc phreg data=ph_data;
   class trt(ref="0");
   model time*cnsr(1)=trt;
```

```
  hazardratio trt / diff=ref;
run;
```

実行結果として、ハザード比は 0.554 、ハザード比の 95%信頼区間は [0.466, 0.658] が得られ、IML プロシジャによる結果と一致することを確認した。これにより、Cox 比例ハザードモデルによる推定過程を IML プロシジャによって正しく再現できていることが確認できた。

3. 時間依存性共変量を含む Cox モデル

3.1. 背景

Cox 比例ハザードモデルは、「ハザード比が時間に依存せずに一定である」という比例ハザード性を仮定している。しかし、臨床試験データにおいては、治療効果が時間とともに変化するという状況がしばしば観察される。例えば、免疫チェックポイント阻害薬に代表される遅発効果がその典型例である。このような状況に対応するための手法として、モデルに時間依存性共変量を導入することを考える。

本章では、治療効果がある時点以降に発現することを仮定した Piecewise モデル、及び治療効果が時間とともに単調増加することを仮定した Log(time)モデルの 2 パターンについて、IML プロシジャを用いて時間依存的なハザード比の推定を行った。

3.2. Piecewise モデルの定式化と実装

本節では、治療効果が 6 か月以降に発現することを仮定したモデルを考える。ハザード関数は次のように表される。

$$h(t|X) = h_0(t)\exp\{\beta_1 X + \beta_2 X \cdot I(t > 6)\} \tag{3.2.1}$$

ここで、$I(t > 6)$ は時間 t が 6 か月を超えた場合は 1、それ以外は 0 を取る指示関数である。また、X は薬物投与群の場合は 1、対照群の場合は 0 を取る変数とする。このとき、(2.1.2)式と同様に、薬物投与群（$X = 1$）と対照群（$X = 0$）で比を取ることでベースラインハザード $h_0(t)$ が相殺されるため、6 か月以前のハザード比は $\exp(\beta_1)$ 、6 か月以降のハザード比は $\exp(\beta_1 + \beta_2)$ で表される。すなわち、β_2 は 6 か月を境とした治療効果の変化を表している。

本モデルでも、Cox 比例ハザードモデルと同様に、部分尤度を用いて推定を行う。イベントが発生した個体 i に対して、部分尤度に寄与する項は次のように表される。

$$\frac{\exp\{\beta_1 X_i + \beta_2 X_i \cdot I(t_i > 6)\}}{\sum_{j\in R(t_i)} \exp\{\beta_1 X_j + \beta_2 X_j \cdot I(t_i > 6)\}} \tag{3.2.2}$$

全てのイベントについて積を取ることで、Piecewise モデルに対応する部分尤度は次式のようになる。

$$L(\beta_1, \beta_2) = \prod_{i:\delta_i=0} \frac{\exp\{\beta_1 X_i + \beta_2 X_i \cdot I(t_i > 6)\}}{\sum_{j\in R(t_i)} \exp\{\beta_1 X_j + \beta_2 X_j \cdot I(t_i > 6)\}} \tag{3.2.3}$$

この対数部分尤度を最大化することにより、パラメータ β_1 、 β_2 の推定を行う。

標準的な Cox 比例ハザードモデルでは、時間に依存しない一定のハザード比を推定するが、Piecewise モデルでは、 X と指示関数 $I(t > 6)$ の交互作用項を導入することで時間依存的な治療効果の変化を表現することが可能である。部分尤度の構築においても、各イベント時点での寄与項に β_2 が組み込まれ、リスク集合の分母においても時間に応じて重み付けが変化する点が異なる。

IML プロシジャでの実装コードを以下に示す。使用するデータセット non_ph_data は付録 A.2 のプログラムを用いて生成した。なお、薬物投与群と対照群のハザード比は 6 か月時点までは 1、それ以降は 0.5 となるように作成した。その他の定義はデータセット ph_data と同様である。

```
proc iml;
  /* #3.1.1. データの読み込み */
  use non_ph_data;
  read all var{trt time cnsr} into non_ph_data;
  call sort(non_ph_data, {2, 3});

  trt = non_ph_data[, 1];
  time = non_ph_data[, 2];
  cnsr = non_ph_data[, 3];
  n = nrow(non_ph_data);

  /* #3.1.2. 部分尤度の構築 */
  start negLogLik(beta) global(trt, time, cnsr, n);
    totalLogLik = 0;
    beta1 = beta[1];
    beta2 = beta[2];
    do i = 1 to n;
      if cnsr[i] = 0 then do;
        ind = (time[i] > 6);
        S0 = 0;
        do j = i to n;
          S0 = S0 + exp(beta1*trt[j] + beta2*trt[j]*ind);
        end;
        totalLogLik = totalLogLik + beta1*trt[i] + beta2*trt[i]*ind - log(S0);
      end;
    end;
    return(-totalLogLik);
  finish;

  /* #3.1.3. ニュートン・ラフソン法による最適化 */
  beta0 = {0, 0};
  optn = {0, 2};
  call nlpnra(rc, betaHat, "negLogLik", beta0, optn);
  HR_pre = exp(betaHat[1]);
  HR_post = exp(betaHat[1] + betaHat[2]);
```

```
    print HR_pre HR_post;
quit;
```

実行結果として、6 か月以前のハザード比 HR_pre は 1.072、6 か月以降のハザード比 HR_post は 0.486 が得られた。これは、データセット non_ph_data で設定した「薬物投与群と対照群のハザード比は 6 か月時点までは 1、それ以降は 0.5 である」という真値を適切に推定できているといえる。

同様の推定を PHREG プロシジャで実装する場合、コードは以下のとおりになる。

```
proc phreg data=non_ph_data;
    class trt(ref="0");
    model time*cnsr(1) = trt trt*i;
    i = (time > 6);
    hazardratio trt / diff=ref at(i=(0 1));
run;
```

実行結果として、IML プロシジャによる結果と同じ値が得られた。これにより、Piecewise モデルによる推定過程を IML プロシジャによって正しく再現できていることが確認できた。

3.3. Log(time)モデルの定式化と実装

本節では、治療効果が時間とともに単調増加することを仮定したモデルを考える。ハザード関数は次のように表される。

$$h(t|X) = h_0(t)\exp\{\beta_1 X + \beta_2 X \cdot \log(t)\} \tag{3.3.1}$$

ここで、X は薬物投与群の場合は 1、対照群の場合は 0 を取る変数とする。このとき、(2.1.2)式と同様に、薬物投与群（$X = 1$）と対照群（$X = 0$）で比を取ることでベースラインハザード $h_0(t)$ が相殺されるため、ハザード比は次のように表される。

$$HR(t) = \exp\{\beta_1 + \beta_2 \log(t)\}$$

$\beta_2 > 0$ の場合は時間とともに治療効果が増大し、逆に、$\beta_2 < 0$ の場合は時間とともに効果が減弱することを意味する。

本モデルでも、Cox 比例ハザードモデルと同様に、部分尤度を用いてβ_1 、 β_2 を推定する。イベントが発生した個体 i に対して、部分尤度に寄与する項は次のように表される。

$$\frac{\exp\{\beta_1 X_i + \beta_2 X_i \log(t_i)\}}{\sum_{j\in R(t_i)} \exp\{\beta_1 X_j + \beta_2 X_j \log(t_i)\}} \tag{3.3.2}$$

全てのイベントについて積を取ることで、Log(time)モデルに対応する部分尤度は次式のようになる。

$$L(\beta_1, \beta_2) = \prod_{i:\delta_i=0} \frac{\exp\{\beta_1 X_i + \beta_2 X_i \log(t_i)\}}{\sum_{j\in R(t_i)} \exp\{\beta_1 X_j + \beta_2 X_j \log(t_i)\}} \tag{3.3.3}$$

この対数部分尤度を最大化することにより、パラメータ β_1 、 β_2 の推定を行う。

Log(time)モデルでは、β_2 の導入によりハザード比が $\log(t)$ に応じて変化する点に特徴がある。つまり、時間依存性を明示的に取り入れることで、比例ハザード性が成立しないデータに柔軟に対応できる。

IML プロシジャでの実装コードを以下に示す。なお、3.2 節と同様に、データセット non_ph_data を使用した。

```
proc iml;
```

```
/* #3.2.1. データの読み込み */

/* #3.2.2. 部分尤度の構築 */
start negLogLik(beta) global(trt, time, cnsr, n);
   totalLogLik = 0;
   beta1 = beta[1];
   beta2 = beta[2];
   do i = 1 to n;
      if cnsr[i] = 0 then do;
         S0 = 0;
         do j = i to n;
            S0 = S0 + exp(beta1*trt[j] + beta2*trt[j]*log(time[i]));
         end;
         totalLogLik = totalLogLik + beta1*trt[i] + beta2*trt[i]*log(time[i]) - log(S0);
      end;
   end;
   return(-totalLogLik);
finish;

/* #3.2.3. ニュートン・ラフソン法による最適化 */
beta0 = {0, 0};
optn = {0, 2};
call nlpnra(rc, betaHat, "negLogLik", beta0, optn);
print betaHat;
quit;
```

実行結果として、$\hat{\beta}_1 = 0.040$、$\hat{\beta}_2 = -0.146$ が得られた。この結果から、ハザード比は時間の経過とともに指数関数的に低下し、初期には治療効果がほとんど見られないが、時間が進むにつれて効果が強まることが示唆された。グラフを図示すると以下のようになる。

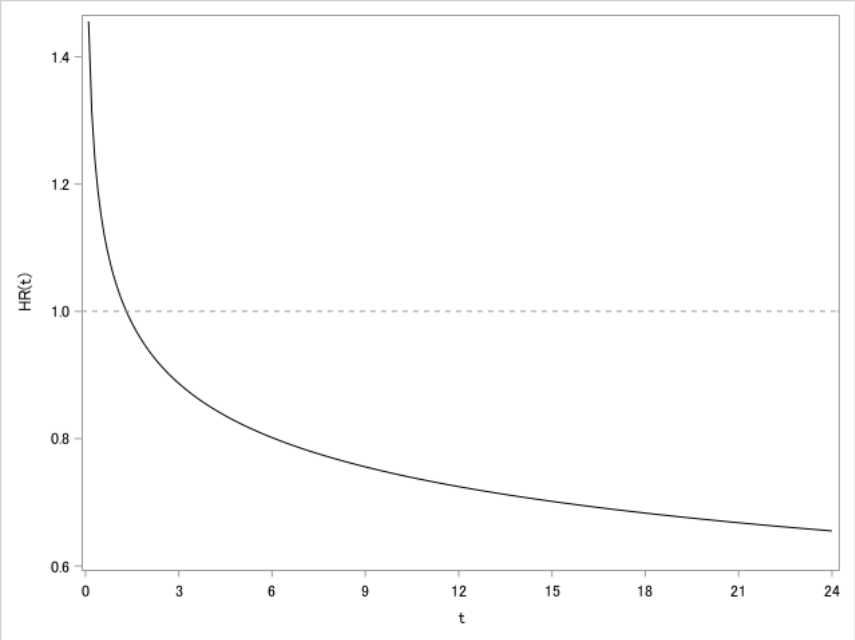


図 3.1. 時間依存ハザード比

データセット non_ph_data では、薬物投与群と対照群のハザード比は 6 か月時点までは 1、それ以降は 0.5 となるように設定していたが、グラフでは、時間経過に伴いハザード比が単調減少しており、その構造を適切に推定できているとはいえない。これは、Log(time)モデルはハザード比が時間に応じて連続的に変化するという滑らかな構造を仮定しているため、実際には段差的に変化している効果をなだらかに近似してしまうからである。したがって、真のデータ構造を反映したいなら Piecewise モデルのほうが適しているが、実際の臨床試験データでは、どの時点で効果が変化するかが事前に分からない場合も多い。このような場合、Log(time)モデルは時間に応じた効果の変化を柔軟にモデリングできるという点で有用といえる。Log(time)モデルの結果は治療効果が遅れて発現するという傾向を捉えることはできており、臨床的解釈に資する近似としては妥当であるといえる。

同様の推定を PHREG プロシジャで実装する場合、コードは以下のとおりになる。

```
proc phreg data=non_ph_data;
   class trt(ref="0");
   model time*cnsr(1) = trt trt*log_time;
   log_time = log(time);
   hazardratio trt / diff=ref;
run;
```

実行結果として、IML プロシジャによる結果と同じ値が得られた。これにより、Log(time)モデルによる推定過程を IML プロシジャによって正しく再現できていることが確認できた。

4. おわりに

本稿では、Cox 比例ハザードモデルによるハザード比の推定を IML プロシジャで再現することにより、PHREG プロシジャの背後にある数理的枠組みを具体的に示した。部分尤度の構築やニュートン・ラフソン法

による最適化の手順を実装することで、Cox 回帰による推定がどのように行われているかを可視化できた点に意義がある。また、比例ハザード性が成立しない場合の一例として、時間依存性共変量を含むモデルを IML で構築し、時間とともに変化する治療効果を捉えるアプローチを提示した。IML プロシジャによる再現は、単なる計算手順の模倣にとどまらず、モデルの数理的理解を深め、応用的な発展を検討する基盤となり得ると考える。

今後は、より複雑な共変量構造や他の非比例ハザードモデルへの拡張などを通じて、IML プロシジャを活用した教育的及び実務的貢献を検討したい。

参考文献

[1] 大橋靖雄, 浜田知久馬, 魚住龍史. 生存時間解析 第 2 版 SAS による生物統計. 東京, 東京大学出版会, 2022, 301p.

[2] SAS Institute Inc. “SAS/IML 15.2 User’s Guide”. SAS Documentation. 2025. https://documentation.sas.com/api/docsets/imlug/15.2/content/imlug.pdf, (2025/08/31 参照).

[3] SAS Institute Inc. ”SAS/STAT User’s Guide The PHREG Procedure”. SAS Documentation. 2025. https://documentation.sas.com/api/docsets/statug/v_037/content/phreg.pdf, (2025/08/31 参照).

付録

A.1

```
data ph_data(keep = trt id time cnsr);
   call streaminit(123);
   do trt = 0 to 1;
      do i = 1 to 400;
         id + 1;
         select(trt);
            when(0) lambda = 0.2;
            when(1) lambda = 0.1;
         end;
         u = rand("Uniform");
         _time = -log(u) / lambda;
         _cnsr = rand("Exponential", 15);
         time = min(_time, _cnsr);
         cnsr = (_cnsr <= _time);
         output;
      end;
   end;
run;
```

A.2

```
data non_ph_data(keep = trt id time cnsr);
   call streaminit(123);
   do trt = 0 to 1;
      do i = 1 to 400;
         id + 1;
         select(trt);
            when(0) do; lambda1 = 0.2; lambda2 = 0.2; end;
            when(1) do; lambda1 = 0.2; lambda2 = 0.1; end;
         end;
         u = rand("Uniform");
         z = -log(u);
         cumHazard = lambda1 * 6;
         if z <= cumHazard then do;
            _time = z / lambda1;
         end;
         else do;
            _time = 6 + (z - cumHazard) / lambda2;
         end;
         _cnsr = rand("Exponential", 15);
         time = min(_time, _cnsr);
         cnsr = (_cnsr <= _time);
         output;
      end;
   end;
run;
```

超多時点のラテン方格 QT 試験データに関する統計解析の新たな方向性 ―ICH-E14/S7B Q&A（2022）への対応―

高橋 行雄
BioStat 研究所（株）

New Directions in Statistical Analysis of Ultra-Multiple Timepoint Latin Square QT Study Data -Response to ICH-E14/S7B Q&A (2022)-

Yukio Takahashi
BioStat Research Co.,Ltd.

要旨： 非げっ歯類を対象にしたラテン方格 *in vitro* QT 試験では，テレメータにより投与前 2 時間から投与後 24 時間まで連続的に収集され，1 分ごとの多変数のデータが得られる．さらに，薬物濃度測定のための追加実験が行われるので，それらの時間にマッチした QTc 間隔について，対照群と各用量群間で，Dunnett の検定（多重比較）が行われてきた．2022 年に ICH-E14/S7B に関する Q&A がステップ 5 になり，Rossman ら（2023）国際共同研究グループから新たな解析法が提示された．元の 1 分ごとのデータに対し，照群と各用量群を別々にした回帰直線付き散布図（QT vs HR，QTc vs HR）が，15 分ごと用量群ごとにまとめた経時データ（HR，QT，QTc，ΔQTc）に対し線グラフが例示されている．対照群と各用量群間の比較については，適当と思われる数時間分をまとめた複数のスーバーインターバル（HR，QT，QTc）それぞれについて，投与前値を含めないラテン方格に対する分散分析が推奨されている．さらに，対照群との各群の比較には，いわゆる *LSD* 法（フィッシャーの最小有意差法）が推奨され，*p* 値は示さずに 95%信頼区間の半幅としての *LSD* を示すように求められている．

キーワード： ICH-E14/S7B，テレメトリーQT 試験，経時データ，ラテン方格，クロスオーバー法

1. はじめに

ICH（医薬品規制調和国際会議）の各種のガイドラインは，Quality：品質，Safety：安全性，Efficacy：有効性，Multidisciplinary：複合領域に分かれ，ICH-E14（QT/QTc 関連），ICH-S7B（QT 間隔延長）も別々に制定されてきた．新たにステップ 5 となった ICH-E14/S7B Q&A（2022）は，2 つの領域にまたがり，薬物が臨床的に QTc 延長を生じないことを示す「ダブルネガティブ」の非臨床シナリオの概念が導入された．

ICH-E14（2009）に関する Q&A は，改訂が繰り返され R3（2017）となっていたが，ICH-S7B（2009）に関する Q&A は，これまで制定されていなかった．新たな ICH-E14/S7B Q&A は，ICH-E14 と ICH-S7B を合わせた Q&A であり，スライド形式のトレーニング マテリアル（全 98 ページ，英文のみ），および，Examples（全 18 ページ，英文のみ）として臨床・非臨床の試験結果を統合したサマリーテーブルの作成方法について詳細に例示されている．サマリーテーブルの Table 1. Integrated Assesment には，QT Asesment pathway に引き続き，Clinical QT Study findinga，*In vitro* findings，および，*In vivo* findings，Conclusion の順に結果のサマリーが例示されている．このような微に入り細を穿つような ICH 関連の Q&A は，見たことがない．ただし，ICH-E3（1996）治験の総括報告書の構成と内容に関するガイドラインは，実に細々した例示がされており，実務を標準化する際に大変役に立った．

さて，このような臨床と前臨床を統合する ICH-E14/S7B Q&A が登場すると，前臨床分野の統計の弱さを臨床分野の統計担当者がカバーせざるを得ない事態となりつつある．前臨床の統計担当者が，歴史的に臨床統計分野に移籍されてきた時代背景も影響している．非げっ歯類を対象にした複数の QT 試験のデータマネージメント（DM），統合した統計解析に 10 年ほど前に関与した経験に基づき，新たに制定された ICH-E14/S7B Q&A の要求を満たすための方法について論じたい．

ICH-E14/S7B Q&A には，*in vivo* QT 試験についての Q3.1～Q3.5 に対する回答が示され，それを補足する Traning Materiars（p61-p79）があるが，抽象的でどのような統計解析をしたらよいのか“標準的解析法”は，まったく見えてこない．そのために，ICH の関係者を巻き込んだ同時並行的な国際研究が行われ．Rossman ら（2023）により，標準とすべき統計解析法，解析結果の表示法が，提言された．以前に関与した *in vivo* QT 試験の統計解析に共通する点もあるが，より緻密な解析結果の提示方が具体的に示されている．そこで，以前に取り扱った *in vivo* QT 試験データを用いて Rossman 論文に準拠した解析法を適用し，結果を示す．

2. イヌに対する QT 試験データの事例

金納（2014）は，陽性対照薬 Moxifloxacin について代表的な試験デザイン（Williams 型ラテン方格，Peace 型漸増法，および，用量漸増法）により，イヌを用いたテレメトリーQT 試験を実施し，日本で実施されている一般的なテレメトリー試験の現状と動向について提示した．板東（2014）は，金納（2014）が実施した QT 試験データを用いて，テレメトリー試験に対する統計学的検討を行なった．このテレメトリーQT 試験データは，第 2 期医薬安全性研究会の Web サイトにアーカイブされており，新たに提言された統計解析にチャレンジするために活用する．

表 1　3 種類の実験計画（表中の 0，3，10，30 は投与量，単位は，mg/kg）

A: Williams 型ラテン方格				
動物番号	1回目	2回目	3回目	4回目
1	0	3	10	30
2	3	30	0	10
3	10	0	30	3
4	30	10	3	0

B: Peace型漸増法				
動物番号	1回目	2回目	3回目	4回目
1	0	3	10	30
2	3	0	10	30
3	3	10	0	30
4	3	10	30	0

C: 用量漸増法				
動物番号	1回目	2回目	3回目	4回目
1	0	3	10	30
2	0	3	10	30
3	0	3	10	30
4	0	3	10	30

テレメトリーを用いたイヌ QT 試験の心電図関連の波形データは，システム内で QT 間隔などが計測され，14 変数が 1 分間の平均値および標準偏差（SD）として csv 形式で出力される．時点数は，26 時間×60 分＝1,560 時点となり，4 頭×4 期の 1 実験単位で 24,960 時点となる．出力されたデータには，多くのアーチファクト，および，飛び離れ値が存在する．実験研究者から提供された “生データ” は，実験の型別 Excel ファイルに，Excel の 4 枚のシートに用量ごと 4 動物のデータが納さめられていた．行方向に 1,560 行，列方向に（時間＋14 変数×4 動物＝57 列）となっていた．このままでは，統計解析を行うことができないので，行方向に，実験の型 3×動物 4×実験時期 4（投与量 4）×時間 1,560＝74,880 行，14 変数に（試験の型，実験番号，動物番号，実験時期，投与量，時間）などの変数を列方向に付与した SAS および JMP データセットを作成した．作成に際しては，（1 動物・1 用量）単位，1,560×14 変数の Excel ファイルを 3×16＝48 個作成し，統計ソフトで一括して結合した．

さて，テレメトリーから出力されたままのデータには，少なからずアーチファクトが混在している．データしては，99999999 のような桁違いの値であれば，それらを欠測値，Excel では“空白”として置き換えた．ただし，一変数ごとにヒストグラムを描いてみれば，図 1 に示すように心拍数 HR=0 のようなデータが混入している．心拍数 HR の赤○印は，他の変数でとんでもないデータが含まれていた場合となっている．このような状況を認知したうえで，試験開始前にデータマネージメント計画（統計解析計画）に取り扱い方について示す必要がある．詳しくは，高橋（2015），坂東（2015）を参照してもらいたい．

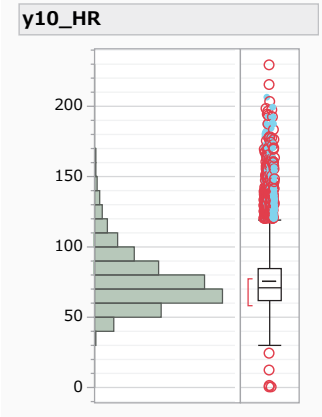


図 1 心拍数

3. 心拍数 HR に対する QT 間隔の補正の可視化

Rossman 論文の Fig. 3 および Fig. 4 では，投与量を増やした場合に，心拍数 HR と QT 間隔の散布図，HR と QTci 間隔の散布図上の回帰直線の傾きを可視化するよう求めている．図 2 に示すのは，Williams 型 4×4 ラテン方格の（0 mg/kg 群 vs. 3 mg/kg 群），（0 mg/kg 群 vs. 10 mg/kg 群），（0 mg/kg 群 vs. 30 mg/kg 群）について JMP のグラフビルダーにより散布図を作成し，回帰直線を重ね書きした結果である．効率よくこれらの散布図を作成するために，JMP ファイルの行属性に対する（非表示かつ除外）を付けたり，はずしたりする機能が役に立つ．

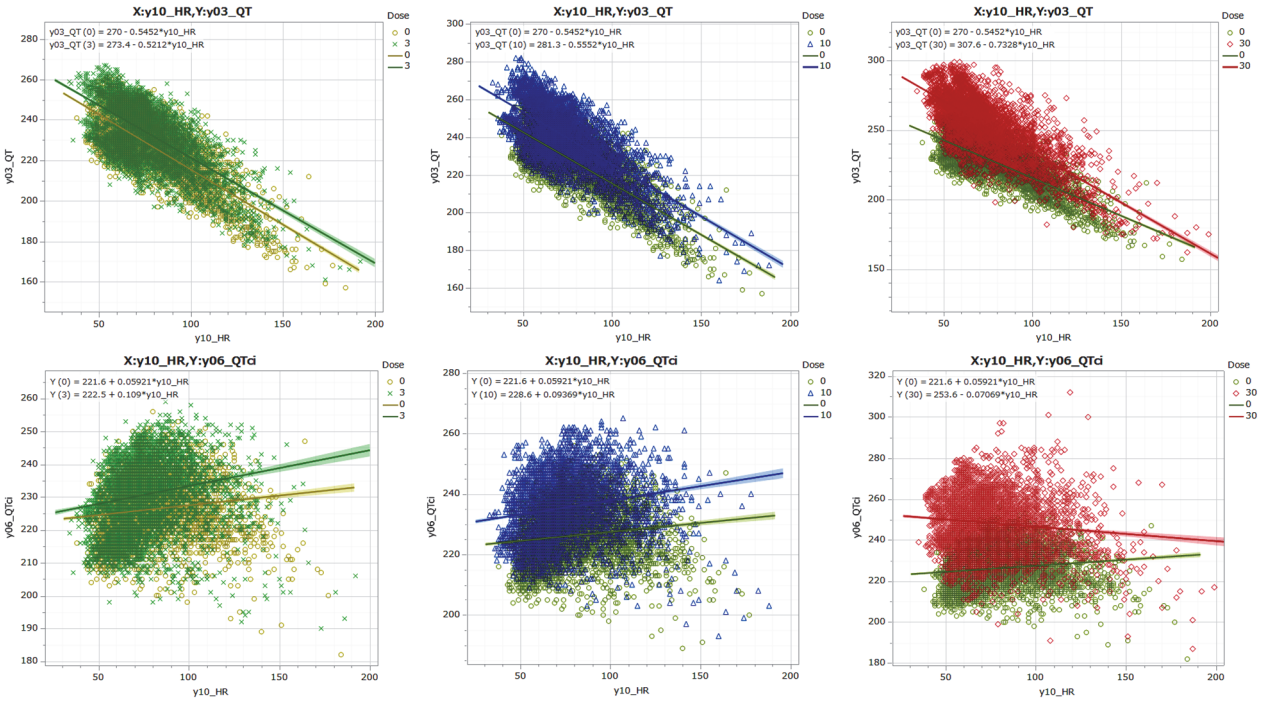


図 2 用量別に溶媒群と各投与群をマッチさせて HR と QT 間隔，HR と QTci 間隔

4. 超多時点の経時データの縮約

テレメトリーQT試験の 1 分刻みの QTci などの経時データには，小刻みなノイズが混入しており，Rossmann 論文では，15 分ごとの平均値に要約した線グラフの例示があるが，ノイズの影響が残っている．板東（2014）は，どこまで広げれば，ノイズの影響を取り除けるかを実際に検討した結果では，1 時間まで広げなくとも 30 分とすれば，ノイズの影響が和らぐことを報告した．Rossmann 論文では，QTci 間隔のみならず心拍数 HR（RR 間隔），QT 間隔についても経時データのプロットを例示している．また，QTci 間隔につ

いては，0 mg/kg 群との時間をマッチした⊿QTci 間隔も例示されている．図 3 に示すのは，1 分ごとのデータのアーチファクト，および，あり得ない値を持つデータを取り除き，30 分ごとの平均値を求め，さらに投与量別に 4 頭の平均値を計算し，JMP のグラフビルダーにより作図した結果である．

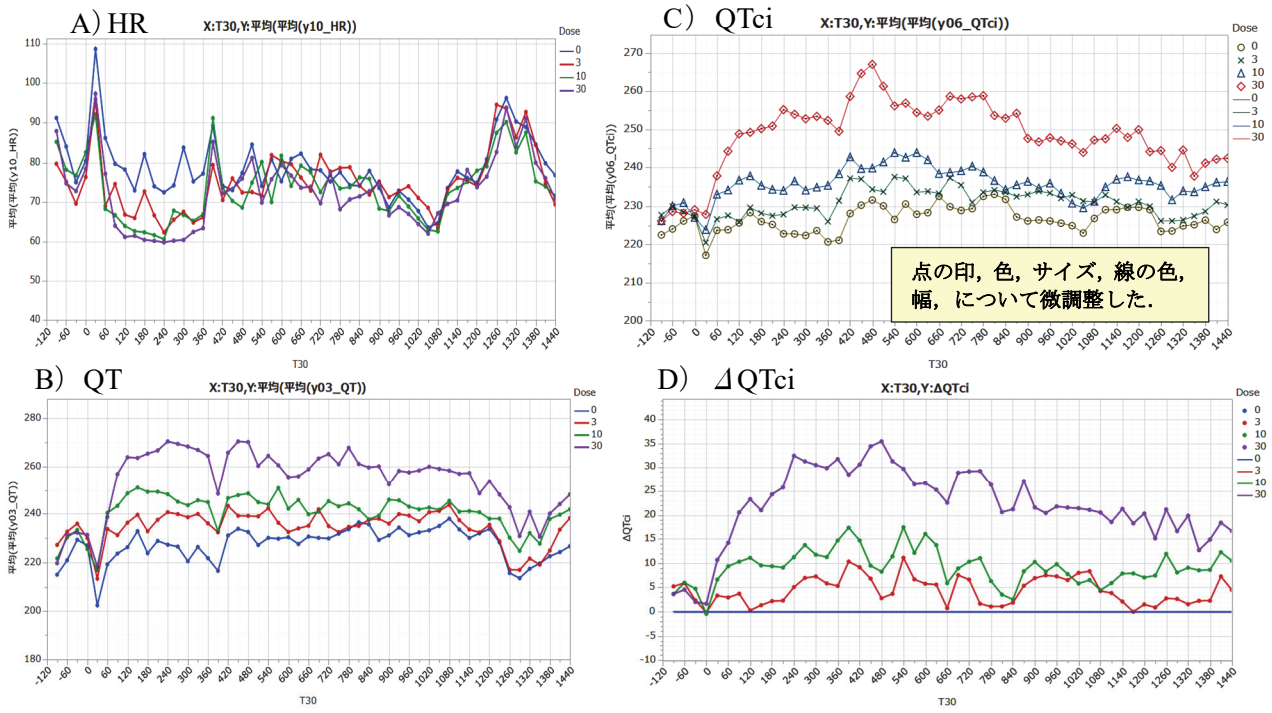


図 3　HR，QT，QTci，⊿QTci に対する 30 分の平均値に対する経時変化

図 3 を作成するためには，1分ごとのデータファイルから幾つかの手順によりデータを要約する必要がある．行方向に展開する 1 分ごと(-119 分～1440 分)の time に列方向に展開する複数の変数に対し 30 分ごとの平均値を求めたい．まず，time=-119～-90 を T30=-90，time=-89～-60 を T30=-60，. . . ，time=1381～1440 を T30=1440 のように 30 分きざみとなる変数(計算式を含む)を設定する．列方向のすべての変数を行方向に積み重ねラベル変数とし，実験の型別・Ainmal 別・Period 別・Dose 別・T30 別・ラベル変数別に平均値を計算する．計算結果ファイルに対し，T30 変数 1 行ごとに行方向のラベル変数を列方法に展開する．SAS の TRANSPOSE プロシジャ，JMP の「列の積み重ね」，「列の分解」，「転置」などを活用する．さらに，Dose 別・T30 別の平均値を計算し，図 3 の元になるデータファイルが完成する．

5. スーパーインターバルよる経時データの縮約と解析法

図 3 の 30 分ごとの経時データに対して何らかの統計解析をしたい．Animal を変量とした Mixed-effects Models for Repeated Measures(MMRM)の典型であり，適用してみたい気持ちになる．時点間の相関構造をどうするのか，交互作用をどこまで考慮するのか，投与前値の扱いをどうするのか，などについて興味が尽きない．Rossman 論文には，反復測定パラダイム(MMRM)と横断的パラダイム(各時間ビンのデータを個別に分析)の利点・欠点を示し，横断的パラダイムによる統計モデルの使用を推奨している．板東(2014)は，このイヌ QT 試験データを用いて，様々な反復測定パラダイム(MMRM)について比較検討を行っている．

Rossman 論文には，横断的パラダイム(各時間ビンのデータを個別に分析)による解析は，30 分より大きな時間間隔(例:2～4 時間)などの平均値(スーパ インターバル)を用いるとの説明がされている．各時間

ビンに対する解析モデルとして，3 種類の解析モデル，標準的な ANOVA（1. “Single delta” ANOVA），投与前値を差し引いた ANOVA（2. “Double-delta” ANOVA），前値を共変量とした（3. ANCOVA）を列挙し，標準的な ANOVA（1. “Single delta” ANOVA）を推奨している．

表2. スーパ インターバル	
S1 -2h~0h	S9 12h~14h
S9 0h~2h	**S4** 14h~18h
S2 2h~6h	S9 18h~20h
S9 6h~8h	**S5** 20h~24h
S3 8h~12h	
S9 は，除外	

どのような時間ビン（スーパ インターバル）にするのかは，試験薬物ごとに設定することが求められている．このイヌ QT 試験の解析計画書には，投与時間，給餌時間，消灯・点灯時間を考慮した表 2 に示すスーパ インターバルの範囲が設定されていた．このように設定されているスーパ インターバルを 30 分ごとの平均値のファイルから複数の変数に対し，一気に求めたい．これができるのは，SAS と JMP に共通するデータハンドリング機能の賜物である．簡潔に示すと，30 分刻みで 26 時間分の 52 行のスーパ インターバル設定ファイルを作成し，元の 30 分ファイルにマージし，1 分データから 30 分間隔ファイルを作成したと同様の操作を行い．整数化した結果を表 3 に示す．

表 3． 整数化した変数別の 5 つのスーパ インターバル（末尾の付表から抜粋）

	Design	Tr_No	Animal	Period	Dose	Dose_G	ラベル	S1_-2h~0h	S2_2h~6h	S3_8h~12h	S4_14h~18h	S5_20h~24h	S9_0h~2h	S9_6h~8h	S9_12h~14h	S9_18h~20h
1	A	1	1	1	0	1	1_RR平均	770	892	781	918	739	664	851	848	892
2	A	2	1	2	3	2	1_RR平均	839	937	754	880	794	851	873	766	943
:							:									
17	A	1	1	1	0	1	2_HR平均	85	75	84	74	87	98	77	78	75
18	A	2	1	2	3	2	2_HR平均	79	72	86	76	82	79	74	85	71
:							:									
33	A	1	1	1	0	1	3_QT平均	233	235	234	239	223	219	235	243	244
34	A	2	1	2	3	2	3_QT平均	241	245	240	245	236	240	247	238	247
:							:									
49	A	1	1	1	0	1	4_QTci平均	235	231	236	234	227	228	233	241	239
50	A	2	1	2	3	2	4_QTci平均	239	238	244	241	237	238	244	240	240
51	A	3	1	3	10	3	4_QTci平均	238	244	251	238	241	240	248	245	243
52	A	4	1	4	30	4	4_QTci平均	227	258	257	247	243	248	261	255	247
:							:									
64	A	16	4	4	0	1	4_QTci平均	226	223	229	224	226	222	229	227	227

Rossman 論文では，HR，QT，QTci の 3 変数に対し，設定された複数のスーパ インターバルに標準的なラテン方格に対する ANOVA を適用し，そのサマリーとしての図表を提示している．そこで，表 3 のデータを用いて，忠実に再現する．まず，QTci について取り上げる．前値に対する「S1_-2h~0h」を除き，S2 から S5 までのそれぞれに対し，Dose ごとの 4×4 個のデータの単純平均 Fitted means と標準誤差 *s.e.m.* を表 4 に示すようにコンパクトに示す．このような，元データから，電卓でも追試できるような単純な統計量を示すように求めていることに敬意を払いたい．

表4. 各スーパー インターバルに対するDose別 QTci の平均と標準誤差

	Moxifloxacin			
	QTci Interval(msec) : Fitted means(s.e.m.)			
	0 mg/kg	3 mg/kg	10 mg/kg	30 mg/kg
S2_2h~6h	224.00 (3.14)	228.50 (4.33)	235.25 (3.82)	252.25 (2.46)
S3_8h~12h	229.25 (2.50)	235.50 (5.32)	241.25 (4.61)	256.75 (2.06)
S4_14h~18h	226.00 (3.63)	232.25 (4.97)	233.50 (3.59)	247.50 (2.33)
S5_20h~24h	225.25 (2.30)	228.50 (5.30)	234.75 (4.48)	242.00 (0.41)

スーパ インターバルに対する 4×4 ラテン方格 16 個のデータに対し “Single delta” ANOVA

$$Sx_xh\sim xh = Animal + Period + Dose + error$$

を適用した結果を表 5 および図 5 のようにまとめる．表 5 は，各スーパ インターバルについて ANOVA を適用し，対照群（0 mg/kg）と各用量間の平均値（最小 2 乗平均）の差について ANOVA で求めた残差の

平均平方の平方根(*RMSE*)を用いた *LSD* が示されている. 平均値の差の下側に *RMSE* を用いた 95%信頼区間が示されている. 各スーパ インターバルの 95%信頼区間の差の半幅が *LSD* となっていることが確認できる. 図 4 は, 表 5 の結果をコンパクトにまとめたものである. 多くの論文で目にする縦棒グラフに *SE* を重ね書きする様式ではなく, 平均値の差を点で示し, *LSD* を用いた 95%信頼区間の表示に親しみを感ずる. なお, 図の点の様式と色の設定は, Rossman 論文に準拠している. もちろん表の形式も Rossman 論文に準拠し, 表中に *p* 値を示さずに, 有意な差があることを明示するために網掛している.

表5. 各スーパ インターバルごとの QTci の平均の差および95%信頼区間

		Moxifloxacin		
QTci Interval(msec) :	Treatment difference from 0 mg/kg (95% C.L.)			
	3 mg/kg	10 mg/kg	30 mg/kg	SI-LSD(msec)
S2_2h~6h	4.50	11.25	28.25	
	(-3.16, 12.16)	(3.59, 18.91)	(20.59, 35.91)	7.66
S3_8h~12h	6.25	12.00	27.50	
	(-2.59, 15.43)	(2.82, 21.18)	(18.32, 36.68)	9.18
S4_14h~18h	6.25	7.50	21.50	
	(-2.07, 14.57)	(-0.82, 15.82)	(13.18, 29.82)	8.32
S5_20h~24h	3.25	9.50	16.75	
	(-4.69, 11.19)	(1.56, 17.44)	(8.81, 24.69)	7.95
			Median Study LSD:	8.135
			Median Study RMSE:	4.700

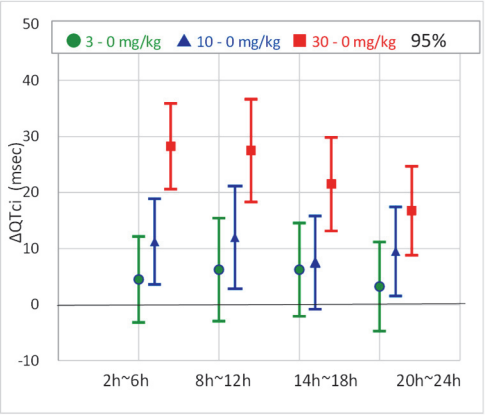


図 4. QTci:対照群との差の 95%信頼区間

6. 統計ソフトによる解析結果の要約

表 5 および図 4 に示した図表を統計ソフトを用いて直接作成する努力を惜しんではならないが, なかなか難儀であり, できるだけ類似の形式の図表を心拍数 HR に対し, 作成を試みる. 表 3 の JMP ファイルから HR を取り出し, JMP の「表の作成」を用い, 各スーパ インターバルに対する平均と標準誤差(*s.e.m.*)を計算した結果を表 6 に示す. 表 4 のように () の中に *s.e.m.* を表示することはできないが, ほぼ同等の結果を得ることができる. なお, Rossman 論文には, QTci だけでなく, HR および未補正の QT についても(表 4, 表 5, 図 4)と同様な図表が提示されている.

表 6. HR に対する平均と標準誤差

		Dose			
		0	3	10	30
S2_2h~6h	平均	76.50	66.50	64.00	61.25
	標準誤差	7.17	5.07	5.55	4.70
S3_8h~12h	平均	79.25	77.25	76.00	74.75
	標準誤差	3.82	3.35	3.42	2.29
S4_14h~18h	平均	70.00	70.75	67.75	67.75
	標準誤差	5.28	2.87	4.82	3.22
S5_20h~24h	平均	86.00	84.50	80.75	81.75
	標準誤差	7.56	6.03	7.22	8.12

ANOVA の分散分析表の表示は, Rossman 論文にないが, 表 5 の下部に, 「Median Study *RMSE*」がさりげなく示されている. *RMSE* (*Root Mean Square Error*)は, 一般的な ANOVA の出力にはないが, 分散分析表の残差(誤差)の平均平方の平方根である. JMP の「モデルのあてはめ」を用い, 「Y」に 4 変数(S2, S3, S4, S5)を, 「モデル効果の構成」に(Animal, Period, Dose)にセットする. 結果の「分散分析表」に対し「連結し

表7. HR に対するの誤差の平均平方の平方根 *RMSE* の計算

	要因	自由度	平方和	平均平方	*F* 値	*p* 値	*RMSE*
S2_2h~6h	モデル	9	2035.06	226.12	24.28	0.0005	
	誤差	6	55.88	**9.31**			**3.05**
	全体	15	2090.94				
S3_8h~12h	モデル	9	474.56	52.73	3.86	0.0570	
	誤差	6	81.88	**13.65**			**3.69**
	全体	15	556.44				
S4_14h~18h	モデル	9	781.06	86.79	6.21	0.0188	
	誤差	6	83.88	**13.98**			**3.74**
	全体	15	864.94				
S5_20h~24h	モデル	9	2490.00	276.67	13.95	0.0023	
	誤差	6	119.00	**19.83**			**4.45**
	全体	15	2609.00				
						Median Study *RMSE*:	**3.716**

たテーブルの作成」により 4 変数の分散分析表の内容が JMP ファイルに出力される．これを Excel に取り込み表 7 に示すように *RMSE* を Sqrt () 関数で計算し，その中央値 3.716 msec を求めている．

HR も表 5 の QTci と同様な表を作成したい．JMP の「分散分析表」に引き続き「効果の詳細」，「Dose」，「最小 2 乗平均差の Student の t 検定」，「差の順位レポート」によって表 8 に示す結果が得られる．

表 8.　HR の S2_2h~6h に対する平均値の差の 95%信頼区間

応答 S2_2h~6h

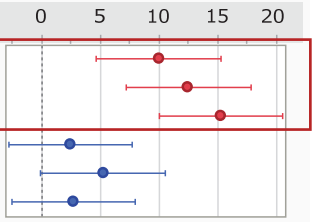
効果の詳細

Dose

最小2乗平均差のStudentのt検定

α= 0.050　t= 2.44691

水準	- 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
0	3	10.00	2.16	4.72	15.28	0.0036*
0	10	12.50	2.16	7.22	17.78	0.0012*
0	30	15.25	2.16	9.97	20.53	0.0004*
3	10	2.50	2.16	-2.78	7.78	0.2907
3	30	5.25	2.16	-0.03	10.53	0.0510
10	30	2.75	2.16	-2.53	8.03	0.2496



JMP の結果は，Dose 群の全ての水準の組み合わせに対し，水準間の差が出力されているので，「連結したテーブルの作成」で作成した JMP ファイルを Excel に取り込み，編集が必要となる．JMP では，先の水準から後の水準の差なので，対照群 0 mg/kg の平均を引く必要がある．表 9 に示すように，水準 1 と水準 2 が入れ替わり，符号がマイナスになっている．したがって，取り込んだ Excel に対し，適切な水準の組み合わせとなるような後処理が必要となる．表 9 の最後の列の *LSD* は，(上限 95%-下限 95%)/2 によって半幅を計算している．そのため，同じスーパ インターバル内では，完全に一致している．

表9. HRに対する対照群との推定平均の差, 95%信頼区間, および, *LSD*

Y	Effect	水準1	水準2	差	標準誤差	下側95%	上側95%	*p* 値	*LSD*
S2_2h~6h	Dose	3	0	-10.00	2.16	-15.28	-4.72	0.0036	5.28
	Dose	10	0	-12.50	2.16	-17.78	-7.22	0.0012	5.28
	Dose	30	0	-15.25	2.16	-20.53	-9.97	0.0004	5.28
S3_8h~12h	Dose	3	0	-2.00	2.61	-8.39	4.39	0.4729	6.39
	Dose	10	0	-3.25	2.61	-9.64	3.14	0.2598	6.39
	Dose	30	0	-4.50	2.61	-10.89	1.89	0.1357	6.39
S4_14h~18h	Dose	3	0	0.75	2.64	-5.72	7.22	0.7862	6.47
	Dose	10	0	-2.25	2.64	-8.72	4.22	0.4274	6.47
	Dose	30	0	-2.25	2.64	-8.72	4.22	0.4274	6.47
S5_20h~24h	Dose	3	0	-1.50	3.15	-9.21	6.21	0.6507	7.71
	Dose	10	0	-5.25	3.15	-12.96	2.46	0.1465	7.71
	Dose	30	0	-4.25	3.15	-11.96	3.46	0.2258	7.71

QTci に対する図 4 は，表 9 と同じ形式の Excel の表から，Excel の散布図によって細かな設定を元に作成したものである．図 5 は，表 9 の Excel 表の右側に点および 95%信頼区間(ひげ)の X 軸上の位置を(1.00, 1.20, 1.40, . . .)のように設定して描いた結果である．点については一括して設定できるが，“ひげ” については，「データの選択」により，1 本ごとに(*x*, *y*)の位置を特定し引いたものである，マーカの形状・色は，個別に設定し，根気よく変更した結果である．ただし，他の変数で設定した図を再利用することにより，効率を上げることができる．

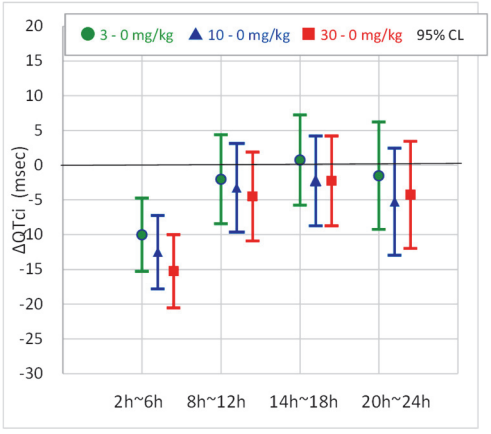


図 5. HR：対照群との差の 95%信頼区間

7. 背景データを用いた分析感度

ICH-E14/S7B Q&A の S7B に関する「質問 3.4 測定法の感度をどのように評価すべきですか。」に対する回答は，背景データを用いた分析感度の立証を求めている．

- 測定法の感度は、最小検出感度（*MDD*, *Minimum Detectable Differences*）の定義及び陽性対照を用いた試験によって説明可能である。
- 統計学的な検出力は、同一試験方法を用いた同一の検査施設における背景を用いて算出することも可能である。
- 陽性対照の背景データを用いて、測定法の感度の正当性を立証する場合又は統計学的な検出力を算出する場合には、実際の試験成績の分散が背景データの分散と一致することが必要である。
- 仮説的な例：検出可能な最小差（*MDD*）は、動物試験における薬物曝露が高い臨床曝露量のみを網羅している場合は 5 ミリ秒であるが、高い臨床曝露量のより大きな倍数が達成された場合はそれより高くなる．（例えば、3 倍高い臨床曝露量が達成された場合は 10 ミリ秒）又は、さらに大きい倍数が達成された場合は、QTc 閾値がさらに高くなる）。

S7B の「質問 3.5 *In vivo* QT 測定の薬力学及び薬物動態から得られた試験成績を（最終報告書中で）示す場合には、どのような内容が推奨されますか。」に対する PD に関する回答には，過去のデータの分散についての要求があり，また，同じ実施計画書及び統計解析計画書であることを求めている．

- 過去の陽性対照を使用する場合、現在のデータの分散は過去に認められた分散と一致していなければならない。
- 新医薬品のデータ及び過去のデータが、同じ実施計画書及び統計解析計画書（statistical analysis plan）に従って収集された旨を表明すること。

8. 背景データの分散および最小検出感度 *MDD*

現在のデータが，過去のデータと同一の分散であること，また，過去のデータから最小検出感度（*MDD*, *Minimum Detectable Differences*）を求め，十分な分析感度を持っているか強く求められている．Williams 型 4×4 ラテン方格実験に先立ち，Peace 型 4×4 クロスオーバー，および，用量漸増法による実験を背景（過去）データとみなした場合の“分散”，および，*MDD* をどのように求めるかについて示す．“分散”は，表 7 に示したラテン方格に対する分散分析表の誤差（残差）項の“平均平方”と解される．QTci については，表 5 の Median Study *RMSE* の平方となる．統計的には，“分散”が好まれるが，Rossman 論文では，分散の平方根 *RMSE* で統一されている．

表 10 から“過去データ”の Median *RMSE* は，3.55 と推定される．現在のデータから Median Study *RMSE* は，4.70 となる．この結果から，現在のデータの分散は，過去のデータの分散と等しいと言えるのだろうか，等しいと言えなくとも“ほぼ同じ”とみなせるのであろうか．

表10．過去のデータから求めた QTci の分散・*RMSE*

スーパ	A:Williams型		B:Peace型		C:用量漸増法	
インターバル	平均平方	*RMSE*	平均平方	*RMSE*	平均平方	*RMSE*
S2_2h~6h	19.58	4.43	18.68	4.32	8.86	2.98
S3_8h~12h	28.15	5.31	8.53	2.92	20.56	4.53
S4_14h~18h	23.15	4.81	13.03	3.61	32.56	5.71
S5_20h~24h	21.08	4.59	9.47	3.08	8.45	2.91
Median Study *RMSE*		4.70		3.34		3.76
	現在のデータ		過去のデータ：Median *RMSE*			3.55

過去のデータから最小検出感度 *MDD* の求め方は，ICH - E14/S7B Q&A の本文には，示されていない．ただし，Training Materials の p71 の説明，Rossman 論文の Fig. 1.および

Fig. 2. から過去データの *RMSE* を用いて検出力曲線を描き，検出力（$1-\beta$）が 80%となる QTc chenge 量の推定値を *MDD* と定義されているが，参考文献を示すだけで計算方法は，示されていない．なお，ICH -E14/S7B Q&A の Training Materials Examples の p10.の注 9．に *MDD* の計算式が示されているが，検出力を考慮した *MDD* ではなく，検出力を求めるために必要な *LSD* の定義式となっていることも *MDD* の算出方法が不透明となる原因である．

良く知られている検出力曲線の描き方は，独立 2 標本平均の検出力であるが，ラテン方格実験では，対応のある 2 標本平均の検定であり，SAS の POWER プロシジャでも対応していない課題である．幸い，4×4 のラテン方格の場合は，誤差の自由度が $n=4$ の独立 2 標本平均の検出力の場合と同じになるので，幸いなことに代用できる．

JMP の DOE の「標本サイズエクスポローラ・検出力・独立二標本平均の検出力」を用い，「$\alpha=0.05$，両側，母標準偏差＝未知，全体の標準サイズ＝8，1 群と 2 群の標準偏差＝3.55（Median *RMSE*），検出力＝80%」を設定すると *MDD*＝8.4517 を得る．ICH -E14/S7B Q&A では，*MDD* が，10 msec 以下であることが求められていので，“過去の実験”は，十分な検出感度を持つと言えるので，同様な“現在”の実験も十分な検出感度を持つことが期待できる．しかし，現在の実験の標準偏差 *RMSE*=4.70 から *MDD* を求めると 11.19 msec となり，10 msec 以上となってしまう．図 4 の結果も参考にしつつ総合的な判断が求められる．

9. 考察

イヌに対する 3 組の *in vivo* QT 試験から得られた知見を元に，小松（2016），高橋（2016）にサルを対象とした 6 社 7 試験の共同研究についの解析結果を報告した．その結果，4×4 の Williams 型ラテン方格による *in vivo* QT 試験データは，十分な検出感度を持つ実験方法であることを確認したのであるが，標準的な解析法についての提言をするまでに至らなかった．ICH-E14/S7B Q&A（2022）に準拠した Rossman ら（2023）は，*in vivo* QT 試験データについて“標準的な解析法”が提言されたので，これまで得た実データを元に追試することにした．ここでは，その一部のデータを用いて，Rossmanら（2023）の提言に忠実に解析を試みた結果を示した．

共同研究の一環として，6 社に対し 2 試験のスーパ インターバル（投与前，投与後）のデータを提供し，対照群と各投与群との差についての統計解析を行うよう依頼した．高橋（2016）に示したように結果は，誤用の見本市ともいえる状況であった．最大の誤用は，4×4 のラテン方格の陽性対照薬に対する実験にもかかわらず，群間比較に対する Dunnett の多重比較を適用した場合であり，対照群と各用量群間に有意な差を検出できないと 3 社から報告された．この結果から，非げっ歯類に対するテレメトリーQT 試験の実施に関して共通プロトコールによる実験が行われ，検出感度も良好であったのであるが，統計解析に関しては，無法地帯ともいうべき状況に愕然とした．是正しなければと思ったが，研究グループの関心事は，PK-PD 解析であったので，なんら対応できずに現在に至っている．

ICH-E14/S7B Q&A（2022）本文，Training Materials，Training Materials Exsamples どこにも統計解析における誤用を防止するような記述は，見当たらない．Rossman ら（2023）は，テレメトリーQT 試験の実施計画書の作成，テレメトリー信号の品質評価に関する注意事項，推奨する統計解析の方法，陰性対照，および，陽性対照薬に対する結果の表示方法に至るまで，事細かに示されていて，“標準作業手順書”のごとくである．高橋（2016）も同様の提言をしていたのであるが，QTc に対する解析のみならず，HRおよび補正前のQTも含めた総合的な解析が具体的に提示されていることに感銘を受けた．ただし，最小検出感度（*MDD*）の算出法については抽象的であり，具体性に欠けるので，今後の検討課題である．

文献

Rossman E.I., Wisialowski T.A., VargasH.M.，et al.(2023),Best practice considerations for nonclinical in vivo cardiovascular telemetry studies in non-rodent species: Delivering high quality QTc data to support ICH E14/S7B Q&As, Journal of Pharmacological and Toxicological Methods 123 (2023) 107270.

ICH-E14（2009），非抗不整脈薬における QT/QTc 間隔の延長と催不整脈作用の潜在的可能性に関する臨床的評価，https://www.pmda.go.jp/files/000156160.pdf

ICH-E14 Q&As R3（2017），「非抗不整脈薬における QT/QTc 間隔の延長と催不整脈作用の潜在的可能性に関する臨床的評価」に関する Q&A（R3），https://www.pmda.go.jp/files/000218172.pdf

ICH-S7B（2009），ヒト用医薬品の心室再分極遅延（QT 間隔延長）の潜在的可能性に関する非臨床的評価，https://www.pmda.go.jp/files/000156281.pdf

ICH-E14/S7B Q&A（2022），「QT/QTc 間隔の延長と催不整脈作用の潜在的可能性に関する臨床的及び非臨床的評価」に関する質疑応答集，https://www.pmda.go.jp/files/000218172.pdf

Traning Materiars，E14/S7B Questions & Answers: Clinical and Nonclinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Petential Initial Training Material，https://www.pmda.go.jp/files/000246889.pdf

Examples，E14/S7B Questions & Answers: Clinical and Nonclinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Petential Initial Training Material - Example Supplemental File，https://www.pmda.go.jp/files/000269816.pdf

ICH-E3（1996），治験の総括報告書の構成と内容に関するガイドライン，https://www.pmda.go.jp/files/000156923.pdf

金納明宏（2014），イヌを用いたテレメトリーQT/QTc 試験の実際，https://biostat.jp/archive_teireikai_2_download.php?id=57

板東正博（2014），超多時点経時データの統計解析，https://biostat.jp/archive_teireikai_2_download.php?id=58

板東正博（2015），SAS を用いたデータクリーニング手順の提案，https://biostat.jp/archive_teireikai_2_download.php?id=74

高橋行雄（2015），JMP を用いた統計的データクリーニング法，https://biostat.jp/archive_teireikai_2_download.php?id=75

小松竜一（2016），実データに基づいたサル QT 延長試験結果の臨床との比較，https://biostat.jp/archive_teireikai_2_download.php?id=138

高橋行雄（2016），実データを用いたサルの QT 延長試験における 4×4 のラテン方格実験の妥当性検証，https://biostat.jp/archive_teireikai_2_download.php?id=137

付表：イヌ QT 試験 Williams 型に対するスーパーインターバル（HR，QTci）

	Design	Tr_No	Animal	Period	Dose	Dose_G	ラベル	S1_-2h~0h	S2_2h~6h	S3_8h~12h	S4_14h~18h	S5_20h~24h	S9_0h~2h	S9_6h~8h	S9_12h~14h	S9_18h~20h
1	A	1	1	1	0	1	1_RR平均	770	892	781	918	739	664	851	848	892
2	A	2	1	2	3	2	1_RR平均	839	937	754	880	794	851	873	766	943
:							:									
17	A	1	1	1	0	1	2_HR平均	85	75	84	74	87	98	77	78	75
18	A	2	1	2	3	2	2_HR平均	79	72	86	76	82	79	74	85	71
19	A	3	1	3	10	3	2_HR平均	83	62	85	73	88	69	79	83	73
20	A	4	1	4	30	4	2_HR平均	81	61	77	72	83	75	79	82	76
21	A	5	2	1	3	2	2_HR平均	64	54	72	64	72	73	70	70	69
22	A	6	2	2	30	4	2_HR平均	67	50	70	60	64	56	74	66	63
23	A	7	2	3	0	1	2_HR平均	69	59	68	60	67	64	69	68	67
24	A	8	2	4	10	3	2_HR平均	62	51	69	66	65	59	67	73	66
25	A	9	3	1	10	3	2_HR平均	79	65	73	55	73	84	82	69	71
26	A	10	3	2	0	1	2_HR平均	80	78	84	63	86	85	87	72	69
27	A	11	3	3	30	4	2_HR平均	74	61	72	65	77	78	80	65	65
28	A	12	3	4	3	2	2_HR平均	69	63	72	68	83	64	71	77	73
29	A	13	4	1	30	4	2_HR平均	91	73	80	74	103	90	74	75	87
30	A	14	4	2	10	3	2_HR平均	98	78	77	77	97	79	75	75	89
31	A	15	4	3	3	2	2_HR平均	88	77	79	75	101	89	83	77	85
32	A	16	4	4	0	1	2_HR平均	96	94	81	83	104	106	81	83	91
:							:									
33	A	1	1	1	0	1	3_QT平均	233	235	234	239	223	219	235	243	244
34	A	2	1	2	3	2	3_QT平均	241	245	240	245	236	240	247	238	247
:							:									
49	A	1	1	1	0	1	4_QTci平均	235	231	236	234	227	228	233	241	239
50	A	2	1	2	3	2	4_QTci平均	239	238	244	241	237	238	244	240	240
51	A	3	1	3	10	3	4_QTci平均	238	244	251	238	241	240	248	245	243
52	A	4	1	4	30	4	4_QTci平均	227	258	257	247	243	248	261	255	247
53	A	5	2	1	3	2	4_QTci平均	218	217	220	218	213	212	220	220	217
54	A	6	2	2	30	4	4_QTci平均	222	253	252	243	242	234	267	250	250
55	A	7	2	3	0	1	4_QTci平均	219	216	224	217	217	218	224	223	217
56	A	8	2	4	10	3	4_QTci平均	217	226	230	224	222	222	232	230	224
57	A	9	3	1	10	3	4_QTci平均	229	238	246	240	241	233	243	242	239
58	A	10	3	2	0	1	4_QTci平均	220	226	228	229	231	221	226	236	234
59	A	11	3	3	30	4	4_QTci平均	229	246	256	246	242	236	248	259	245
60	A	12	3	4	3	2	4_QTci平均	229	230	238	236	232	228	237	235	236
61	A	13	4	1	30	4	4_QTci平均	233	252	262	254	241	240	265	260	254
62	A	14	4	2	10	3	4_QTci平均	230	233	238	232	235	232	237	233	240
63	A	15	4	3	3	2	4_QTci平均	228	229	240	234	232	223	239	236	233
64	A	16	4	4	0	1	4_QTci平均	226	223	229	224	226	222	229	227	227

Wilcoxon-Mann-Whitney（WMW）検定の問題点とBrunner Munzel検定への展望：SASを用いたBrunner Munzel検定の実装

○橋本 隆正、舟尾 暢男、水間 浩太郎

（武田薬品工業株式会社 日本開発センター 生物統計室）

Issues of the Wilcoxon-Mann-Whitney (WMW) Test and Insights for the Brunner-Munzel Test: Implementation of the Brunner-Munzel Test Using SAS

Takamasa Hashimoto, Nobuo Funao, Kotaro Mizuma
Statistical & Quantitative Sciences, Takeda Pharmaceutical Company Limited

要旨

WMW 検定と Brunner Munzel 検定の概要と理論的な考察を行い、Brunner Munzel 検定を SAS で実施する方法を紹介した後、シミュレーションにより両手法を比較する。

キーワード: Wilcoxon-Mann-Whitney（WMW）検定、Brunner Munzel 検定、t 検定、Welch 検定、Behrens-Fisher 問題

1. はじめに

投与群が 2 群ある臨床試験を考える。群 1（被験薬群）の応答が群 2（対照薬群）の応答よりも大きい（又は小さい）かどうかに興味を持つことが多く、この場合「2 群の母集団の位置パラメータは異なる」という対立仮説に対して「2 群の母集団の位置パラメータは等しい」という帰無仮説をおき、本帰無仮説に関する検定を実施・棄却されたことをもって群間の応答が異なる、すなわち群 1 の群 2 に対する効果を示すことが一般的な流れとなる。この設定の下、本稿は Wilcoxon-Mann-Whitney（WMW）検定に対して非臨床や臨床試験の現場で誤認されている以下の誤った考え方を正すことを目的としている。

- 2 群の中央値に差があることを示すために WMW 検定を適用する
- 2 群の平均値に差があることを示すために t 検定を適用したいが、データが正規分布に従っているかどうか自信が無い、又はデータ数が少ないので WMW 検定を適用する
- WMW 検定はノンパラメトリック検定であるので、データがどんな形であっても（ t 検定よりも少し検出力は落ちるが）誤った結果を導かない、頑健な検定である
- 2 群の母集団は正規分布に従っているが（2 群の等分散性の検定結果等で判断し）2 群の母集団の分散が異なることが示唆されるため t 検定の代わりに WMW 検定を適用する

まず、2 群の母集団が正規分布に従っていると仮定する場合、「2 群の母集団の平均値は等しい」という帰無仮説に関する検定を 2 標本 t 検定を用いて実施することが多い。両群に対して正規母集団を仮定したとしても、2 群の母集団の分散が異なる場合はこの検定では厳密な有意水準の下で検定を行うことはできず、Behrens-Fisher 問題として知られているが、「2 群の母集団の分散は等しい」と仮定するものが通常の t 検

定、この仮定をおかない（ただし Welch-Satterthwaite の近似式を用いた）手法が Welch 検定となる。これらの検定手法は正規母集団を仮定（パラメトリックモデルを前提と）していることから、パラメトリック検定とよばれる。

次に、WMW 検定について考える。WMW 検定では 2 群の母集団に対して特定の確率分布を仮定しない（ノンパラメトリックモデルを前提とする）ことから、ノンパラメトリック検定とよばれる。WMW 検定における帰無仮説・対立仮説については後述するが、t 検定や Welch 検定とは通常全く異なる仮説と関連しているにも拘らず、WMW 検定を、t 検定や Welch 検定の代用として適用されることが少なくない。WMW 検定は一般的には「2 群の母集団の平均値は等しい」という帰無仮説に関する検定ではないことから、以下の様な状況において、2 群の平均値に差があることを示す目的で t 検定の代わりに WMW 検定を適用することは適切でない。

- 2 群の母集団が正規分布に従っていないことが示唆されるため t 検定の代わりに WMW 検定を適用する
- 2 群の母集団は正規分布に従っているが例数が小さい又は分散が異なることが示唆されるため t 検定の代わりに WMW 検定を適用する

また、WMW 検定を適用する際、以下の様な前提をおく、あるいは思い込むことがあるが、場合によっては問題が生じる。

1. 2 群の母集団が連続分布かつ同じ形状であることのみ仮定し、「2 群の母集団の位置パラメータは異なる」という対立仮説に対して「2 群の母集団の位置パラメータは等しい」という帰無仮説を設定・実施する
2. 2 群の母集団に（連続分布であることは仮定するが）特定の確率分布を仮定しないことから頑健な検定手法であり、有意水準を適切に制御することができる
3. #1 より、帰無仮説が棄却された場合は「2 群の母集団の位置パラメータは異なる」こととなるが、位置パラメータは 2 群間の中央値を含む任意の分位点のずれ、また平均値のずれを表しているため、特に、群 1（被験薬群）の群 2（対照薬群）に対する効果を平均値や中央値にて考察する

「WMW 検定は、t 検定や Welch 検定よりも仮定が少ないため頑健な検定手法である」と考えてしまいがちだが、#1 の「2 群の母集団が同じ形状である」という仮定は非常に強いものとなっており、異質性（例えば Behrens-Fisher 問題である「2 群の母集団の分散が異なる」ことに対応）が存在する場合は検定結果が適切でなく、異質性に対しては t 検定や Welch 検定よりも敏感に反応する（有意水準が適切に制御できない）ことに加え、2 群の母集団の分散が等しい場合であっても、高次のモーメントが異なる状況では検定結果が適切でない場合も生じる。また、#2 は非臨床や臨床試験の現場でよく見受けられる考え方であるが、一般的には成り立たず、有意水準が制御できない場面も少なくない（Karch, 2021）。さらに、WMW 検定の結果が有意であった場合、群 1（被験薬群）の群 2（対照薬群）に対する効果があると解釈できるが、一方で「WMW 検定が注目しているものは平均値や中央値の群間比較である」と思われることがあるものの、上で記載したように、これはかなり特殊なモデルが正しい場合であって、実際の解析において考えていた帰無仮説が棄却されたとしても、#3 での解釈は難しいものとなることがある。なお、WMW 検定の欠点を解消するために Brunner Munzel 検定が提案されており、ノンパラメトリックな Behrens-Fisher 問題に対して適切な手法となっているが、SAS のプロシジャでは用意されていない。

本稿では、WMW 検定及び Brunner Munzel 検定の概要と、Brunner Munzel 検定を SAS で検定を実施する方法を紹介した後、いくつかの場面における「t 検定、Welch 検定、WMW 検定、Brunner Munzel 検定」の 4 手法の比較・考察を試みる。

2. WMW 検定

群 1 のデータ $y_{\mathrm{F}1},\ldots,y_{Fn_1}$ 及び群 2 のデータ $y_{\mathrm{G}1},\ldots,y_{Gn_2}$とした際、$\mathrm{W}=\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}I\left(y_{Fi},y_{Gj}\right)$を、Mann-Whitney 型の WMW 検定と呼ぶ。ただし、$I(y_F,y_G)$は、$y_F>y_G$の時 1 となり、それ以外の時は 0 となる指示関数とする。WMW 検定は、臨床試験や非臨床試験でしばしば適用されているが、多くの、特に理論的なアプローチを重視した成書においては、ロケーションシフト問題の枠組みで位置母数の検定（$G(y)=F(y+\Delta);\ \Delta=0\ \mathrm{v.s.}\ \Delta\neq 0$. なお、$F$ 及び G はそれぞれ群 1 及び群 2 における確率分布関数とする）としての観点から詳細に書かれているためか（Hájek and Šidák（1967）、Cox and Hinkley（1974）、柳川（1982） 等）、WMW 検定は中央値や平均値の違いを見出すものであると理解されていることがある（Hart（2001）等）。一方、Shirahata（1993）では、検定統計量の定義から、帰無仮説は $H_0\colon \Pr(Y_F>Y_G)=\frac{1}{2}$ である（群 1 から無作為に選んだ標本が､群2から無作為に選んだ標本より大きくなる確率は $\frac{1}{2}$ である)はずであると記載されている。また、この検定統計量が $\Pr(Y_F>Y_G)$ の自然な推定量になっていることから、Cox and Hinkley（1974）においても上述したロケーションシフト問題の枠組みに加えて、$H_0\colon \Pr(Y_F>Y_G)=\frac{1}{2}$ からの記述がみられる（なお、タイが存在する場合は、$\Pr(Y_F>Y_G)+\frac{1}{2}\Pr(Y_F=Y_G)=\frac{1}{2}$ である)。しかしながら、Fay and Proschan（2010）や Lehmann and Romano（2022）には、WMW 検定を $H_0\colon \Pr(Y_F>Y_G)=\frac{1}{2}$ に対して適用すると、第 1 種の過誤確率を制御できないことが記載されており、Shirahata（1993）においても、棄却点は $F=G$ の時のみ正しいと記載されている。Fay and Proschan（2010）は、WMW 検定が適用され得る様々な仮説間の関係を整理し、実務家がこれらの関係性を理解できるようなフレームワークを提供しており､一読するに値するが､ここでは、最小限の記述で WMW 検定についての考察を試みる。なお、話を簡単にするために、平均値や分散等の適当なモーメントが存在する確率分布のみを対象にし、同じ順位となるデータ（タイデータ）が存在しない前提で本項の議論を進めるが、タイデータが存在する場合でも同様の結論となることを申し添える。

準備

以降の議論に直接関係ないものも含め有用な命題を挙げておく。いずれも簡単な計算から得ることができる。なお、$F<_{st}G$ は、G は F より確率順序の意味で大きいということを表す（即ち、任意の y について $G(y)\leq F(y)$が成立し、ある点では厳密な不等号が成立する)。

- **Proposition 1**. $F=G\Longrightarrow \Pr(Y_F>Y_G)=\frac{1}{2}$
- **Proposition 2**. $Y_F<_{st}Y_G\Longrightarrow E(Y_F)\neq E(Y_G)$
- **Proposition 3**. $Y_F<_{st}Y_G\Longrightarrow \Pr(Y_F>Y_G)\neq\frac{1}{2}$
- **Proposition 4**. $G(y)=F(y+\Delta),\Delta\neq 0\Longrightarrow \Pr(Y_F>Y_G)\neq\frac{1}{2}$
- **Proposition 5**. 同じ点において対称な確率密度関数を持つ 2 つの確率変数に対して、$\Pr(Y_F>Y_G)=\frac{1}{2}$ が成り立つ（これにより、$F\neq G$ だが、$\Pr(Y_F>Y_G)=\frac{1}{2}$ となる例が作れる。なお、このモデルの下では平均値と中央値は一致し、Y_F と Y_G の間でもこれらは等しくなる）

WMW 検定を適用する動機

竹内（1975）には､「ノンパラメトリック検定といっても、分布や対立仮説の下での分布のずれについて全く何も想定されないということはない。むしろ対立仮説について大体の形を想定して検定方式を選ばねばならない。しかしながら、対立仮説の下での検出力について、特定の型の問題、例えば位置母数のずれについてのみ、やたらと細かい議論をしても、あまり意味がない。現実にはある程度の大きな傾向を想定して、その

前提の下で、簡単でわかりやすい検定方式を選ぶべきである。」と記載されている。この観点から、Cox and Hinkley（1974）や Shirahata（1993）においても指摘されているように、WMW 検定は $\Pr(Y_F > Y_G) = \frac{1}{2}$ からのずれを検討したい場合に適用する検定であると考えられる。

WMW 検定を適用する際の帰無仮説と対立仮説

従って、$H: \Pr(Y_F > Y_G) = \frac{1}{2}$ v.s. $K: \Pr(Y_F > Y_G) \neq \frac{1}{2}$ という問題を考える時に、WMW 検定を適用しようとすることは自然である。なお、これが 2 群間の中央値や平均値の違いの仮説問題となるためには、F と G にさらなる仮定が必要であり、現実問題でそれらを仮定することの妥当性を正当化することは中々困難である。さて、前述の通り WMW 検定を $H: \Pr(Y_F > Y_G) = \frac{1}{2}$ に対して適用すると、第 1 種の過誤確率を制御できないため、$F = G$ という条件を帰無仮説に加えないといけない。Proposition 1 より、$\{F = G\} \subset \left\{\Pr(Y_F > Y_G) = \frac{1}{2}\right\}$ であるので、結果的に $H: F = G$ となる。従って、WMW 検定を適用する場合の仮説は以下となる。

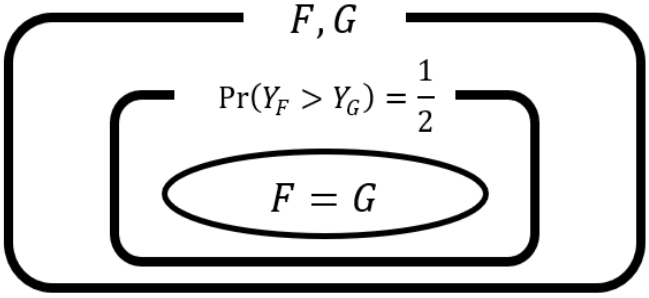


図 2.1 確率分布の仮定に関する包含関係

$$H_1: \{F, G: F = G; F \in \Psi_c\} \text{ v.s. } K_1: \left\{F, G: Pr(Y_F > Y_G) \neq \frac{1}{2}; F, G \in \Psi_c\right\}$$

ここで $Y_F \sim F, Y_G \sim G, \Psi_c$: 連続関数からなる集合である。WMW 検定はこの仮説空間の下では valid で consistent であるが、仮説空間が奇妙で実用上正当化しにくいことを以下で説明する。今、$P = H_1 \cup K_1$ とする。P は、全ての分布 F と G の集合から $\left\{F \neq G, Pr(Y_F > Y_G) = \frac{1}{2}\right\}$ を除いた空間になっている。Proposition 5 より、例えば平均値は等しいが分散が異なる正規分布などが、この除かれた空間に含まれる。この除かれた空間には真のモデルは存在しないという仮定が成り立つ、もしくは存在しても問題とならないのであればこの除かれた空間が無視できるが、前者を正当化するのは現実データを適用する際には難しく、後者は WMW 検定により棄却されることがあり問題となる。同じような理由で、

$$H_2: \{F, G: F = G; F \in \Psi_c\} \text{ v.s. } K_2: \{F, G: F <_{st} G \text{ or } G <_{st} F; F, G \in \Psi_c\}$$

も実用上正当化しにくく、また

$$H_3: \{F, G: F = G; F \in \Psi_c\} \text{ v.s. } K_3: \{F, G: G(y) = F(y + \Delta); \Delta \neq 0; F \in \Psi_c\}$$

については、前述の如く仮定している確率モデルが特殊すぎて現実的でない。そうすると、現実的な仮説問題としては、

$$H_4: \{F = G\} \text{ v.s. } K_4: \{F \neq G\}$$

が考えられるが、「$Pr(Y_F > Y_G) = \frac{1}{2}$ からのずれを見たい」という motivation に直接対応するものではなくなる。なお、WMW 検定はこの仮説問題において consistent でないことが知られているが（4.4 項にて考察する）、これは Shirahata（1993）より $\left\{F \neq G, Pr(Y_F > Y_G) = \frac{1}{2}\right\}$ が真のとき WMW 検定の検定統計量の漸近分布が、帰無仮説の下でのそれと分散のみが異なる正規分布となることが示すことができるので、直感的には明らかである。以上より、$H: \Pr(Y_F > Y_G) = \frac{1}{2}$ v.s. $K: \Pr(Y_F > Y_G) \neq \frac{1}{2}$ という問題に対して、WMW 検定を適用することは適切ではないが、Brunner and Munzel（2000）はこの問題に対して valid で consistent な検定統計量を提案した。従って、$Pr(Y_F > Y_G) = \frac{1}{2}$ からのずれを検討したい場合に適用する検定としては、以下にまとめた通り、WMW 検定ではなく Brunner -Munzel 検定を用いるべきである。

- WMW 検定の帰無仮説：$\boldsymbol{F = G}$ **（2 群の確率分布が等しい）**、対立仮説：$Pr(Y_F > Y_G) \neq \frac{1}{2}$
- Brunner -Munzel 検定の帰無仮説：$Pr(Y_F > Y_G) = \frac{1}{2}$ 、対立仮説：$Pr(Y_F > Y_G) \neq \frac{1}{2}$

3. Brunner Munzel 検定

Brunner and Munzel（2000）に倣い、以下のノンパラメトリックモデルを考える。F と G は任意の分布関数とするが、1 点分布の様な特殊なものは除く。

$$N = n_1 + n_2\,,\ Y_{1k} \sim F\ with\ \text{i.i.d.}\,,\ Y_{2k} \sim G\ with\ \text{i.i.d.}\,,\ k = 1, \dots, n_i\,,\ i = 1, 2\ \text{（群）} \tag{1}$$

「2 群の応答が等しい」というノンパラメトリックな仮説を定式化するため、以下の効果量を考える（前項と不等号の向きが逆である点に注意）。

$$p = P(Y_{11} < Y_{21}) + \frac{1}{2}P(Y_{11} = Y_{21}) \tag{2}$$

前項で述べた通り、上記に関する帰無仮説は $H\colon p = \frac{1}{2}$ で表せる。さて、N 個の Y_{ij} に順位付けを行ったものを R_{ij} とし（タイは mid-rank ）、各群の順位の平均を$\bar{R}_{i\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i} R_{ik}\ \ (i = 1,\ 2)$ とすると、p の不偏かつ一致推定量として以下が得られる。

$$\hat{p} = \frac{1}{n_1}\left(\bar{R}_{2\cdot} - \frac{n_2 + 1}{2}\right) = \frac{\bar{R}_{2\cdot} - \bar{R}_{1\cdot}}{N} + \frac{1}{2} \tag{3}$$

ここで、各群内で Y_{ik} に順位付けを行ったものを $R_{ik}^{(i)}$ とすると、以下の検定統計量 W_N は帰無仮説の下で漸近的に正規分布に従うが、小標本の際は、Satterthwaite の方法にて自由度 $\hat{f} = \frac{\left(n_1 S_1^2 + n_2 S_2^2\right)^2}{\left(n_1 S_1^2\right)^2/(n_1 - 1) + \left(n_2 S_2^2\right)^2/(n_2 - 1)}$ なる t 分布に近似させる。これを Brunner Munzel 検定と呼ぶが、本稿では後者の t 分布に近似させる方法で議論を進める。

$$\begin{aligned} W_N &= \frac{\sqrt{N}\left(\hat{p} - \frac{1}{2}\right)}{\hat{\sigma}_N} = \frac{1}{\sqrt{N}} \cdot \frac{\bar{R}_{2\cdot} - \bar{R}_{1\cdot}}{\hat{\sigma}_N} \\ S_i^2 &= \frac{1}{n_i - 1}\sum_{k=1}^{n_i}\left(R_{ik} - R_{ik}^{(i)} - \bar{R}_{i\cdot} + \frac{n_i + 1}{2}\right)^2 \\ \hat{\sigma}_N^2 &= N\left(\frac{S_1^2}{n_1 n_2^2} + \frac{S_2^2}{n_1^2 + n_2}\right) \end{aligned} \tag{4}$$

Brunner Munzel 検定の実行例を紹介する目的で、Lumley（1996）にて紹介されている「腹腔鏡手術を行った患者 25 例に対して群 1（14 例）又は群 2（11 例）を行い、術後 3 日目における疼痛スコアを測定したデータ」に対して本検定を適用する。

```
data MYDATA ;
  input TRTPN AVAL @@ ; cards ;
  1 1  1 2  1 1  1 1  1 1  1 1  1 1  1 1  1 1  1 1  1 2  1 4  1 1
  1 1  2 3  2 3  2 4  2 3  2 1  2 2  2 3  2 1  2 1  2 5  2 4
 ;
run ;
```

```
%brunner_munzel(_DATA=MYDATA, _GROUP=TRTPN, _VAL=AVAL) ;
```

SAS マクロ「%brunner_munzel」の引数は以下の通り。

- _DATA：使用するデータセット
- _GROUP：群変数
- _VAL：解析変数
- _GROUP1_LABEL=1：解析変数にて群 1 を表す値
- _GROUP2_LABEL=2：解析変数にて群 2 を表す値
- _ALPHA=0.05：有意水準

結果としてデータセット RESULT が出力される。中身は以下の通り。

P_EST	STATISTIC	DFBM	LOWER_CL	UPPER_CL	P_VALUE
0.78896	3.13747	17.6828	0.59522	0.98271	.005786209

- P_EST：p の推定値、効果量は $p = P(Y_{11} < Y_{21}) + \frac{1}{2}P(Y_{11} + Y_{21})$ である（不等号の向きに注意）
- STATISTIC：検定統計量
- DFBM：自由度
- LOWER_CL：（95%）信頼区間の下限
- UPPER_CL：（95%）信頼区間の上限
- P_VALUE：両側 p 値

4. シミュレーションによる考察

WMW 検定や Brunner Munzel 検定の性能を評価した文献・研究は多数あるが、いくつかの先行研究（特にシミュレーション研究）の結果のいくつかは注意して解釈する必要がある。例えば Fagerland（2012）では「WMW 検定は、平均値や中央値の差を見ているわけではない」という警鐘を鳴らしている点では意義がある論文だが、一方で WMW 検定の帰無仮説について正しくない記載が見られたり、「平均値が等しい」あるいは「中央値が等しい」2 つの分布に対して「WMW 検定が有意となった割合」に関するシミュレーション研究を行っているが、シミュレーション手順の詳細が記載されていないため結果の妥当性を評価することができない。

2 項及び 3 項において WMW 検定、Brunner Munzel 検定に関する適切な仮説問題についての考察を行ったが、本項ではシミュレーションによりこれまでの考察の正しさを示したい。なお、有意水準は両側 5%とした。

4.1. 2 群間の中央値が等しい場合

過去の SAS ユーザー総会の論文の引用とその拡張

工藤（2015）よりシミュレーション結果を引用する。各投与群の応答変数の分布と要約統計量が下記である場合、すなわち応答変数の実現値は整数であり分布の中央値は両群で等しく、かつ中央値付近の密度が両群とも高く、分布の裾が一方の分布だけ広がる場合を考える。

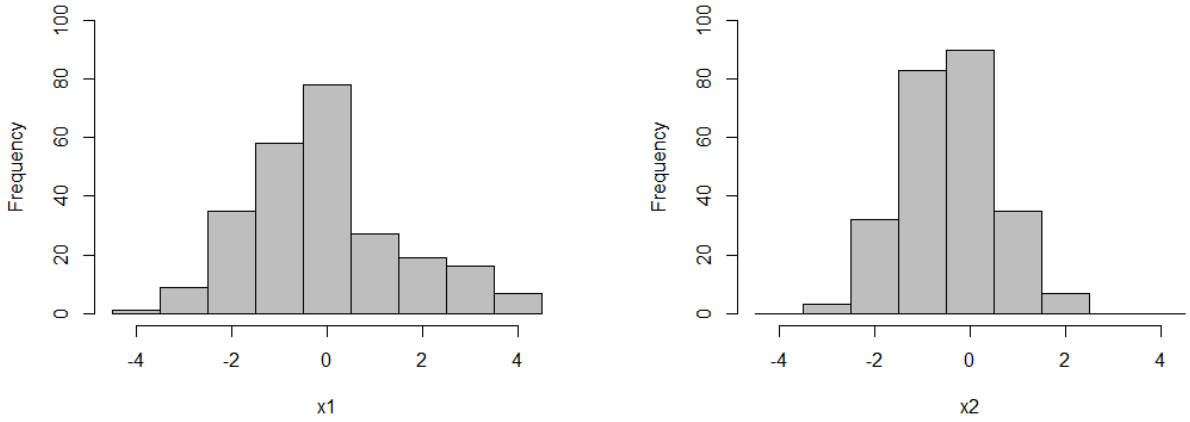


図 4.1 各群の応答変数の分布（左：群 1、右：群 2）

表 4.1 各群の要約統計量

群	例数	平均値	中央値	標準偏差	最小値	最大値
1	250	-0.1	0	1.6	-4	4
2	250	-0.4	0	1.0	-3	2

上記のデータに対して、npar1way プロシジャにより Wilcoxon 順位和検定及び Hodges-Lehmann 型による δ の推定を実施すると、Wilcoxon 順位和検定の結果は p=0.0471、δ の点推定値及び両側 95%信頼区間の推定結果は全て 0 となり、解釈が難しい結果であった（以下の結果は抜粋）。

Wilcoxon の順位和検定(2 標本)

正規近似		
Z		1.9855
片側	Pr > Z	0.0235
両側	Pr > |Z|	0.0471

Hodges-Lehmann 推定値

Location Shift	0.0000

95% 信頼限界		区間の中間点	漸近標準誤差
0.0000	0.0000	0.0000	0.0000

同様のシミュレーション例として、応答変数の実現値は整数であり分布の中央値は両群で等しく、かつ分布の裾が右方向のみ広がった分布を考える。

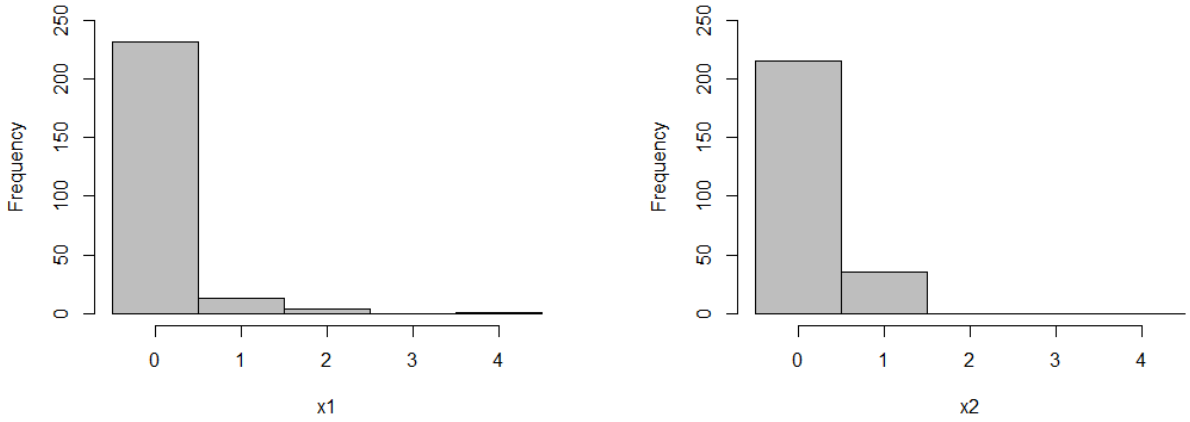


図 4.2 各群の応答変数の分布（左：群 1、右：群 2）

表 4.2 各群の要約統計量

群	例数	平均値	中央値	標準偏差	最小値	最大値
1	250	0.1	0	0.4	0	4
2	250	0.1	0	0.3	0	1

上記のデータに対して、npar1way プロシジャにより Wilcoxon 順位和検定及び Hodges-Lehmann 型による δ の推定を実施すると、Wilcoxon 順位和検定の結果は p=0.0181、δ の点推定値及び両側 95%信頼区間の推定結果は全て 0 となり、やはり解釈が難しい結果であった（以下の結果は抜粋）。

Wilcoxon の順位和検定（ 2 標本）

正規近似		
Z		-2.3644
片側	Pr < Z	0.0090
両側	Pr < |Z|	0.0181

Hodges-Lehmann 推定値

Location Shift	0.0000

95% 信頼限界		区間の中間点	漸近標準誤差
0.0000	0.0000	0.0000	0.0000

混合分布を用いたシミュレーション

Proposition 5 より、同じ点において対称な確率密度関数を持つ 2 つの確率変数は、2 群間の中央値が等しいが、$\Pr(Y_F > Y_G) = \frac{1}{2}$ が成り立つため非対称な分布で 2 群間の中央値が等しい分布を考えたいが、非対称な分布の代表的な分布であるガンマ分布の中央値は閉形式（closed form）で表現できないため、以下のアプローチにより中央値が等しい非対称な 2 つの分布を生成し各検定のパフォーマンスをシミュレーションにより検討する。以下の確率変数を考える。中央値は $inf\{x: F(x) \geq 0.5\}$ で定義されるため、この確率変数の中央値は m となることに注意する。ここで $F(x)$ は確率変数 X の累積分布関数を表す。なお、以下のモデルは一般にハードルモデルと呼ばれる。

$$X = \begin{cases} m & (\text{確率 } p\ (\geq 0.5)) \\ x\ (> m) & (\text{確率 } 1-p) \end{cases} \tag{5}$$

この確率変数 X に従う乱数は以下のように生成できる。確率変数 Z の確率密度関数に含まれる $g(z)$ は、確率密度関数であれば何でも良く、例えば $g(z)$ としてパラメータの異なるガンマ分布を指定することで、同じ中央値 m を持つが異なる確率分布に従う乱数が容易に発生できる。

$$X = mP + (1 - P)Z \tag{6}$$

- 確率変数 Z の確率密度関数：$f(z) = \frac{g(z)}{\int_m^\infty g(z)dz}\ (z \geq m)$
- 確率変数 P：パラメータ $p\ (\geq 0.5)$ のベルヌーイ分布に従う
- P と Z は独立

確率変数 X に従う乱数を生成する SAS マクロ「%Rand」の引数は以下の通り。

- _DATA：出力するデータセット名
- _P：確率変数 P のパラメータ p（デフォルトでは 0.6）
- _N：生成する乱数の個数
- _M：式 (6) 中の m（中央値）
- _DIST：$g(z)$ が従う確率密度関数、CDF 関数に指定できる確率分布が指定できる
  SAS Support - CDF 関数：https://support.sas.com/documentation/cdl_alternate/ja/lefunctionsref/67960/HTML/default/n0n7cce4a3gfqkn1vr0p1x0of99s.htm
- _PARAM：引数_DIST で指定した確率分布のパラメータ

例として、$g(z)$ としてガンマ分布を指定した場合（$m = 1$、ガンマ分布の形状パラメータと尺度パラメータは、それぞれ $(2, 1)$ ）の乱数の生成例を示す。

```
%Rand(_M=1, _DIST='GAMMA', _PARAM=%str(2, 1)) ;
proc means data=wk mean median stddev min q1 q3 max ;
  var x ;
run ;
```

分析変数 : x

平均	中央値	標準偏差	最小値	下側四分位点	上側四分位点	最大値
1.5973053	1.0000000	1.1095590	1.0000000	1.0000000	1.8107937	13.0377718

上記の混合分布を用いて、WMW 検定や Brunner Munzel 検定が中央値に関する検定手法でないことを確認するためのシミュレーションを行う。

表 4.3 シミュレーション結果

検定手法	混合分布			
	(1)	(2)	(3)	(4)
WMW	5.02	**17.20**	**40.44**	**53.59**
Brunner Munzel	5.06	**16.99**	**38.71**	**50.63**
t	5.00	**83.02**	**99.98**	**100**
Welch	4.98	**82.89**	**99.98**	**100**

2 群（各群 100 例）の母集団が上記の混合分布に従っており、$g(z)$ としてガンマ分布を指定（m=1）、4 つのケース（1）～（4）を想定、すなわちガンマ分布の形状パラメータと尺度パラメータをぞれぞれ(2, 1)、(2, 2)、(2, 4)、(2, 8)を想定する。2 群のうち一方が（1）、他方が（1）～（4）のいずれかであり、両群の中央値が共に $m = 1$ で等しい下で「WMW 検定、Brunner Munzel 検定、t 検定及び Welch 検定の各手法にて有意差がみられる」回数を 10 万回のシミュレーションにて評価した。結果は表 4.3 となっており、特にケース（2）～（4）の結果より、WMW 検定と Brunner Munzel 検定は「2 群の中央値が等しい」という帰無仮説に対する検定ではないことが分かる。

4.2. 2 群間の 3 次以下のモーメントが全て等しく 4 次以上のモーメントが異なる場合

WMW 検定の誤用の一つに「2 群間の分散が等しくない場合に WMW 検定が使える」という主張があるようであるが、次項のシミュレーションにて示す通りこの主張は正しくない。一方、本項では 2 群の確率分布において分散を含む 3 次以下のモーメント（平均値、分散及び歪度）が全て等しい場合でさえも、4 次以上のモーメント（尖度他）が異なっている場合は WMW 検定が適切でないことを示したい。
まず、標準正規分布に従う独立な確率変数 X, Z に対し、独立な確率変数 L、Y を考える。つまり、「平均値、分散及び歪度（に加え中央値）は等しいが尖度が異なる」確率変数を考える。

- 確率変数 $X, Z \sim$ 標準正規分布：平均値＝中央値＝0、分散＝1、歪度＝0、**尖度＝0**
- 確率変数 $L \sim$ ラプラス分布$(0,\ 1/\sqrt{2})$：平均値＝中央値＝0、分散＝1、歪度＝0、**尖度＝3**
- 確率変数 $Y = Z^3/\sqrt{15}$：平均値＝中央値＝0、分散＝1、歪度＝0、**尖度＝43.2**

上記の確率変数を用いて、2 群間で分散が等しくても $F = G$ が成り立っていない状況では、WMW 検定は適切でないことを確認するためのシミュレーションを行う。2 群の母集団が、それぞれ「① 確率変数 X と L に従う」「② 確率変数 X と Y に従う」状況の下で、2 群の例数を (100,100)、(20,100)、(100,20) とした場合に「WMW 検定、Brunner Munzel 検定、t 検定及び Welch 検定の各手法にて有意差がみられる」かを判定することを 10 万回のシミュレーションにて行った。まず、①（4 次以上のモーメントの差が大きくない場合）では、2 群の例数が (100,100)、(100,20) の場合は WMW 検定を含むいずれの検定手法も有意水準が概ね制御できた一方、2 群の例数が (20,100) の場合では WMW 検定にて有意水準が制御できないことが示唆された。次に、②（4 次以上のモーメントの差が大きい場合）では、2 群の例数が (100,100) の場合は WMW 検定にて、2 群の例数が (20,100) の場合は WMW 検定及び t 検定にて有意水準が制御できないことが示唆された。一方で、②において 2 群の例数が (100,20) の場合では WMW 検定を含むいずれの検定手法も有意水準が概ね制御できた。注目すべきは、2 群の例数が (20, 100) や (100, 20) 等の不等例数の場合、WMW 検定が有意水準を制御できるか否かが異なる結果となった点である。なお、不等例数の場合にて Brunner Munzel 検定で名目有意水準から外れる傾向が確認された点については次項で考察する。
Wetherill（1960）では、一方が標準正規分布、他方が標準正規分布にノイズを加えたある意味で正規分布に近い分布についての局所対立仮説における検出力に及ぼす歪度と尖度の影響を数値計算によって考察しているが、WMW 検定は t 検定に比べ歪度と尖度の影響を強く受けることを示した。また、この場合の WMW 検定の検定統計量の分散も算出しており、両群の例数が異なる場合には、帰無仮説の下での分散より大きくなる場合もあれば小さくなる場合もあることが示されている。そしてその大小を決めているのが両群の例数の差及び尖度の符号である。一方、上で記載したように我々のシミュレーションにおいても、尖度が正の時、2 群の例数が (20, 100) や (100, 20) 等の不等例数の場合、WMW 検定が有意水準を制御できたり出来なかったりすることが示された。対象としている分布は異なるため、Wetherill（1960）の結果をそのまま当てはめることはできないが、もし考えている分布の下での WMW 検定の理論的な分散式が導出できれば、Wetherill（1960）で示された分散式と同じようなものを求められるかもしれず、この現象を説明できるかもしれない。

表 4.4 シミュレーション結果

検定手法	①			②		
	(100,100)	(20,100)	(100,20)	(100,100)	(20,100)	(100,20)
WMW	4.92	**6.58**	3.66	**6.80**	**12.93**	1.44
Brunner Munzel	4.88	5.26	5.26	4.99	5.08	5.47
T	4.82	5.27	4.81	4.79	**7.34**	3.82
Welch	4.82	5.02	4.78	4.78	5.07	3.71

4.3. $\Pr(Y_F > Y_G) = \frac{1}{2}$ が成り立っている場合

<u>シミュレーション設定</u>

WMW 検定、Brunner Munzel 検定、t 検定及び Welch 検定の 4 手法について「各検定における帰無仮説が棄却される割合（以下、帰無仮説の棄却割合）」に関する比較・考察を試みる。同様の検討はいくつかの先行研究にて既になされており、例えば Fagerland and Sandvik（2009）、Fagerland and Sandvik 他（2011）、Fagerland（2012）、及び名取（2014）等があるが、本項では 2 群の母集団が「正規分布、一様分布、対数正規分布、ロジスティック分布、アーラン分布、テーブル分布」のいずれかに従っており、$\Pr(Y_F > Y_G) = \frac{1}{2}$ が成立している状況の下、2 群の分散が異なると仮定した場合において、有意水準 5%に対する帰無仮説の棄却割合を算出する目的でシミュレーションを行う。なお、各確率分布において $\Pr(Y_F > Y_G) = \frac{1}{2}$ を仮定した場合、2 群間の関係は以下となり、対数正規分布やアーラン分布では 2 群間での平均値が等しくないため、t 検定と Welch 検定の結果の解釈を行う際には留意する必要がある。

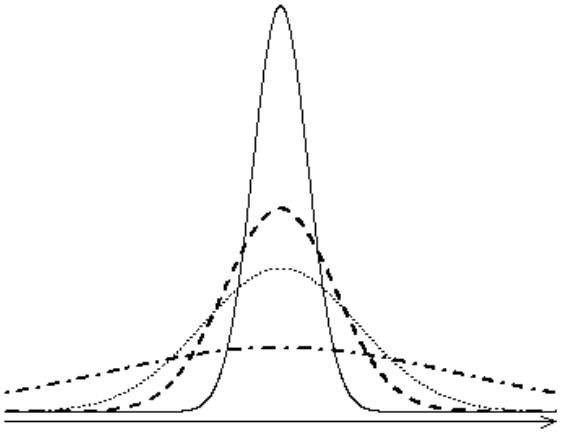
表 4.5 各確率分布に対する 2 群間の関係

確率分布	2 群間の関係
正規分布、一様分布、ロジスティック分布、テーブル分布	等平均値、等中央値、$\Pr(Y_F > Y_G) = \frac{1}{2}$
対数正規分布	等中央値、$\Pr(Y_F > Y_G) = \frac{1}{2}$
アーラン分布	$\Pr(Y_F > Y_G) = \frac{1}{2}$

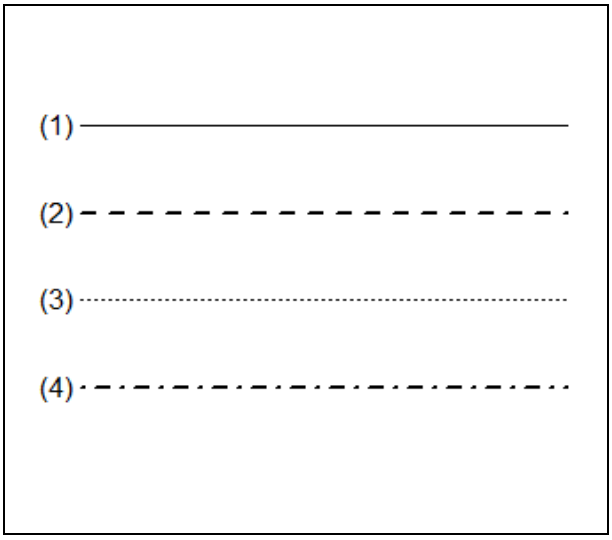
まず、各連続分布について（1）～（4）のケースを想定し、（1）に対する（2）～（4）の分散比をそれぞれ 4、8、40 とした。なお、アーラン分布の rate に関するパラメータは、$\Pr(Y_F > Y_G) = \frac{1}{2}$ かつ分散比が設定通りとなるよう、シミュレーションごとに数値計算にて算出しているため下表には掲載しないこととした。

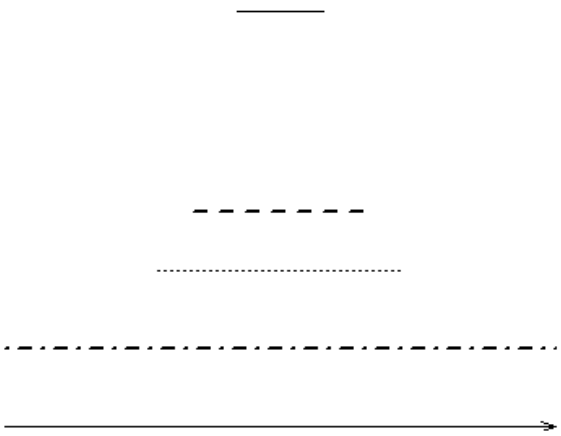
表 4.6 シミュレーションで用いる連続分布

ケース	正規分布	一様分布	ロジスティック分布	対数正規分布	アーラン分布
(1)	N(0,1)	$U(-\sqrt{3}, \sqrt{3})$	location=0, scale=1	$LogN(0, 1^2)$	shape=20
(2)	N(0,4)	$U(-2\sqrt{3}, 2\sqrt{3})$	location=0, scale=2	$LogN(0, 1.25667410^2)$	shape=5
(3)	N(0,8)	$U(-2\sqrt{6}, 2\sqrt{6})$	location=0, scale=$2\sqrt{2}$	$LogN(0, 1.37553188^2)$	shape=3
(4)	N(0,40)	$U(-2\sqrt{30}, 2\sqrt{30})$	location=0, scale=$2\sqrt{10}$	$LogN(0, 1.62839619^2)$	shape=1

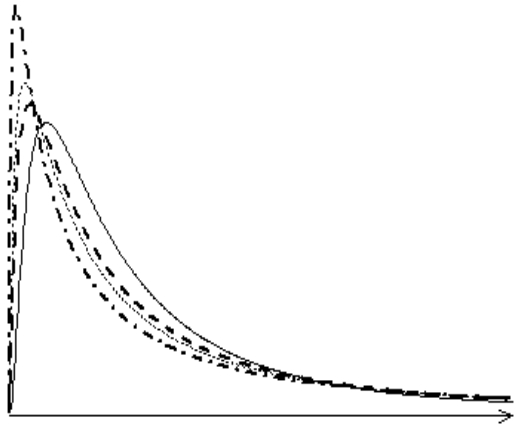


【 正規分布 】

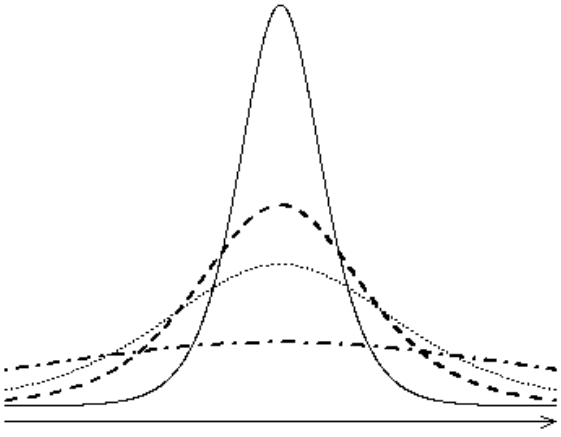




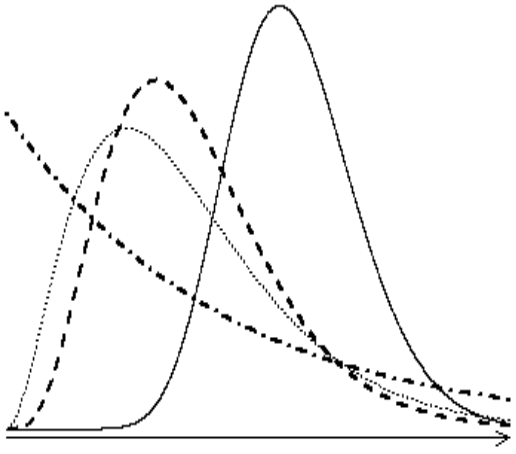
【 一様分布 】



【 対数正規分布 】



【 ロジスティック分布 】



【 アーラン分布 】

図 4.3 シミュレーションの対象とした各確率分布

次に、2 種類のテーブル分布（①：応答が 0～4 の 5 段階、②：応答が 0～10 の 11 段階）について（1）～（4）のケースを想定した。テーブル分布①では、（1）に対する（2）～（4）の分散比はそれぞれ 2.0、3.3、4.5、テーブル分布②では、（1）に対する（2）～（4）の分散比はそれぞれ 2.0、4.7、6.5 となる。

表 4.7 テーブル分布①：応答が 0～4 の 5 段階

ケース	応答	0	1	2	3	4
(1)	確率	0.05	0.10	0.70	0.10	0.05
(2)	確率	0.10	0.20	0.40	0.20	0.10
(3)	確率	0.20	0.20	0.20	0.20	0.20
(4)	確率	0.30	0.15	0.10	0.15	0.30

表 4.8 テーブル分布②：応答が 0～10 の 11 段階

ケース	応答	0	1	2	3	4	5	6	7	8	9	10
(1)	確率	0.01	0.02	0.03	0.04	0.05	0.70	0.05	0.04	0.03	0.02	0.01
(2)	確率	0.02	0.04	0.06	0.08	0.10	0.40	0.10	0.08	0.06	0.04	0.02
(3)	確率	0.09	0.09	0.09	0.09	0.09	0.10	0.09	0.09	0.09	0.09	0.09
(4)	確率	0.15	0.12	0.10	0.06	0.05	0.04	0.05	0.06	0.10	0.12	0.15

<u>2 群の例数が等しい場合</u>

各確率分布について、2 群（各群 100 例）のうち一方が（1）、他方が（1）～（4）のいずれかであり、$\Pr(Y_F > Y_G) = \frac{1}{2}$ が成立している状況の下で「WMW 検定、Brunner Munzel 検定、t 検定及び Welch 検定の各手法にて有意差がみられる」か判定することを 10 万回くり返すことで、帰無仮説の棄却割合をシミュレーションにて算出した。

表 4.9 シミュレーション結果（2 群の例数が等しい場合：1 群 100 例）

検定手法	正規分布				一様分布			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
WMW	4.97	**5.80**	**6.69**	**8.62**	5.08	**6.58**	**7.56**	**9.29**
Brunner Munzel	5.05	4.96	4.99	5.01	5.16	5.17	5.12	5.12
t	4.99	4.97	5.08	5.14	5.15	5.15	5.18	5.22
Welch	4.99	4.92	4.98	5.03	5.15	5.11	5.10	5.11

検定手法	対数正規分布				ロジスティック分布			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
WMW	4.97	5.05	5.14	5.40	5.08	**5.91**	**6.71**	**8.63**
Brunner Munzel	5.05	5.03	5.03	4.99	5.16	5.15	5.16	5.12
t	4.53	**18.80**	**33.08**	**61.73**	5.08	5.18	5.14	5.20
Welch	4.49	**18.60**	**32.71**	**61.17**	5.08	5.12	5.06	5.07

検定手法	アーラン分布			
	(1)	(2)	(3)	(4)
WMW	4.92	**5.89**	**6.52**	**8.08**
Brunner Munzel	5.00	5.00	4.99	4.91
t	4.97	**15.83**	**31.49**	**87.73**
Welch	4.96	**15.67**	**31.20**	**87.46**

検定手法	テーブル分布①				テーブル分布②			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
WMW	4.97	**5.84**	**7.39**	**8.72**	4.98	**5.80**	**8.41**	**9.51**
Brunner Munzel	5.14	5.09	5.18	5.16	5.15	5.08	5.17	5.13
T	4.99	4.99	5.19	5.24	5.06	5.02	5.19	5.25
Welch	4.98	4.97	5.13	5.17	5.05	5.01	5.13	5.18

以上、正規分布、一様分布、ロジスティック分布及びテーブル分布に関する結果より、WMW 検定では（1）を除いて有意水準が制御できていない一方、Brunner Munzel 検定はほぼ制御できており（以下、追加の考察あり）、2 項の理論的考察の正当性を裏付けるものとなった。

また、対数正規分布やアーラン分布における結果では、表 4.9 シミュレーション結果（2 群の例数が等しい場合：1 群 100 例）や上述の通り、WMW 検定と Brunner Munzel 検定の比較が主となるが、Brunner Munzel 検定についてはいずれのケースでも適切に有意水準が制御できていることが確認できる。一方で WMW 検定についてはアーラン分布では（1）を除いて有意水準が制御できておらず、対数正規分布の場合についても最大 10%以下ではあるものの、（1）から（4）へ異質性が大きくなるについて有意水準がより溢れる傾向になっていることが確認された。

なお、上記考察では、Brunner Munzel 検定はいずれのケースでも適切に有意水準が制御できたと考察しており、2 項にて示した理論的考察と整合しているが、帰無仮説の棄却割合は 5%を少し上回るものも散見された（例えば、一様分布においては 5.12%～5.17%）。各群 100 例という例数が本考察においては少ないことが一因であると考えられ、下表のとおり各群 500 例として同様のシミュレーションを行ったところ、この傾向が無くなることが確認できた。

表 4.10 シミュレーション結果（2 群の例数が等しい場合：1 群 500 例）

検定手法	正規分布				一様分布			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
WMW	5.01	**5.89**	**6.79**	**8.80**	4.99	**6.41**	**7.41**	**9.12**
Brunner Munzel	5.03	4.98	5.00	5.02	5.00	5.00	5.05	5.03
t	5.01	5.02	5.04	5.09	4.98	5.00	5.02	5.05
Welch	5.01	5.01	5.03	5.07	4.98	4.99	4.99	5.03

検定手法	対数正規分布				ロジスティック分布			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
WMW	5.01	5.08	5.19	5.43	4.99	**5.77**	**6.57**	**8.48**
Brunner Munzel	5.03	4.99	4.99	4.95	5.00	5.00	5.02	5.03
t	4.84	**79.81**	**97.12**	**99.25**	5.02	5.00	5.04	5.03
Welch	4.84	**79.79**	**97.11**	**99.25**	5.02	4.99	5.02	5.00

検定手法	アーラン分布			
	(1)	(2)	(3)	(4)
WMW	4.91	**5.89**	**6.68**	**8.19**
Brunner Munzel	4.93	4.94	5.01	4.93
t	4.97	**61.77**	**93.58**	**100.0**
Welch	4.97	**61.73**	**93.56**	**100.0**

検定手法	テーブル分布①				テーブル分布②			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
WMW	4.98	**5.82**	**7.23**	**8.53**	4.92	**5.81**	**8.26**	**9.39**
Brunner Munzel	5.01	5.01	5.00	5.02	4.95	5.04	5.03	5.01
t	5.01	4.98	5.11	5.03	5.05	5.04	5.09	5.08
Welch	5.01	4.98	5.10	5.02	5.05	5.03	5.08	5.06

2 群の例数が小さい／等しくない場合（等分布）

前項と同じく、2 群の母集団が「正規分布、一様分布、対数正規分布、ロジスティック分布、アーラン分布、テーブル分布」のいずれかに従っており、$\Pr(Y_F > Y_G) = \frac{1}{2}$ が成立している状況の下で、帰無仮説の棄却割合を算出する目的でシミュレーションを行うが、2 群の分布が等しいが例数が小さい又は異なる場合（2 群の例数：(10,10)、(10,100)、(20,20)、(20,100) ）を想定する。具体的には、各連続分布については 2 群とも（1）のケース、2 種類のテーブル分布（①：応答が 0～4 の 5 段階、②：応答が 0～10 の 11 段階）のそれぞれについては 2 群とも（3）のケースを想定した。

表 4.11 シミュレーション結果（2 群の例数が小さい／等しくない場合（等分布））

検定手法	正規分布				一様分布			
	(10,10)	(10,100)	(20,20)	(20,100)	(10,10)	(10,100)	(20,20)	(20,100)
WMW	5.27	4.89	4.87	4.97	5.22	4.76	4.93	4.95
Brunner Munzel	**5.51**	**5.59**	5.22	5.27	5.47	**5.53**	5.27	5.19
t	4.99	4.99	5.04	5.09	5.12	4.77	5.02	4.99
Welch	4.84	5.16	4.99	5.01	5.01	5.36	4.99	5.12

検定手法	対数正規分布				ロジスティック分布			
	(10,10)	(10,100)	(20,20)	(20,100)	(10,10)	(10,100)	(20,20)	(20,100)
WMW	5.27	4.89	4.87	4.97	5.22	4.76	4.93	4.95
Brunner Munzel	**5.51**	**5.59**	5.22	5.27	5.47	**5.53**	5.27	5.19
t	3.43	5.19	3.92	4.77	4.80	4.99	4.93	5.04
Welch	2.70	**12.78**	3.58	**8.58**	4.61	4.72	4.89	4.95

検定手法	アーラン分布			
	(10,10)	(10,100)	(20,20)	(20,100)
WMW	5.32	4.96	4.94	4.96
Brunner Munzel	**5.60**	**5.71**	5.31	5.36
t	4.98	5.06	5.12	5.06
Welch	4.79	**5.53**	5.06	5.24

検定手法	テーブル分布①				テーブル分布②			
	(10,10)	(10,100)	(20,20)	(20,100)	(10,10)	(10,100)	(20,20)	(20,100)
WMW	4.82	4.81	4.90	4.94	4.78	4.77	4.89	4.98
Brunner Munzel	**5.55**	**5.58**	5.27	5.24	**5.62**	**5.53**	5.30	5.25
t	5.08	4.85	5.05	4.98	5.09	4.79	5.05	5.00
Welch	4.97	5.45	5.03	5.13	4.98	5.41	5.03	5.11

このシミュレーションは両群の分布が同一である下での各検定のパフォーマンスを検討していることになるが、対象としている分布が正規分布、一様分布、ロジスティック分布及びテーブル分布の場合は、WMW 検定は小標本の影響は受けているが不等例数の影響はあまり大きくないように見受けられた。一方、Brunner Munzel 検定はそれらの影響をより大きく受けているように見受けられた。

また、対数正規分布やアーラン分布における結果では、まず、WMW 検定及び t 検定に関してはその他の分布の場合の結果と同様に小標本又は例数比に偏りがある場合でも比較的適切に名目有意水準が制御されていることが確認された。一方で、Brunner Munzel 検定や Welch 検定に関しては、分布が 2 群間で等しい場合であっても例数比に偏りが大きい場合は（保守的な場合も含めて）名目有意水準から外れる傾向が確認された。

なお、Brunner Munzel 検定において、上記の様な例数が非常に小さい場合にて liberal な傾向となる問題点は Neubert and Brunner（2007）にて既に指摘されており、この場合は Brunner Munzel 検定での「studentized permutation test based on the rank statistic」を推奨している。

2 群の例数が小さい／等しくない場合（不等分散）

前項と同じく、2 群の母集団が「正規分布、一様分布、対数正規分布、ロジスティック分布、アーラン分布、テーブル分布」のいずれかに従っており、$\Pr(Y_F > Y_G) = \frac{1}{2}$ が成立している状況の下で、帰無仮説の棄却割合を算出する目的でシミュレーションを行うが、2 群の分布の分散が異なり、かつ例数が小さい又は異

なる場合（2 群の例数：(10,10)、(100,10)、(20,20)、(100,20) ）を想定する。具体的には、各分布について、それぞれの群で（1）と（3）のケースを想定した。

表 4.12 シミュレーション結果（2 群の例数が小さい／等しくない場合（不等分散））

	正規分布				一様分布			
検定手法	(10,10)	(100,10)	(20,20)	(100,20)	(10,10)	(100,10)	(20,20)	(100,20)
WMW	**6.97**	**15.97**	**6.62**	**14.34**	**7.53**	**18.04**	**7.26**	**15.79**
Brunner Munzel	5.30	5.47	5.24	5.20	5.05	5.44	5.20	5.17
t	**5.80**	**36.14**	5.46	**27.79**	**6.13**	**36.70**	**5.56**	**27.28**
Welch	5.06	5.11	5.07	4.95	5.41	**5.58**	5.17	5.15

	対数正規分布				ロジスティック分布			
検定手法	(10,10)	(100,10)	(20,20)	(100,20)	(10,10)	(100,10)	(20,20)	(100,20)
WMW	**5.50**	**8.59**	5.09	**7.94**	**6.74**	**15.32**	**6.54**	**13.52**
Brunner Munzel	**5.56**	**5.67**	5.22	5.26	5.34	**5.61**	5.24	5.21
t	3.36	**25.24**	5.23	**29.55**	**5.50**	**36.40**	5.32	**27.24**
Welch	2.50	**6.91**	4.61	4.26	4.76	4.90	4.93	4.91

検定手法	アーラン分布			
	(10,10)	(100,10)	(20,20)	(100,20)
WMW	**6.77**	**15.54**	**6.53**	**13.99**
Brunner Munzel	5.34	5.44	5.23	5.31
t	**5.93**	**40.26**	**7.87**	**37.04**
Welch	5.06	5.08	**7.24**	**7.24**

	テーブル分布①				テーブル分布②			
検定手法	(10,10)	(100,10)	(20,20)	(100,20)	(10,10)	(100,10)	(20,20)	(100,20)
WMW	**7.06**	**21.03**	**7.20**	**17.67**	**7.88**	**24.41**	**8.16**	**20.26**
Brunner Munzel	**5.59**	**5.60**	5.41	5.12	5.36	5.48	5.31	5.19
t	**5.82**	**23.20**	5.33	**18.23**	**6.00**	**29.06**	5.43	**22.16**
Welch	5.28	**5.56**	5.15	5.07	5.37	**5.56**	5.14	5.12

対象としている分布が正規分布、一様分布、ロジスティック分布及びテーブル分布の場合は、WMW 検定及び t 検定では概ね有意水準が制御できておらず、特に例数比に偏りが大きい場合にはその影響が大きくなる傾向が見られた。一方、Brunner Munzel 検定では小標本または例数比に偏りが大きい場合に有意水準が溢れる傾向が見られた。

また、対数正規分布やアーラン分布における結果では、WMW 検定と Brunner Munzel 検定の比較が主となるが、2 群の例数が等しい場合（表 4.9 シミュレーション結果（2 群の例数が等しい場合：1 群 100 例））と同様に WMW 検定では概ね有意水準が制御できておらず、特に例数比に偏りが大きい場合にはその影響も大きくなる傾向が見られた。一方、Brunner Munzel 検定については WMW 検定のような有意水準の増大は確

認されなかったものの、表 4.11 シミュレーション結果（2 群の例数が小さい／等しくない場合（等分布））の結果と同様、特に小標本または例数比に偏りが大きい場合では有意水準が溢れる傾向が見られた。

4.4. WMW 検定の inconsistency について

2 項で述べた通り、WMW 検定は仮説問題：H_4: $\{F = G\}$ v.s. K_4: $\{F \neq G\}$ において consistent ではないことが直感的には明らかであるが、それをシミュレーションにて確認してみたい。2 群の母集団のうち一方が（1）N(0, 1)、他方が（2）N(0, 4)、（3）N(0, 8)、（4）N(0, 40) のいずれかの時、「WMW 検定にて有意差がみられる」か判定することを 5 万回くり返す。これは、$\left\{F \neq G, \Pr(Y_F > Y_G) = \frac{1}{2}\right\} \subset \{F \neq G\}$に真のモデルが含まれる場合に「WMW 検定にて有意差がみられる」割合であり、検出力に相当する。2 群の例数として 100 例、1000 例、10000 例（等例数）とした結果が下表となり、シミュレーション結果を見る限りサンプルサイズを大きくしても検出力は 5%は超えるが 100%には近づかないことが伺える。

表 4.13 シミュレーション結果

1 群あたりの例数	正規分布		
	(2)	(3)	(4)
100 例	5.59	6.53	8.49
1000 例	6.08	6.95	8.85
10000 例	5.73	6.60	8.56

5. まとめ

まず、H: $\Pr(Y_F > Y_G) = \frac{1}{2}$ v.s. K: $\Pr(Y_F > Y_G) \neq \frac{1}{2}$ という問題に対して、WMW 検定を適用することは適切ではないことを文献引用及びシミュレーションにて示した。巷では「WMW 検定は t 検定や Welch 検定よりも仮定が少ない（例えば、2 群の母集団に特定の確率分布を仮定しない）ため頑健な検定手法であり、有意水準を適切に制御することができる」と考えられることがあるが、2 項で詳述した通り、H: $\Pr(Y_F > Y_G) = \frac{1}{2}$ に対して WMW 検定を適用すると第 1 種の過誤確率が制御できず、$F = G$ という条件を帰無仮説に加える必要があるが、$F = G$ は強い仮定であり、「WMW 検定は t 検定や Welch 検定よりも仮定が少ない」点は否定したいと考える。$F = G$ の場合にのみ WMW 検定は第 1 種の過誤確率が制御できるという点について、もう一つ強調したいことがある。WMW 検定の誤用の一つに「2 群間の分散が等しくない場合に WMW 検定が使える」という主張があるようであるが、4.3 項及び、4.2 項のシミュレーションで示したように、分散が両群で異なる場合や、分散を含む 3 次以下のモーメントが全て等しい場合でさえ、4 次以上のモーメントが異なれば WMW 検定は適切でないことが分かり、この主張は正しくない。さらにここでのシミュレーションでは、不等例数の影響も大きいことが示唆され、不等例数の際に WMW 検定を実際に適用する際には注意を要する。

次に、「WMW 検定は中央値（や平均値）の違いを見出すもの」とすると誤った結論を導く場合があることも文献引用及びシミュレーションにて示した。すなわち、WMW 検定は特殊な状況を除くと位置母数の検定ではないため、仮に WMW 検定にて有意差がみられたとしても、その後の推定・解釈が困難であることを意味し、例えば、WMW 検定は「位置母数に対する Hodges-Lehmann 推定量」や「適当な統計手法（例えば bootstrap 法）を用いた中央値の群間差の推定」とは対応しない。WMW 検定に対応する推定・解釈を行う際は、Pocock and Ariti（2012）で提案された win ratio や、工藤（2015）で紹介された Wilcoxon-Mann-Whitney オッズ等を用いた推定量を用いることが一案であるが、t 検定や Welch 検定に対応する推定量が「平均値の群間差」である

ことに比べると、win ratio や Wilcoxon-Mann-Whitney オッズを用いた推定・解釈は使いにくい印象があり、特に臨床試験における統計解析にて薬剤の効果を論じる際、統計を専門としない担当者が解釈を行うことは難しいと思われる。ちなみに、この問題意識は Brunner Munzel 検定においても同様であること、logrank 検定において有意差がみられた後、推定・解釈をハザード比にて行うことと類似した状況であることを申し添える。また、$H: \Pr(Y_F > Y_G) = \frac{1}{2}$ v.s. $K: \Pr(Y_F > Y_G) \neq \frac{1}{2}$ という問題に対しては、WMW 検定ではなく Brunner - Munzel 検定を用いるべきであることを述べたが、Brunner -Munzel 検定も万能ではなく、4 項の「2 群の例数が小さい／等しくない場合（等分布）」等でも見た通り、Brunner Munzel 検定が小標本の設定で性能が悪くなる。この点は既に Brunner and Munzel（2000）でも指摘されており、検定統計量の漸近分布の近似精度が悪くなる旨が言及され、Neubert and Brunner（2007）や Nowak et al.（2022）では並べ替え検定や t 分布の自由度の推定方法の改良等が提案されている。本稿では取り扱わなかったものの、小標本等の特殊な設定におけるこれらの手法との比較及び改良については今後の課題と考える。

参考文献

- Karch, J. D., 2021. Psychologists should use Brunner-Munzel's instead of Mann-Whitney's U test as the default nonparametric procedure. Advances in Methods and Practices in Psychological Science, 4(2), 2515245921999602.
- Hajek J., Sidak Z., 1967. Theory of Rank Tests, Academic Press, New York.
- Cox, D.R. Hinkley, D.V., 1974. Theoretical statistics. Chapman and Hall.
- 柳川 堯, 1982. ノンパラメトリック法. 培風館. 新統計学シリーズ〈9〉.
- Hart, A., 2010. Mann-Whitney test is not just a test of medians: differences in spread can be important, BMJ, 323, 391-3.
- Shirahata S., 1993, Estimate of Variance of Wilcoxon-Mann-Whitney Statistics, J. Japanese Soc. Comp. Statist., 6.2, 1-10.
- Fay MP, Proschan MA, 2010. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Statistics Surveys, 4(0), 1-39.
- Lehmann E.L., Romano J. P., 2022. Testing Statistical Hypotheses. Springer Cham.
- 竹内 啓, 1975. 確率分布と統計解析. 日本規格協会.
- Brunner E., Munzel U., 2000. The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation. Biometrical J. 42, 17-25.
- Lumley T., 1996. Generalized estimating equations for ordinal data: A note on working correlation structures. Biometrics 52, 354-361.
- Nadarajah S., Kotz S., 2006. R. Programs for Computing Truncated Distributions. Journal of Statistical Software. August, Volume 16, Code Snippet 2.
- Wetherill, G. B., 1960. The Wilcoxon Test and Non-Null Hypotheses. Journal of the Royal Statistical Society. Ser B, 1960, 22, 2, 402-418.
- Fagerland MW, Sandvik L., 2009. The Wilcoxon-Mann-Whitney test under scrutiny. Statistics in Medicine, 28:1487-1497.
- Fagerland MW, Sandvik L., Mowinckel P., 2011. Parametric methods outperformed non-parametric methods in comparisons of discrete numerical variables. BMC Medical Research Methodology, 11:44.
- Fagerland MW, 2012. t-tests, non-parametric tests, and large studies - a paradox of statistical practice? BMC Medical Research Methodology, 12:78.
- 名取 真人, 2014. マン・ホイットニーの U 検定と不等分散時における代表値の検定法, 霊長類研究 30:173-185.

- Neubert K., Brunner E., 2007. A studentized permutation test for the non-parametric Behrens-Fisher problem. Computational Statistics & Data Analysis. 51, 5192-5204.
- SAS Support - CDF 関数：https://support.sas.com/documentation/cdl_alternate/ja/lefunctionsref/67960/HTML/default/n0n7cce4a3gfqkn1vr0p1x0of99s.htm
- Pocock, SJ, Ariti CA, et. al., 2012. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities, European Heart Journal 33, 176-182
- O'Brien RG, Castelloe J., 2006. Exploiting the link between the Wilcoxon-Mann-Whitney test and a simple odds statistic. Proceedings of the Thirty-First Annual SAS Users Group International Conference.
- 工藤 健太郎, 他, 2015. Wilcoxon-Mann-Whitney オッズ（WMWodds）と Wilcoxon 順位和検定について. SAS ユーザー総会 2015.
- Nowak, C.P., Pauly, M. and Brunner, E., 2022. The nonparametric Behrens-Fisher problem in small samples. arXiv preprint arXiv:2208.01231.

補遺：本文に掲載していないが引用したプログラム

プログラム 3.1：SAS マクロ「%Brunner Munzel」

```
%macro brunner_munzel(_DATA, _GROUP, _VAL, _GROUP1_LABEL=1, _GROUP2_LABEL=2, _ALPHA=0.05) ;
  proc iml ;
    use work.&_DATA. ;
    read all var{&_VAL.} where(&_GROUP. = &_GROUP1_LABEL. & &_VAL. ^= .) into x ;
    read all var{&_VAL.} where(&_GROUP. = &_GROUP2_LABEL. & &_VAL. ^= .) into y ;
    close work.&_DATA. ;

    alpha     = &_ALPHA. ;
    n1        = nrow(x) ;
    n2        = nrow(y) ;
    r1        = ranktie(x) ;
    r2        = ranktie(y) ;
    r         = ranktie( x // y) ;
    R1_bar    = mean(r[1:n1]) ;
    R2_bar    = mean(r[(n1+1):(n1+n2)]) ;
    p_est     = (R2_bar-R1_bar)/(n1+n2)+1/2 ;

    tmp_1     = r[1:n1]          -r1+j(n1,1,-R1_bar+(n1+1)/2) ;
    tmp_2     = r[(n1+1):(n1+n2)]-r2+j(n2,1,-R2_bar+(n2+1)/2) ;
    S2_1      = sum(tmp_1 # tmp_1)/(n1-1) ;
    S2_2      = sum(tmp_2 # tmp_2)/(n2-1) ;
    Sigma2_N  = (n1+n2)*(S2_1/(n1*n2**2)+S2_2/(n2*n1**2)) ;
    statistic = (R2_bar-R1_bar)/sqrt(n1+n2)/sqrt(Sigma2_N) ;
    dfbm      = (n1*S2_1+n2*S2_2)**2/((n1*S2_1)**2/(n1-1)+(n2*S2_2)**2/(n2-1)) ;
    p_value_l = 1 - probt(statistic, dfbm) ;
    p_value_g =     probt(statistic, dfbm) ;
    p_value   = 2 * min(p_value_l, p_value_g) ;
    lower_cl  = p_est-tinv(1-alpha/2,dfbm)*sqrt(Sigma2_N/(n1+n2)) ;
    upper_cl  = p_est+tinv(1-alpha/2,dfbm)*sqrt(Sigma2_N/(n1+n2)) ;

    create RESULT var {p_est  statistic dfbm lower_cl upper_cl p_value} ;
    append ;
  quit ;
  proc print noobs ; run ;
%mend ;
```

プログラム 4.1：シミュレーション①

```
data SAMPLEDATA1 ;
  call streaminit(777) ;
  do TREAT=1 to 2 ;
    do I=1 to 250 ;
      X =  floor( rand("NORMAL",0,1) + 3*(2-TREAT)*floor(0.15+rand("UNIFORM")) ) ;
      output ;
    end ;
  end ;
run ;

proc means data=SAMPLEDATA1 nonobs n mean median stddev min max maxdec=1;
  class TREAT  ;
  var   X ;
run ;

proc npar1way data=SAMPLEDATA1 correct=no hl ;
  class TREAT  ;
  var   X ;
run ;
```

プログラム 4.2：シミュレーション②

```
data SAMPLEDATA2 ;
  call streaminit(777) ;
  do TREAT=1 to 2 ;
    do I=1 to 250 ;
      X =  floor( rand("WEIBULL", 0.8*TREAT, 0.35*TREAT) ) ;
      output ;
    end ;
  end ;
run ;

proc means data=SAMPLEDATA2 nonobs n mean median stddev min max maxdec=1;
  class TREAT  ;
  var   X ;
run ;

proc npar1way data=SAMPLEDATA2 correct=no hl ;
  class TREAT  ;
  var   X ;
run ;
```

プログラム 4.3：SAS マクロ「%Rand」

```
%macro Rand(_DATA=wk, _N=100000, _P=0.6, _M=1, _DIST='GAMMA', _PARAM=%str(2, 1), _SEED=12345) ;
  data &_DATA.(keep=i x) ;
    call streaminit(&_SEED.) ;
    do i=1 to &_N. ;
      u=rand('UNIFORM') ;
      y=cdf(&_DIST., &_M., &_PARAM.)+u*(1-cdf(&_DIST., &_M., &_PARAM.)) ;
      z=quantile(&_DIST., y, &_PARAM.) ;
      p=rand('BERNOULLI', &_P.) ;
      x=&_M.*p+(1-p)*z ;
      output ;
    end ;
  run ;
%mend ;
```

ベイズ流ワイブル回帰に基づく Commensurate Prior 法のシミュレーションによる動作特性の検討

○高戸 俊一[1,4], 田中 優一[2,4], 吉本 拓矢[3,4]
([1]杏林製薬株式会社　開発推進部、[2]大鵬薬品株式会社　データサイエンス部、[3]中外製薬株式会社　バイオメトリクス部、[4]日本製薬工業協会　医薬品評価委員会　データサイエンス部会)

Evaluation of Operating Characteristics of the Commensurate Prior Method Based on Bayesian Weibull Regression through Simulation

Shunichi Takado[1,4], Yuichi Tanaka[2,4], Takuya Yoshimoto[3,4]
([1]Clinical Sceience & Affairs, Kyorin Pharmaceutical Co., Ltd., [2]Data Science Dept., Taiho Pharmaceutical Co., Ltd., [3]Biometrics Dept., Chugai Pharmaceutical Co., Ltd. [4]Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association)

要旨

臨床試験のデザイン・解析において、新規試験の対照群と外部対照データを併合して対照群を構成する方法は hybrid control アプローチと呼ばれる。hybrid control アプローチを利用した事例として、FDA の Complex Innovative Trial Design pilot meeting program に採択された、びまん性大細胞型 B 細胞リンパ腫を対象とした臨床試験がある。当該臨床試験の重要な副次評価項目である全生存時間に対する解析では、ベイズ流ワイブル回帰に基づく Commensurate Prior 法（CMP 法）が利用されている。CMP 法は、外部対照データと臨床試験の対照群の類似性をパラメータ化し、両者の共通性に応じて動的に外部データの借用程度を調整するベイズ流手法である。本稿では、当該臨床試験を参考に、外部対照データを利用した際の CMP 法の動作特性をシミュレーション実験により検討する。具体的には、外部データを利用しない場合およびすべて利用する場合の解析結果と CMP 法を比較し、第一種の過誤確率、検出力、バイアス、平均二乗誤差、Effective Historical Sample Size への影響を評価・報告する。

キーワード：シミュレーション, Commensurate Prior 法, ベイズ流ワイブル回帰モデル

1. はじめに

医薬品の開発において、実際の医療環境下で取得されたデータを含む、実施する試験以外のデータソースから得られたデータの利活用を試みる動きが活発化している[1]。ICH E8（R1）臨床試験の一般指針では、過去の臨床研究、全国的な死亡データベース、疾患及び薬剤レジストリ、医療保険請求データ、並びに日常的な医療行為からの医療及び管理の記録を含む外部データの二次利用について記載されており、試験計画時に考慮すべき内容が明記された[2]。

外部データの利用は、生命を脅かす希少疾患や、不可逆的な転帰を辿る疾患、小児医薬品開発等、同時対照群の設定が倫理的な観点から困難な場合に薬効を評価する方法の 1 つとして考えられている [3]。ICH-E10 ガイドラインでは、臨床試験における対照群の種類の 1 つとして試験に参加して治療を受ける集団とは別の対照集団（外部対照群）を設けることが述べられている [4]。臨床試験での外部対照の利用方法については、単群試験の外部対照群とすることや、事前に定義したアルゴリズムやマッチング等の統計手法を用いて、二重盲検試験の対照群と併合することが考えられる。このように実施しようとする臨床試験のデザイン・解析に外部データを対照群として利用する考え方をヒストリカルコントロールアプローチと呼び、その中でも後者の二重盲検試験の対照群と外部データを併合して対照群を構成する考え方を hybrid control アプローチと呼ぶ [5]。

ICH-E10 ガイドラインでは、外部対照デザインの使用は、治療効果が劇的であり、疾患の通常の経過が十分に予測可能である場合に認められると限定されている。さらに、外部対照を採用するのは、エンドポイントが客観的であり、評価項目に対するベースラインや治療変数の影響の特徴が十分に分かっている場合に限ることが述べられている。また新規試験の対照群の一部として利用可能な既存試験の条件として Pocock（1976）[6] は 6 つの条件を提唱している。しかし提唱された条件を全て満たす外部対照データが存在することは現実的にはほとんどない。そこで、全ての条件が厳密に満たされていない状況でヒストリカルデータの情報を活用する有効な方法を考える必要がある [7]。単一の外部対照データしかない場合、全ての条件が満たされていない状況下だと、新規試験の対照群と外部対照データの間に異質性が生じる可能性が高く、試験間で結果が変動（drift）してしまう場合、信頼性をもって治療効果を推定することが難しい。その場合、解析担当者が事前に drift に対する情報を付与する必要があり、これが検出力や第一種の過誤確率に大きな影響を与える。Hobbs（2011）[8] らは、外部対照データと新規試験の対照群の類似性をパラメータ化し、外部対照データと新規試験の対照群の共通性（commensurability）に応じて、借用の程度を調整する hybrid control アプローチを提案している（commensurate prior 法。以下、CMP 法）。本手法は，Complex Innovative Trial Design（CID）の pilot program に採択されたびまん性大細胞型 B 細胞リンパ腫（Diffuse large B-cell lymphoma; DLBCL）を対象とした臨床試験でその適用が検討された。

ベイズ流試験デザインにおいて，事後分布が解析的に解けないケースはしばしばみられ、適用にあたって試験開始前に広範なシミュレーション実験を行い、提案するデザインの動作特性を精緻に評価しておくことが必要となる。一方で複数のシナリオに基づくシミュレーション実験が必要となる場合などは、計算負荷の課題が存在するため、より効果的にその評価性能を考察することが肝要である。本稿では，DLBCL の事例を取り上げ，CMP 法の適用並びに動作特性の評価を可能とする R パッケージを用いて当該デザインの評価性能を考察する方法を紹介する。

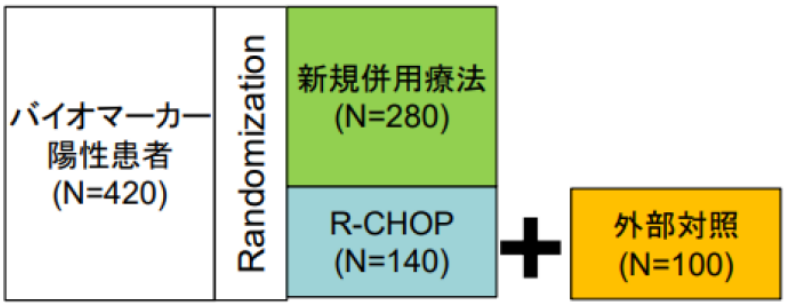
2. 本稿で取り上げる臨床試験

2.1 臨床試験デザイン

CID のウェブサイトで CID pilot meeting program（https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program）に採択された試験の 1 つとして公開されている、DLBCL を対象とした臨床試験の事例を取り上げる。本試験では、バイオマーカー陽性患者（N=420）を 2:1 の割付比で新規併用療法（N=280）、対照群として標準的な化学療法レジメンである R-CHOP（N=140）に組み入れ、さらに外部対照データとして 100 例分のデータを用意する。本試験では主要な副次的評価項目として全生存期間（OS）を設定し、OS の解析集団は外部対照群からの患者を加えることで補強される。OS の解析ではワイブル分布を仮定したベイズ流の CMP 法を採用し、外部対照群からの情報を動的に借用する。さらに、傾向スコアマッチングを行い、解析に含める外部対照患者を選定する。治療効果の推定は、ハザード比の事後分布の事後平均と 95%信用区間に基づいて行われ、ハザード比が 1 を下回る事後確率が 97.5%を超える場合に治療効果があると判断する。

本試験のデザインのシェーマを図 1 に示す。

図 1　臨床試験デザイン



2.2 解析方法

主要な副次評価項目である OS に対する解析では、ワイブル分布を仮定した commensurate prior を用いたベイズ流手法を利用する。commensurate prior の数理的側面を以下に示す。

新規試験（Internal Control: IC）および外部対照群（External Control: EC）の被験者iの time to event アウトカムの確率変数をT_iとする。

また被験者iにおけるイベントの発生または打ち切りの情報を変数$\delta_i \in \{0,1\}$で表し、以下のように定義する。

$$\delta_i = \begin{cases} 1\ (\textit{イベントが観察された場合};\ uncensored) \\ 0\ (\textit{打ち切りの場合};\ censored) \end{cases}$$

被験者iに関連する共変量ベクトルを$\boldsymbol{X}_i$とし、$\boldsymbol{X}_i$は当該被験者が外部対照データか否かを示すダミー変数（$X_{ext,i}$）、治療介入の有無を示すダミー変数（$X_{trt,i}$）、およびその他の患者背景因子（$\boldsymbol{X}_{cov,i}$）から構成される。本モデルにおける確率分布のパラメータ全体を$\boldsymbol{\theta}$とし、以下のように定義する。

$$\boldsymbol{\theta} = (\beta_0, \beta_{ext}, \beta_{trt}, \boldsymbol{\beta}_{cov}, \gamma)$$

- β_0：ワイブル回帰における切片項

- β_{ext}：外部対照群であることを示すダミー変数に対応する回帰係数
- β_{trt}：治療群に属することを示すダミー変数に対応する回帰係数
- $\boldsymbol{\beta_{cov}}$：患者背景に対応する回帰係数ベクトル
- γ： ワイブル分布の形状パラメータ

ワイブル分布は、ハザード関数が時間とともに増加または減少する可能性を許容するため、ハザードが一定である指数分布よりも柔軟な生存時間モデルとして広く用いられている。
ワイブル分布の確率密度関数は$f(t_i) = \gamma\lambda_i t_i^{\gamma-1}\exp\left(-\lambda_i t_i^{\gamma-1}\right)$と表すことができ、ワイブル分布の尺度パラメータ$\lambda_i$に以下を設定することで、ワイブル回帰モデルを構築する。

$$\lambda_i = \exp\left(\beta_0\left(1 - X_{ext,i}\right) + \beta_{ext}X_{ext,i} + \beta_{trt}X_{trt,i} + \boldsymbol{\beta_{cov}}\boldsymbol{X}_{cov,i}\right)$$

ベイズ流アプローチにより各パラメータを推定する際には$\boldsymbol{\theta}$に対して事前分布を設定する。また各パラメータが従う事前分布を規定するパラメータ（超パラメータ）の分布（超事前分布）も併せて設定する。ここでτは外部対照データと新試験データの類似性を表す指標である。

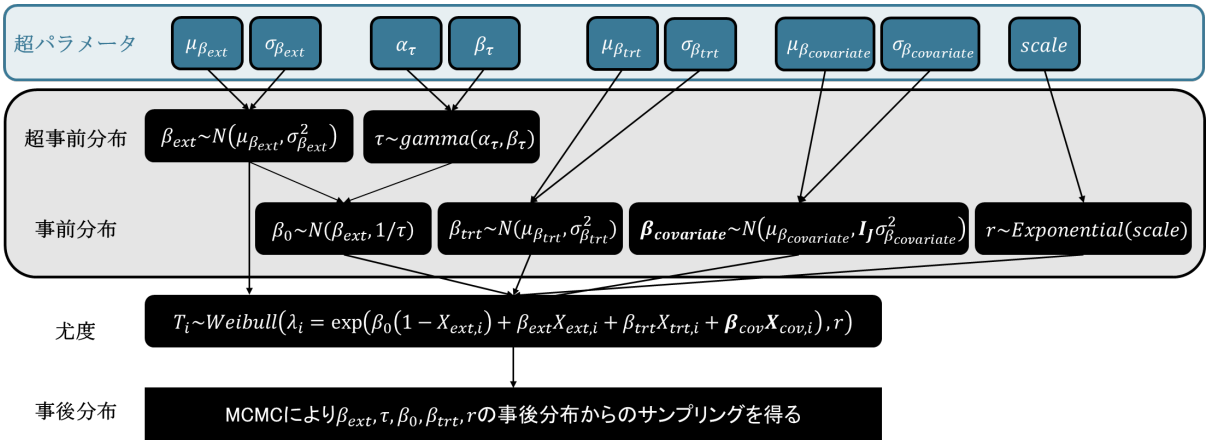
・ 超事前分布

$$\beta_{\text{ext}} \sim N(0, 1000), \qquad \tau \sim gamma(1, 0.001)$$

・ 事前分布

$$\beta_0 \sim N(\beta_{\text{ext}}, 1/\tau), \qquad \beta_{\text{trt}} \sim N(0, 1000), \qquad r \sim \exp(1),$$
$$\boldsymbol{\beta_{cov}} \sim N(0, 1000)$$

以上の設定値を概念図として示すと下記になる。



3. シミュレーション

3.1 シミュレーションの目的

本シミュレーション実験では、CMP 法を用いて外部対照データを利用した際の第一種の過誤確率、検出力、バイアス、平均二乗誤差（mean squared error：MSE）、effective historical sample size （EHSS）への影響を検討するために、複数の効果の想定のもとでその動作特性を確認する。

シミュレーションの際に生成する仮想データは実際の試験を想定し、外部対照データと新規臨床試験データの患者背景因子が異なることを想定する。

DLBCL の臨床試験では傾向スコアマッチングにて、借用する外部対照データを特定しているが、本シミュレーション実験ではマッチングした集団を特定することはせず、共変量はモデルで調整する。

3.2 シミュレーションの設定

本シミュレーションでは DLBCL の臨床試験を参考としているが、詳細なデザインが不明であった被験者の登録期間などは任意で設定している。また症例数など一部の設定値は参考試験のデザインと異なるものを使用している。本シミュレーションで実施した試験デザインの設定値を示す。

新規試験に組み入れる被験者数は 300 名とし、新規併用療法群と対照群への割り付け比率を 2：1 とする（新規併用療法群：200 名、対照群：100 名）。

新規試験と外部対照データのそれぞれの被験者を登録する期間は 24 ヶ月と想定し、一定の割合・間隔で登録されると想定する。また、最後の被験者の登録から 36 ヶ月間を追跡期間とし、当該期間を終えた時点で解析することとする。

3.2.1 患者背景（共変量）の設定

はじめに、評価項目 OS に対する共変量$\boldsymbol{\beta}_{cov}$は 5 つあると想定し、新規試験と外部対照データの患者背景分布に違いが生じるように設定する。5 つの共変量のうち、3 つは 2 値のカテゴリカル変数で、2 つは連続変数とする。

・新規試験の患者背景の分布

共変量が多次元正規分布に従うと仮定し、平均および共分散構造を以下の通り設定する。

$$\begin{bmatrix} covatiate1 \\ covatiate2 \\ covatiate3 \\ covatiate4 \\ covatiate5 \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} 0 \\ 0.5 \\ 0.5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.7 & 0 & 0 \\ 0.5 & 1.2 & 0.9 & 0 & 0 \\ 0.7 & 0.9 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.7 \\ 0 & 0 & 0 & 0.7 & 0.7 \end{bmatrix} \right)$$

このとき、連続変数を閾値で区切り、2 値のカテゴリカル変数を作成する事とする。covariate1、covariate 2、covariate 4 を以下のルールに基づき変換し、2 値のカテゴリカル変数 category1、category2、category4 を作成する。

$$\text{category}_{1,\text{i}} = \begin{cases} 1(covariate_{1,i} \geq 0.45) \\ 0(covariate_{1,i} < 0.45) \end{cases} \quad \text{category}_{2,\text{i}} = \begin{cases} 1(covariate_{2,i} \geq 0.55) \\ 0(covariate_{2,i} < 0.55) \end{cases} \quad \text{category}_{4,\text{i}} = \begin{cases} 1(covariate_{4,i} \geq 0.50) \\ 0(covariate_{4,i} < 0.50) \end{cases}$$

・外部対照試験の患者背景の分布

新規試験と同様に、共変量が多次元正規分布に従うと仮定し、平均及び共分散構造を以下の通り設定する。

$$\begin{bmatrix} covariate1 \\ covariate2 \\ covariate3 \\ covariate4 \\ covariate5 \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} 0.7 \\ 0.5 \\ 0.9 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.7 & 0 & 0 \\ 0.5 & 1.2 & 0.9 & 0 & 0 \\ 0.7 & 0.9 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.7 \\ 0 & 0 & 0 & 0.7 & 0.7 \end{bmatrix} \right)$$

新規試験と同様に、covariate1、covariate 2、covariate 4 を以下のルールに基づき変換し、2 値のカテゴリカル変数 category1、category2、category4 を作成する。

$$\text{category}_{1,i} = \begin{cases} 1(covariate_{1,i} \geq 0.65) \\ 0(covariate_{1,i} < 0.65) \end{cases} \quad \text{category}_{2,i} = \begin{cases} 1(covariate_{2,i} \geq 0.55) \\ 0(covariate_{2,i} < 0.55) \end{cases} \quad \text{category}_{4,i} = \begin{cases} 1(covariate_{4,i} \geq 0.50) \\ 0(covariate_{4,i} < 0.50) \end{cases}$$

3.2.2　生存時間データの生成

次に、新規試験と外部対照データにおける評価項目 OS のデータ発生条件を検討する。OS はワイブル分布に従うと仮定し、尺度パラメータλ_iと形状パラメータ$\gamma = 0.9$のワイブル分布に基づいてデータを発生させる。

ワイブル分布の尺度パラメータλ_iの設定について、新規試験の治療群間差（ハザード比：HR）と新規試験の対照群を基準としたときの外部対照データの治療群間差（drift HR）を用いて以下のように表現する。なおλ_cは新規試験の対照群における尺度パラメータである（$\lambda_c = \exp(\beta_0)$）。

- Internal control: $\lambda_i = \lambda_c \exp(\boldsymbol{X_{cov,i}}\boldsymbol{\beta_{cov}}) = \lambda_c \times \exp(0.5category1_i + 0.5category2_i + 0.5covariate3_i + 0.5category4_i + 0.5covariate5_i)$
- Internal treatment: $\lambda_i = \lambda_c \times HR \times \exp(0.5category1_i + 0.5category2_i + 0.5covariate3_i + 0.5category4_i + 0.5covariate5_i)$
- External control: $\lambda_i = \lambda_c \times driftHR \times \exp(0.5category1_i + 0.5category2_i + 0.5covariate3_i + 0.5category4_i + 0.5covariate5_i)$

生存時間データの生成には逆関数法を用いる。まず一様分布$U_i \sim Uniform(0,1)$を発生させ、生存関数$S_i(t) = \exp(-\lambda_i t^{\gamma})$に対し$U_i = S_i(T_i)$とおく。この関係式から下記のように$T_i$を算出することができる。

$$T_i = \left(\frac{-\log(U_i)}{\lambda_i}\right)^{\frac{1}{\gamma}}$$

検討するシナリオとして、HR 及び drift HR を変更した場合の動作特性を評価する。以下に OS の中央値と検討する HR 及び drift HR を示す。

パラメータ	設定値
新規試験の対照群の効果（λ_c）	Log(2)/24（OS の中央値 24ヶ月に対応する値）
新規試験の治療群間差（HR）	1, 0.65
新規試験の対照群と外部対照データの差（drift HR）	1, 1.2

最後に、転院等により OS が追跡できない状況を想定し、1 年あたり 0.05 人/年でランダムな脱落が生じると仮定する。ランダムな脱落時点または追跡期間終了時までにイベントが起こっていない場合、脱落時点と追跡期間終了時のいずれか早い時点で打ち切りとする。

3.2.3　ベイズ解析の実行

本シミュレーション実験では、外部対照データを借用しない no borrowing 法、外部対照データの全てを借

用する full borrowing 法を参考に、CMP 法の動作特性を評価する。共変量を含めたモデルに基づく評価をすることを目的とするが、共変量を利用しないモデルに基づく評価も合わせて実施する。そのため、本シミュレーションで設定するシナリオは HR（0.65 or 1.0）、drift HR（1.0 or 1.2）、解析モデル中の共変量（有り、無し）を掛け合わせた8パターンである。

CMP 法を用いて HR を推定するために MCMC 法を用いる。MCMC 法にて設定する事後分布推定のための条件は以下の通りとする。

- Burn in 期間：1000
- iteration 数：10000
- chain 数：2
- thinning 数：1

3.2.4 シミュレーションの要約指標

CMP 法の動作特性の評価に際して、上記で設定したシナリオに応じて、第一種の過誤確率、検出力、バイアス、MSE 及び EHSS を以下の通り評価する。ただし、それぞれの評価に際してシミュレーション実験回数は1000回とする。

- 第一種の過誤確率、検出力
  - no borrowing 法、full borrowing 法：推定された HR の95%信頼区間の上限が1を下回った回数に基づき評価
  - CMP 法：ハザード比の事後分布の95%信用区間の上限は1を下回った回数基づき評価
- バイアス、MSE
  - no borrowing 法、full borrowing 法：推定された HR（最尤推定値）と新規試験の真の治療群間差（HR）との差に基づき評価
  - CMP 法：HR の事後平均値と新規試験の治療群間差（HR）との差に基づき評価
- EHSS
  - no borrowing 法：計算しない
  - CMP 法、full borrowing 法：治療の有無を表す回帰パラメータβ_{trt}の精度（分散の逆数、以下 Prec）を用いて算出

    ここで EHSS は以下の式により得られる、事後 effective sample size から新規試験の被験者数N_{IC}を差し引く方法を用いる。

$$EHSS = N_{IC}\left(\frac{\mathrm{Prec}(\beta_{\mathrm{trt}}|\text{all data})}{\mathrm{Prec}(\beta_{\mathrm{trt}}|\text{Concurrent trial data})} - 1\right)$$

4. シミュレーション結果

3.で設定したシミュレーションについて、各要約指標の結果を以下に示す。

・第一種の過誤確率

解析モデルの共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
有り	1	1	0.027	0.025	0.029
有り	1	1.2	0.027	0.086	0.102
無し	1	1	0.039	0.053	0.060
無し	1	1.2	0.039	0.105	0.134

・検出力

解析モデルの共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
有り	0.65	1	0.817	0.929	0.929
有り	0.65	1.2	0.817	0.971	0.984
無し	0.65	1	0.482	0.690	0.719
無し	0.65	1.2	0.482	0.79	0.855

・バイアス

解析モデルの共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
有り	1	1	0.013	0.016	0.008
有り	1	1.2	0.013	-0.050	-0.068
有り	0.65	1	0.003	0.005	0.001
有り	0.65	1.2	0.003	-0.037	-0.048
無し	1	1	0.008	-0.012	-0.023
無し	1	1.2	0.008	-0.057	-0.076
無し	0.65	1	0.112	0.096	0.088
無し	0.65	1.2	0.112	0.062	0.048

・MSE

解析モデルの共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法
有り	1	1	0.024	0.034	0.017
有り	1	1.2	0.024	0.033	0.019
有り	0.65	1	0.010	0.014	0.007
有り	0.65	1.2	0.010	0.014	0.008
無し	1	1	0.025	0.031	0.016
無し	1	1.2	0.025	0.033	0.020
無し	0.65	1	0.027	0.027	0.017
無し	0.65	1.2	0.027	0.021	0.010

・EHSS

解析モデルの共変量	HR	drift HR	no borrowing 法	CMP 法	full borrowing 法※
有り	1	1	-	124.2	134.8
有り	1	1.2	-	117.2	139.8
有り	0.65	1	-	119.4	129.2
有り	0.65	1.2	-	111.8	130.5
無し	1	1	-	114.6	153.5
無し	1	1.2	-	110.0	154.8
無し	0.65	1	-	110.7	146.4
無し	0.65	1.2	-	104.9	147.9

※full borrowing については外部対照データをすべて活用した場合におけるベイズ流ワイブル回帰に基づく EHSS を出力している。このため，頻度流の full borrowing の解析方法とは異なる点に注意が必要である。

5. シミュレーション結果の考察

第一種の過誤確率について、drift HR = 1 でありかつ適切に共変量にて調整した場合を除いて，いずれの手法でも名義水準である 2.5%を超える結果であり、no borrowing 法に対する CMP 法や full borrowing 法の増大傾向は顕著であった。特に、Drift HR が 1.2 の場合には、共変量による調整によらず一貫して名義水準を大き

く上回る傾向であった。本傾向は、外部対照データを利用することで、新規併用療法群と外部対照データを組み合わせた対照群の OS に違いが生じたため、治療群間差 HR が 1 の状況にて第一種の過誤確率が名義水準を上回ったと考える。

次に CMP 法や full borrowing 法の検出力について、no borrowing 法と比較して検出力が大きくなった。CMP 法と full borrowing 法の間では大きな検出力の差はないが、外部対照データの利用に伴い情報量が増加して、no borrowing 法と比較して検出力が増加したと考える。特に、drift HR = 1.2 の場合、外部対照データを利用することで治療間差が大きくなるシナリオのため、drift HR = 1 の場合と比較して検出力の増加の程度は大きかった。

バイアスについて、共変量による調整の有無によりその傾向の向きが変わるが、CMP 法のバイアスはおおむね no borrowing 法と full borrowing 法の間に位置することが確認できた。特に drift HR = 1 の場合でかつ共変量による調整を行った場合、full borrowing 法のバイアスが最も 0 に近い結果となるが、drift HR = 1.2 の場合では最も 0 から離れる傾向が確認できた。外部対照データとの drift が存在する場合にはバイアスの程度が上昇することは明らかであり、その傾向が結果にも示されたと考える。

MSE について、drift HR に関わらず、no borrowing 法と full borrowing 法と比較して CMP 法の MSE が最も大きい傾向であった。MSE についても，バイアスと同じく no borrowing 法と full borrowing 法の間に位置することを期待したが，結果として最も MSE が大きかった。本傾向はベイズ流と頻度流に基づく違いと想像され、特に階層的な事前分布を用いることで結果の不安定性が上昇したものと考える。no borrowing 法と full borrowing 法を比較すると、driht HR に関わらず full borrowing 法の MSE 小さくなり、外部対照データを用いることで推定精度の上昇は確認された。

EHSS について、CMP 法おおよそ 110 付近の結果であり、full borrowing 法と比較して小さい値であった。このことから、CMP 法では外部データの全ての情報量を利用していないことが分かる。加えて、共変量の設定に応じて、CMP 法と full borrowing 法の EHSS の差の程度は異なり、共変量を適切に調整することでデータ間の患者背景の違いを含めた drift を考慮した情報借用ができていることが伺える。

6. プログラムの紹介

6.1 R package の紹介（psborrow, psborrow2）

time to event データを対象に CMP 法を使用できる R package として psborrow[9] と psborrow2[10] がある。psborrow という名前は、傾向スコアとベイズ動的借用法を組み合わせたものであり、いずれのパッケージも動的借用と傾向スコア調整/重み付け法を組み合わせるために使用することが可能である。psborrow パッケージが先に開発され、その改訂版として psborrow2 が存在する。psborrow2 は現在も開発中のパッケージであり、状況に応じて psborrow も利用可能と考えたため、それぞれのパッケージの特徴を以下に示す。特に、利用可能なサンプリングプログラムが異なるため、利用者の開発環境に応じて適切なパッケージを選択いただきたい。更新状況を考慮すると、基本的には psborrow2 を利用してシミュレーション実験することで問題無いが、Stan が使用できない場合には psborrow を利用することを提案する。

	psborrow	psborrow2
バージョンと作成年月日	v0.2.1（2023/03/03）	v0.0.3.4 （2024/5/1）
パッケージのサポート	なし	あり

サンプリングプログラム	JAGS	Stan
利用可能な Outcome	・ time to event	・ continuous data ・ categorical data ・ time to event data
傾向スコアを用いた共変量の利用	可	可
傾向スコアモデルに基づくマッチング	不可	可

3.2.1 患者背景（共変量）の設定の実装

create_baseline_object 関数で、治療群、対照群、外部対照群の N 数と共変量の設定を同時に行う。ここで covariate1、covariate2、covariate4 はカテゴリカル変数（バイナリ変数）に変換する必要があるため、transformations の部分にて変換を行う。

```
baseline <- create_baseline_object(
  n_trt_int = 200,
  n_ctrl_int = 100,
  n_ctrl_ext = 100,
  covariates = baseline_covariates(
    name = c("covariate1", "covariate2", "covariate3", "covariate4", "covariate5"),
    means_int = c(    0, 0.5, 0.5, 0, 0),
    means_ext = c(0.7, 0.5, 0.9, 0, 0),
    covariance_int = covariance_matrix(
      diag = c(1, 1.2, 1, 0.7, 0.7),
      upper_tri = c(0.5, 0.7,0.9, 0,0,0, 0,0,0,0.7)
    ),
    covariance_ext = covariance_matrix(
      diag = c(1, 1.2, 1, 0.7, 0.7),
      upper_tri = c(0.5, 0.7,0.9, 0,0,0, 0,0,0,0.7)
    )
  ),
  transformations = list(
    category1 = binary_cutoff("covariate1", int_cutoff = 0.45, ext_cutoff = 0.65),
    category2 = binary_cutoff("covariate2", int_cutoff = 0.55, ext_cutoff = 0.55),
    category4 = binary_cutoff("covariate4", int_cutoff = 0.50, ext_cutoff = 0.50)
  )
)
```

3.2.2 生存時間データの生成の実装

生存時間の従う関数，治験の登録・脱落に関する情報と生存時間・打ち切りの情報を設定する。3.2.1 で作成した患者背景が生存時間に与える程度として各共変量に対する係数を指定し、その後に生存時間の従う関数を指定する。

治験の登録期間、観察期間、脱落情報として下記を仮定し、 パイプ演算子を用いて 1 つのオブジェクト指向の変数内（data_sim）にまとめる。

- 新規試験、外部対照データでは登録期間24ヵ月内に均一に登録される
- Clinical cut off：追跡期間は36ヶ月とし最後の被験者登録から36ヶ月でデータカットオフ
- 脱落における想定：1年あたり0.05人の脱落が発生すると想定

ここで、ランダムな脱落の設定における注意点を明記する。psborrow2 パッケージでは、イベント時間と打ち切り時間を比較した上で、イベントと打ち切りの判定を含めた採用する時間を決定している。しかし、psborrow2 のバージョン v0.0.3.4 では、イベント時間と打ち切り時間の大小関係に応じたイベントと打ち切りの判定及びその時間の取り扱いが逆転している。そのため、本パッケージでのランダムな脱落の設定は困難であり、今回のシミュレーション実験では代案としてほとんど drop する状況（すなわちハザードを大きくする状況）を作り上げるため、設定するハザードを 1 からを引いた値を用いた。

```
#Generating Survival Data setting
data_sim <- create_data_simulation(
  baseline = baseline,
  coefficients = c(category1 = 0.5, category2 = 0.5, covariate3 = 0.5, category4 = 0.5, covariate5 = 0.5),
  event_dist = create_event_dist(dist = "weibull", lambdas = log(2) / (24), gammas = 0.9)
 # Enrollment setting
  ) %>% set_enrollment(
    internal = custom_enrollment(fun = function(n) runif(n, min = 0, max = 24), label = "Uniform enrollment distribution"),
    external = custom_enrollment(fun = function(n) runif(n, min = 0, max = 24), label = "Uniform enrollment distribution")
  # Clinical cut-off setting
  ) %>% set_cut_off(
    internal = cut_off_after_last(time = 36),
    external = cut_off_after_last(time = 36)
  ) %>%
  # Drop out setting
  set_dropout(
    internal_treated = create_event_dist(dist = "exponential", lambdas = 1 - 0.05/12),
    internal_control = create_event_dist(dist = "exponential", lambdas = 1 - 0.05/12),
    external_control = create_event_dist(dist = "exponential", lambdas = 1 - 0.05/12)
)
```

generate 関数を用いて、各被験者の生存時間データをシミュレートする。このとき、評価する治療群間差（HR）とデータ間の差（Drift HR）も設定する。

```
data_list <- generate(data_sim, n = 1000, treatment_hr = c(1, 0.65), drift_hr = c(1, 1.2))
```

3.2.3 ベイズ解析の実行の実装

評価する解析モデルを設定する。構成として、ベースラインモデルの指定、治療効果モデルの指定、共変量モデルの指定、borrowing の方法の指定を行う。ベースラインモデルの指定では、アウトカムの従う分布に応じた関数を用いて、アウトカムデータの特定とアウトカムの従う分布のパラメータの事前分布を指定する。本シミュレーション実験ではワイブル分布に従うと仮定し、形状パラメータの事前分布に指数分布を設定し、尺度パラメータの事前分布に正規分布を設定した。なお、正規分布に渡す引数は平均と標準偏差であることに注意する。

```
outcome <- outcome_surv_weibull_ph(
    time_var = "eventtime",
    cens_var = "cens",
    baseline_prior = prior_normal(0, sqrt(1000)),
    shape_prior = prior_exponential(1)
)
```

次に治療効果モデルの指定を行う。具体的には、治療群を表すダミー変数を特定し、その係数の事前分布を指定する。

```
treatment <- treatment_details(
        trt_flag_col = "trt",
        trt_prior = prior_normal(0, sqrt(1000))
)
```

次に共変量モデルの指定を行うが、基本的には治療効果モデルの指定と同じで、係数の事前分布の指定が主な作業である。

```
    covariate <- sim_covariate_list(
        list(
            "None" = NULL,
            "All"  =  add_covariates(c("category1",  "category2",  "covariate3",  "category4",  "covariate5"),
  prior_normal(0, sqrt(1000)))
        )
)
```

最後に borrowing の方法を指定する。シミュレーション実験では、CMP 法の評価に際して、外部対象データを利用しない no borrowing 法と外部対照データの全てを利用する full borrowing 法との比較を実施する。この 3 つのモデルについて以下の手順にてそれぞれ指定する。CMP 法の設定では、τ の事前分布の指定もこの時点で行う。プログラムでは新規試験と外部対照群のアウトカムの差が小さいと仮定した場合に利用できるガンマ分布をしているが、アウトカムの差が大きいと仮定する場合には、よりロバストな半コーシー分布 prior_half_cauchy(mu = , sigma =)を指定する事ができる。

```
    borrowing_list <- sim_borrowing_list(
      list(
        "No borrowing" = borrowing_none("ext"),
        "CMP" = borrowing_hierarchical_commensurate(ext_flag_col = "ext", prior_gamma(alpha = 1, beta =
  0.001)),
        "Full borrowing" = borrowing_full("ext")
      )
)
```

3.2.3で作成したデータと指定した解析モデルを用いて、Stan で解析できる状態にする。

```
sim_obj <- create_simulation_obj(
    data_matrix_list = data_list,
    outcome = outcome,
    covariate = covariate,
    borrowing = my_borrowing_list,
    treatment = treatment
)
```

作成されたシミュレーション実験用のオブジェクト指向の変数（sim_obj）を mcmc_sample()に渡して、ベイズ推測を実施する。ここで、parallel_chains は rstan にて用意される chain 別に並行で mcmc を実行するための宣言であり、若干の計算の高速化が期待される。

```
results <- mcmc_sample(
    sim_obj,
    posterior_quantiles = c(0, 0.975),
    iter_warmup = 1000,
    iter_sampling = 10000,
    chains = 2,
    parallel_chains = 2,
    seed = 2024
)
```

mcmc_sample()にて推測された結果 results を get_results()に渡す。

7. まとめ

本稿では、DLBCL の臨床試験を例に取り、ベイズ流ワイブル回帰モデルに基づく CMP 法の実装について、その理論とシミュレーションおよび動作特性評価の実装を紹介した。CMP 法は、外部対照データと新規試験データとの類似性をパラメータ化し、その共通性の程度に応じて動的に情報を借用できるという、非常に合理的な枠組みを提供する。一方で、その数理的な解釈には一定の難解さが伴い、実際の臨床試験解析への応用には実装上のハードルも存在する。本稿で取り上げた R パッケージ psborrow および psborrow2 は、臨床試験デザインの設定、共変量の組み込み、ベイズ流解析の実行に加えて、シミュレーションによる動作特

性評価までを比較的容易に実装できる環境を提供している。本稿が、読者にとって効率的な臨床試験計画の立案や、外部データを活用した臨床試験デザインの性能評価を進める上での一助となれば幸いである。

付録

本論文および製薬協報告書では CMP 法の比較的簡便な実装として R パッケージを用いた方法を紹介してきた。本付録では、CMP 法を含むベイズ階層モデルを用いた生存時間データの解析を、SAS の proc mcmc により実装する。簡単のため、本項での解析対象は単一のシミュレーション結果に限られ、モデルの構造（パラメータ指定・事前分布・尤度関数）や full borrowing 法 / no borrowing 法 との比較構成を明示することを主目的とする。

なお本プログラムは、ベイズ推定手順を再現することに重点を置いており、いわゆる統計的性能評価（たとえば、第一種の過誤確率や 検出力、バイアス、EHSS などの要約指標）は出力対象としておらず、検証的なシミュレーション評価を目的とした設計とはなっていない。この点は、本プログラムの limitation として留意されたい。

3.2.1 患者背景（共変量）の設定および 3.2.2 生存時間データの生成の実装

proc IML を用いて、共変量データセットを生成する事ができる。なお、シミュレーションの設定では新規試験と外部対照試験で異なる共変量分布を仮定しているため、シミュレーション条件を統一させるためには、下記プログラム（新規臨床試験を想定した設定）のパラメータを外部対照群の設定に変更し、生成したデータセットを縦結合する必要がある。

```
%let n_trt_int = 200;
%let n_ctrl_int = 100;
%let n_ctrl_ext = 100;
%let shape = 0.9;
%let lambda = %sysevalf( %sysfunc(log(2)) / 24 );  /* 尺度パラメータ */
%macro gen_sim_data(hr, drift);

*************新規臨床試験*****************;
proc iml;
   call randseed(12345);

   /* 共変量生成（新規臨床試験用） */
   mu_int = {0, 0.5, 0.5, 0, 0};
   sigma = {
                  1.0   0.5   0.7   0.0   0.0,
                  0.5   1.2   0.9   0.0   0.0,
                  0.7   0.9   1.0   0.0   0.0,
                  0.0   0.0   0.0   0.7   0.7,
                  0.0   0.0   0.0   0.7   0.7
        };

   n_int = &n_trt_int + &n_ctrl_int;
   X_int = randnormal(n_int, mu_int, sigma);

   /* 2値化 */
   category1 = (X_int[,1] >= 0.45);
   category2 = (X_int[,2] >= 0.55);
   category4 = (X_int[,4] >= 0.50);
   covariate3 = X_int[,3];
```

```
   covariate5 = X_int[,5];

   /* 群割付：最初のn_trt_intが治療群 */
   trt = j(&n_trt_int,1,1) // j(&n_ctrl_int,1,0);
   ext = j(n_int,1,0);

   /* 線形予測子 */
   linpred = 0.5#category1 + 0.5#category2 + 0.5#covariate3 + 0.5#category4 +
0.5#covariate5;
   linpred=linpred + log(&hr)#trt;  /* 外部対照の実装ではlog(&drift)#extとする*/

   /* ワイブル生存時間 */
   U = randfun(n_int, "Uniform");
   T_event = (-log(U) / (&lambda # exp(linpred)))##(1/&shape);

   /* 登録時刻（0〜24） */
   enrollment = randfun(n_int, "Uniform") * 24;

   /* ドロップアウト時間（指数分布） */
   dropout_rate = 0.05 / 12;
   T_drop = randfun(n_int, "Exponential") / dropout_rate;

   /* カットオフ（36ヶ月） */
   cutoff = j(n_int,1,36);

   /* 実観察期間 */
   T_obs = T_event;
   event = j(n_int,1,1);
   do i = 1 to n_int;
      if T_event[i] > T_drop[i] then do; /*ドロップアウトによる打ち切り*/
         T_obs[i] = T_drop[i];
         event[i] = 0;
      end;
      if T_obs[i] > (cutoff[i] - enrollment[i]) then do; /*カットオフによる打ち切り*/
         T_obs[i] = cutoff[i] - enrollment[i];
         event[i] = 0;
      end;
   end;


   /* 出力 */
   final = trt || ext || enrollment || T_obs || event || category1 || category2
|| category4 || covariate3 || covariate5;
   varNames = {"trt" "ext" "enroll" "time" "event" "cat1" "cat2" "cat4" "cov3"
"cov5"};

   create sim_int from final[colname=varNames];
   append from final;
   close sim_int;
quit;

%mend gen_sim_data;

%gen_sim_data(hr=0.65, drift=1.2);
%gen_sim_data(hr=0.65, drift=1.0);
%gen_sim_data(hr=1.0, drift=1.2);
%gen_sim_data(hr=1.0, drift=1.0);
```

3.2.3 ベイズ解析の実行の実装

proc mcmc を用いてベイズ流解析を実行する。2 章で説明した各パラメータの事前分布や超事前分布、回帰モデル式、尤度関数を宣言する。なおベイズ解析を用いて full borrowing を実装する際には beta0 や tau を削除し、全観察データを含める。no borrowing を実装する際には beta_ext や tau を削除し、beta0 の事前分布に正規分布 `normal(0, prec = 0.001)` を直接指定する。プログラムの詳細は吉本ら（2025）[11] の論文も参照のこと。

```
data monitor;
  set final;
  cnsr = 1 - event; /* MCMC用の打ち切りフラグ */
run;

ods output PostSummaries = post_summary;    /* 事後要約統計量 */
ods output PostIntervals = post_intervals;  /* 95%信用区間 */

proc mcmc data = monitor outpost = posterior_out nbi = 3000 nmc = 30000 thin = 3
seed = 1234;

  /* パラメータ宣言 */
  parms beta_ext 0;        /* 外部データベースライン */
  parms tau 1;             /* commensurability 精度 */
  parms beta_trt 0;        /* 治療効果 */
  parms beta0 0;           /* 内部対照のベースライン */
  parms beta1-beta5 0;     /* 共変量効果：cat1, cat2, cov3, cat4, cov5 */
  parms r 1;               /* shape パラメータ */

  /* 事前分布 */
  hyperprior beta_ext ~ normal(0, prec = 0.001);
  hyperprior tau ~ gamma(1, is = 0.001);
  prior beta_trt ~ normal(0, prec = 0.001);
  prior beta0 ~ normal(beta_ext, prec = tau);
  prior beta1-beta5 ~ normal(0, prec = 0.001);
  prior r ~ expon(is = 1);

  /* 線形予測子 */
  linpred = beta0 * (1 - ext) + beta_ext * ext + beta_trt * trt
          + beta1 * cat1 + beta2 * cat2 + beta3 * cov3 + beta4 * cat4 + beta5 *
cov5;

  lambda = (1 / exp(linpred)) ** (1 / r);

  llike = (1 - cnsr) * logpdf("weibull", time, r, lambda)
       + cnsr * logsdf("weibull", time, r, lambda);

  model general(llike);
run;

ods output close;
```

参考文献

1. 厚生労働省. (2021). 承認申請等におけるレジストリの活用に関する基本的考え方. https://www.mhlw.go.jp/web/t_doc?dataId=00tc5763&dataType=1&pageNo=1
2. 厚生労働省. (2022). 「臨床試験の一般指針」の改正について. 医薬審第 1223 第 5 号.

https://www.pmda.go.jp/files/000250244.pdf
3. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. "Rare disease の治療効果の推測法." （2022）
4. 厚生労働省. (2001). 「臨床試験における対照群の選択とそれに関連する諸問題」について. 医薬審第136号. https://www.pmda.go.jp/files/000156634.pdf
5. 野村尚吾, 大東智洋, 澤本涼.（2022）. Hybrid control アプローチを用いるランダム化比較試験の計画と解析：外部データが要約統計量の場合. 計量生物学43(1), 63-96.
6. Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. Journal of Chronic Diseases 29, 175-188.
7. 武田健太朗, 大庭真梨, 橋爪智行, 坂巻顕太郎, 田栗正隆, 森田智視（2015）臨床試験におけるヒストリカルコントロールデータの利用. 計量生物学会36(1), 25-50
8. Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., & Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047-1056.
9. Gower-Page, C., Lu, Y., Lin, A. (2023). _psborrow: Bayesian Dynamic Borrowing with Propensity Score_. R package version 0.2.1.
10. Secrest M and Gravestock I (2024). _psborrow2: Bayesian Dynamic Borrowing Analysis and Simulation. https://github.com/Genentech/psborrow2, https://genentech.github.io/psborrow2/index.html.
11. 吉本拓矢, 田中優一, 高戸俊一（2025）MCMC プロシジャによる commensurate prior 法を用いた外部対照データの動的借用法. SAS ユーザー総会2025

ワクチン試験での必要イベント数の設定について

○浦狩保則[1]

([1]シミック（株）データサイエンス事業本部）

On Determination of Required Number of Events in Vaccine Study

Yasunori,Uragari
Data Science Div., CMIC Co.,Ltd.

要旨

本稿では、ワクチン群とプラセボ群の 2 群比較試験の計画において、2 群のイベント発現数合計に条件付きでワクチン群のイベント発現数が従う二項分布に基づいて、ワクチン有効性についての仮説検定で必要なイベント発現数を算出する方法（条件付き法）に関する検討結果を記述した。本稿では、条件付き法の適用が妥当であるための条件を、両群の試験期間中の発現割合が 0.3 以下であることと設定した。必要被験者数ではなく、必要イベント数に注目したのは、イベント数による試験の進捗管理が行いやすいためである。そのことを、イベント発現数合計についての条件付けを行わず必要被験者数を算出する方法との比較に基づいて示した。さらに、ワクチン有効性の推定において、Senn(2022)によるケース・プロポーションの 95%信頼区間算出法が優れた性質をもつことを示したうえで、ワクチン有効性の推定精度の観点から必要イベント数を算出するための一案を提示した。

キーワード：ワクチン試験、必要イベント数、ワクチン有効性、リスク比、ケース・プロポーション

1. ワクチン試験の特徴

ワクチン群とプラセボ群のイベント（例えば、感染、発症）の発現割合（あるいは発現率）の比較試験を考える。最初に、そのようなワクチン試験について、3 つの特徴を踏まえておきたい。第 1 の特徴は、有効性の評価指標として、ワクチン有効性が用いられることである。ワクチン有効性は、1-リスク比（=1-相対リスク）であり、リスク比は、(ワクチン群のイベント発現割合)/(プラセボ群のイベント発現割合)で定義する。イベント発現割合の大小に関わらず、イベント発現割合の比によって、リスク比、そしてワクチン有効性が決定されることに注意したい。そして、リスク比が 1 より小さな値であるほど、ワクチン有効性が高くなる。第 2 の特徴は、ワクチンの有効性の証明にあたって、対プラセボ優越性を示すこと、具体的にワクチン有効性=0 の帰無仮説を棄却し、ワクチン有効性>0 の対立仮説を採択すること（あるいは、リスク比=1 の帰無仮説を棄却し、リスク比<1 の対立仮説を採択すること）だけでは不十分であると見なされるのが通常になっていることである。第 3 の特徴は、ワクチン試験の検出力に対し、両群で発現するイベント数が直接的な影響力をもつことである。本稿のタイトルの通り、必要イベント数を設定すること、および両群で生起

しているイベント総数を用いて試験の進捗管理を行うことが有用である(Senn,2022)。イベント発現割合（あるいは発現率）の大小が影響するのは、必要被験者数あるいは 1 被験者あたりの観察期間である。イベント発現割合（あるいは発現率）が小さいほど、より多くの被験者あるいはより長期の観察が必要となる。

本稿の以降の構成は次の通りである。2 節では、ワクチン試験における必要イベント数の設定法として、両群の発現総イベント数に条件付きで、ワクチン群の発現イベント数が従う条件付き二項分布に基づく方法（条件付き法と略す）に着目するとともに、その方法の適用が妥当であるための条件について述べる。3 節では、試験の進捗管理において、条件付き法により算出される必要総イベント数を用いる方法が、必要総被験者数を用いる方法よりも優れたものであることを示す。4 節では、ワクチン有効性の推定方法に目を向け、ワクチン有効性の推定精度の観点からの必要イベント数の設定法の一案を提示する。最後の 5 節で、上記事項のまとめを示すとともに、関連するいくつかの事項について触れる。

2. ワクチン有効性についての仮説検定と必要イベント数

2.1 ワクチン試験のデザイン時の諸特性についての規定

ワクチン群をv、プラセボ群をpで表し、両群への被験者数の割付け比率は 1 対 1 であるとし、そのもとでの両群の被験者数を$n_v = n_p = n$で表す。これらの群の発現率（件/人年）をそれぞれΛ_v、Λ_pで表す。試験の追跡期間がD（年）の場合、試験期間中の発現割合はワクチン群、プラセボ群でそれぞれ$P_v = \Lambda_v D$、$P_p = \Lambda_p D$と表せる。（ここでは、すべての被験者について追跡期間Dの全体にわたってイベントが観測されること、追跡期間中にイベント未発現の状態で試験から脱落することがないこと、各被験者におけるイベント発現数は 0 回または 1 回であることを仮定している。）

このとき、両群のイベント発現数Y_v、Y_pの確率分布はそれぞれ二項分布$Y_v \sim Bin(n, P_v)$、$Y_p \sim Bin(n, P_p)$で表される。nが大きく、P_v、P_pが適度に小さいとき、Y_v、Y_pの確率分布が Poisson 分布$Po(\lambda_v)$、$Po(\lambda_p)$で近似されることがよく知られている。ここで、$\lambda_v = nP_v$、$\lambda_p = nP_p$である。

今、リスク比 (あるいは相対リスク)を $RR = P_v/P_p$で表すと、ワクチン有効性（Vaccine Effect）は$\pi = 1 - RR$で与えられる。これらと密接な関係をもつ指標として、ケース・プロポーション

$$\theta = \frac{nP_v}{nP_v + nP_p} = \frac{\lambda_v}{\lambda_v + \lambda_p} = \frac{1-\pi}{2-\pi} \qquad (1)$$

がある(Chan & Bohider,1998 ; Loiacono,2018)。この割合は、(ワクチン群の期待イベント数)/(両群の期待総イベント数)であり、このことがθをケース・プロポーションと呼ぶ理由になっている(Senn,2022)。(1)から$\pi = (2\theta - 1)/(\theta - 1)$の関係が成立する。

今、両群のイベント数がそれぞれ独立に Poisson 分布に従う、すなわち$Y_v \sim Po(\lambda_v), Y_p \sim Po(\lambda_p)$である場合、それらの同時分布は、Poisson 分布の再生性を通して

$$p(Y_v, Y_p | \lambda, \theta) = \frac{\lambda^T e^{-\lambda}}{T!} \times \frac{T!}{Y_v!(T-Y_v)!} \theta^{Y_v} (1-\theta)^{T-Y_v} \qquad (2)$$

で示される(Przyborowski & Wilenski,1940)。ここで、$T = Y_v + Y_p$は両群の総イベント数、および$\lambda = \lambda_v + \lambda_p$である。(2)式右辺の第 1 因数は、パラメータ$\lambda$の Poisson 分布のもとで、$T$が起こる確率であり、第 2 因数は$T$を与えたもとでのワクチン群のイベント数$Y_v$の確率分布を表し、それがパラメータ$\theta$の二項分布であることがわかる。すなわち、$T$に条件付きで、$Y_v$の確率分布は

$$Y_v | T \sim Bin(T, \theta) = \binom{T}{k} \theta^k (1-\theta)^{T-k} \qquad (3)$$

で表される(Przyborowski & Wilenski,1940; Chan & Bohidar, 1998; Senn,2022)。前節で述べた条件付き法は、(3)式の条件付き二項分布に基づく方法を指している。

2.2 ワクチン有効性の仮説検定で必要なイベント数

ワクチン有効性についての検定仮説

$$H_0{:}\pi \le \pi_0 \ \text{ vs } \ H_1{:}\pi > \pi_0$$

を考える（π_0は 0 とは限らない）。これは、ケース・プロポーションについての検定仮説

$$H_0{:}\theta \ge \theta_0 \ \text{ vs } \ H_1{:}\theta < \theta_0 \ \text{、ただし } \theta_0 = (1-\pi_0)/(2-\pi_0)$$

で置き換えることができる。そして、両群の総イベント数Tおよびワクチン群のイベント数$y_{v,obs}$が得られた場合、検定の p 値は(3)式を用いて

$$p = \Pr\left(Y_v \le y_{v,obs} \middle| \theta_0\right) = \sum_{k=0}^{Y_{v,obs}} \binom{T}{k} \theta_0^k (1-\theta_0)^{T-k} \qquad (4)$$

で算出される。検定の片側有意水準をαで表すと、検定の棄却限界値Y_cは

$$Y_c = \ Max_Y[\Pr(Y_v \le Y|\theta_0) < \alpha] \qquad (5)$$

となり、このときの実際の α レベルは

$$\Pr(Y_v \le Y_c|\theta_0) = \sum_{k=0}^{Y_c} \binom{T}{k} \theta_0^k (1-\theta_0)^{T-k} \qquad (6)$$

となる。ワクチン有効性についての期待値（対立仮説の期待値）をπ_1とすると、それに対応するケース・プロポーション値は$\theta_1 = (1-\pi_1)/(2-\pi_1)$であり、検定の検出力は

$$1-\beta = \Pr(Y_v \le Y_c|\theta_1) = \sum_{k=0}^{Y_c} \binom{T}{k} \theta_1^k (1-\theta_1)^{T-k} \qquad (7)$$

で評価することができる。

検定仮説を、ワクチン有効性（π_1:期待値、π_0：帰無仮説値）と対応するケース・プロポーション（θ_1:期待値、θ_0：帰無仮説値）で記述したもとで、片側有意水準 0.025 で 0.9 以上の検出力を得るために必要な総イベント数（T_{req}）と棄却限界値（Y_c *:* 有意差が得られるワクチン群のイベント数の最大値）を表 1 に示す。

表 1.ワクチン有効性の仮説検定ための必要総イベント数T_{req}と棄却限界値Y_c

（片側有意水準:0.025、必要な検出力:0.90）

ワクチン有効性による記述		ケース・プロポーションによる記述		必要総イベント数：T_{req}	棄却限界値：Y_c	α レベル：(6)式	検出力：(7)式
π_1	π_0	θ_1	θ_0				
0.1	0.0	0.474	0.500	3,845	1,861	0.025	0.903
0.2	0.0	0.444	0.500	880	410	0.023	0.906
	0.1	0.444	0.474	3,115	1,420	0.024	0.903
0.3	0.0	0.412	0.500	350	156	0.024	0.910
	0.1	0.412	0.474	700	305	0.024	0.907
	0.2	0.412	0.444	2,457	1,043	0.024	0.904
0.4	0.0	0.375	0.500	178	75	0.021	0.911
	0.1	0.375	0.474	279	115	0.023	0.910
	0.2	0.375	0.444	548	220	0.023	0.907
	0.3	0.375	0.412	1,901	740	0.024	0.904
0.5	0.0	0.333	0.500	99	39	0.022	0.916
	0.1	0.333	0.474	138	53	0.021	0.911
	0.2	0.333	0.444	216	81	0.023	0.914
	0.3	0.333	0.412	419	152	0.023	0.908
	0.4	0.333	0.375	1,431	500	0.024	0.906

＜必要イベント数算出時の留意点＞

必要イベント数の算出場面では、両群の総イベント数Tを、例えば 1 から 5000 まで 1 刻みで動かし、(6)の実際のレベルが α を超えず、(7)の検出力が目標検出力を上回るように、Tとそのもとでの棄却限界値Y_cを見出すことになる。ただし、そのような総イベント数Tを見出すことができたとしても、それ以上の総イベント数のもとで検出力が目標値を下回ることが起こり得る。例えば、表 1 において$\pi_1 = 0.4, \pi_0 = 0.0$の場合について、

α レベルは常に名目有意水準以下に保たれる。しかし、検出力は、$T = 169$のとき初めて 0.9 を超えるが、それ以上のいくつかのTの値（170、172、177）で 0.9 を下回り、$T = 178$以上で常に 0.9 を超えることが観察される。これは、二項分布の離散的性質（discrete nature）によるものである。そのため、十分に大きい総イベント数のもとでの検出力も考慮して、必要総イベント数を設定するべきである。

＜必要被験者数＞

必要総イベント数T_{req}が算出されれば、当該試験における必要被験者数の算出へ移る。その際、試験期間D（年）、ワクチン群およびプラセボ群の期待イベント発現率Λ_v、Λ_p（件/人年）を用いることが通常である。(なお、期待イベント発現率の単位を件/人年としているが、同一被験者において初発のイベントに注目し、初発イベントの発現までの期間で人年をカウントする場合は、期待イベント発現率の単位は人/人年となり、すなわち期待イベント発現率は試験期間当たりのイベント発現割合と見なすことができる。) また、期待発現率の比がリスク比に等しいこと、すなわちΛ_v/Λ_p=RRであることを仮定する。このとき、Λ_v=$\Lambda_p RR = \Lambda_p(1-\pi_1)$である。$D$(年)の試験期間中に、2 群合計で$T_{req}$件のイベントを観察するための 1 群あたりの必要被験者N_{req}は

$$N_{req} = \frac{T_{req}}{\{D(\Lambda_p+\Lambda_v)\}} = \frac{T_{req}}{\{D(2-\pi_1)\Lambda_p\}} \qquad (8)$$

で算出される。割付比が 1 対 1 であるので、必要被験者数は 2 群合計で$2N_{req}$となる。

表 1 で示した各ケースについて、プラセボ群の期待イベント発現率をΛ_p=0.2 件/人年と想定した場合で、D=0.5 年、D=1 年とする場合の必要被験者数（2 群合計）を表 2 に示す。

表 2．表 1 の各ケースの必要被験者数:プラセボ群の期待イベント発現率をΛ_p=0.2 件/人年と想定

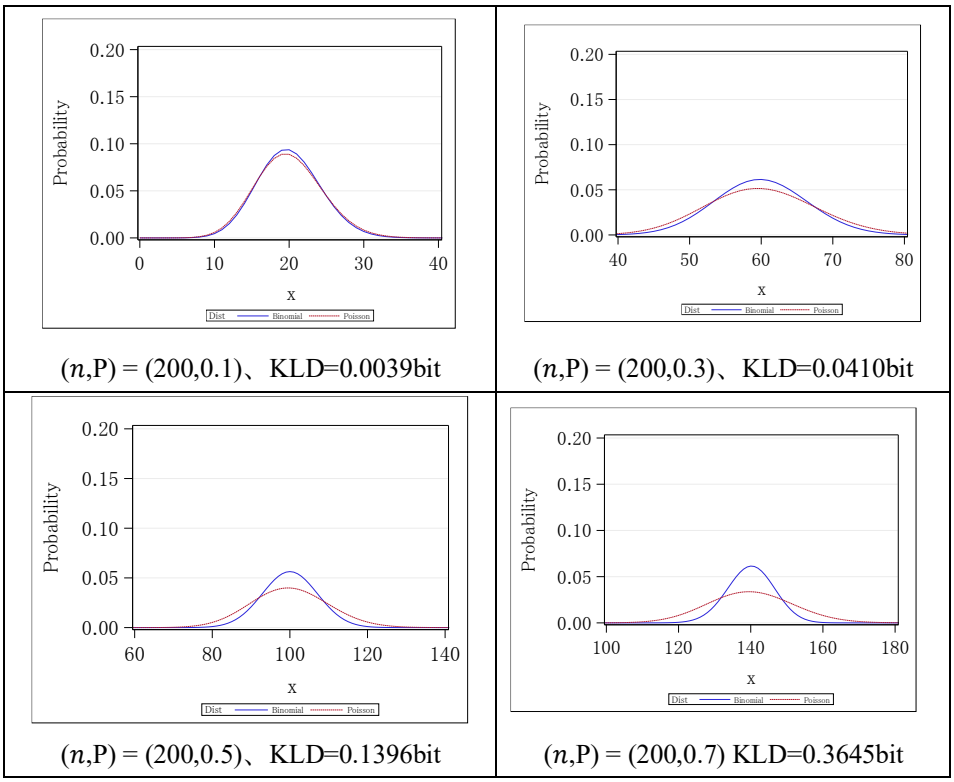
ワクチン有効性		プラセボ群	ワクチン群	必要被験者数（2 群合計）	
π_1	π_0	発現率（件/人年）	発現率（件/人年）	試験期間 0.5 年	試験期間 1 年
0.1	0.0	0.2	0.18	40,474	20,238
0.2	0.0	0.2	0.16	9,778	4,890
	0.1	0.2	0.16	34,612	17,306
0.3	0.0	0.2	0.14	4,118	2,060
	0.1	0.2	0.14	8,236	4,118
	0.2	0.2	0.14	28,906	14,454
0.4	0.0	0.2	0.12	2,226	1,114
	0.1	0.2	0.12	3,488	1,744
	0.2	0.2	0.12	6,850	3,426
	0.3	0.2	0.12	23,764	11,882
0.5	0.0	0.2	0.1	1,320	660
	0.1	0.2	0.1	1,840	920
	0.2	0.2	0.1	2,880	1,440
	0.3	0.2	0.1	5,588	2,794
	0.4	0.2	0.1	19,080	9,540

試験期間を長く設定すれば、必要イベント数を達成するために必要な被験者数は少なくなることが伺える。

2.3 条件付き法の適用条件

条件付き法は(2)式に基づいているが、(2)式が妥当であるためには、両群のイベント発現数Y_v、Y_pの確率分布が Poisson 分布で十分に近似できることが前提であった。2.1 節では、Poisson 近似のための条件を「nが大きく、P_v、P_pが適度に小さい」と記載したが、ここではさらに詳しく調べることとした。具体的に、被験者数nを 20~500 まで（20 刻み）の範囲、イベント発現割合Pを 0.1~0.8 まで（0.1 刻み）の範囲、で動かしたときの各組合せで、イベント発現数Yについての二項分布上の理論確率と Poisson 分布上の理論確率の相違を、カルバック・ライブラー距離（Kullback-Leibler Distance）

KLD=$\sum_{x=0}^{n}[p(x|n,P)\log_2\{p(x|n,P)/q(x|nP)\}]$,　$p(x|n,P)$：二項確率、$q(x|nP)$：Poisson 確率で評価した(杉山,2013)。(n,P) = (200,0.1)、(200,0.3)、(200,0.5)、(200,0.7)の各場合における二項分布と Poisson 分布を重ねて描写するとともに、KLD 値(単位は bit)を図 1 に示した。また、0.1~0.8 までPのそれぞれで、nの変化に伴う KLD 値の挙動を図 2 に示した。図 2 より、二項分布と Poisson 分布の相違を KLD 値で計量する場合、KLD 値に大きく影響するのはPの値であり、nの大小の影響は小さいことが見て取れる。KLD 値がどの程度に小さい値であれば、二項分布への Poisson 分布近似による情報損失が十分に小さいかという明確な基準は存在しないと思われる。ただ、図 1 において$P = 0.3$の場合に KLD 値が 0.05bit 未満で、Poisson 近似がほぼ有効であると見なすことができるであろう。したがって、**条件付き法の適用が妥当であるための条件として、イベント発現割合が 0.3 以下であることを提案したい**。



(n,P) = (200,0.1)、KLD=0.0039bit　　(n,P) = (200,0.3)、KLD=0.0410bit

(n,P) = (200,0.5)、KLD=0.1396bit　　(n,P) = (200,0.7) KLD=0.3645bit

図 1.　二項分布と Poisson 分布および両者間の KLD 値

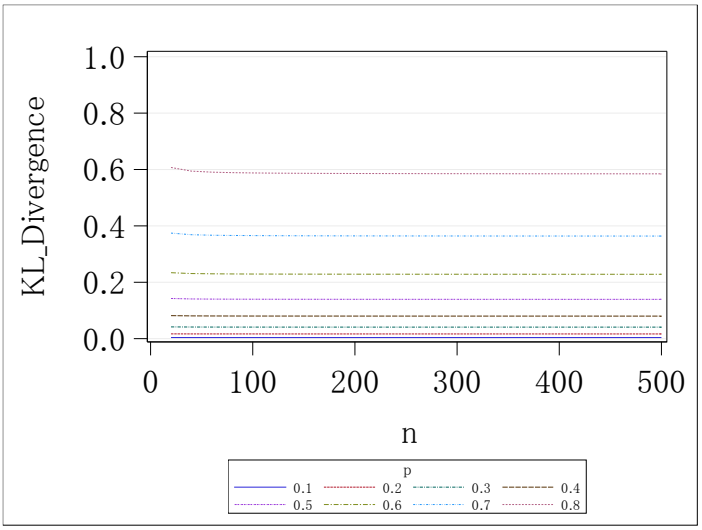


図 2.　nの変化に伴う KLD 値の挙動：P＝0.1~0.8

3. 条件付き法と無条件法の比較

前節で述べた条件付き法は、両群の総イベント数Tへの条件付けのもとでワクチン群のイベント発現数が従う条件付き二項分布に基づいて、必要イベント数を算出するものであった。一方、そのような条件付けを行わず、必要被験者数を算出する方法も多く用いられている。このように、総イベント数Tについての条件付けを行わず、（必要イベント数ではなく）必要被験者数を算出する方法を「無条件法」と呼ぶことにする。ここでは、無条件法の一例として、Farrington & Manning(1990)のスコア検定に基づく、リスク比についての仮説検定のための必要標本サイズの算出法をとりあげる。この方法は、SAS PROC POWER(V15.1)を用いて、以下のコードにより実行でき、使用にあたってはリスク比についての帰無仮説値、対立仮説の期待値、およびプラセボ群のイベント発現割合の想定値を指定することが必要である。

```
proc power;
    twosamplefreq test=fm_rr
        nullrelativerisk = 0.9     /* リスク比についての帰無仮説値：RR0 */
        relativerisk = 0.6         /* リスク比についての対立仮説の期待値：RR1 */
        refproportion = 0.05       /* プラセボ群のイベント発現割合の想定値：Pp */
        sides = L ntotal = . alpha=0.025 power = 0.9;
run;
```

ここでは、リスク比について、帰無仮説値がRR_0=0.9、対立仮説の期待値がRR_1=0.6 である仮説検定で、2 群の発現割合が、リスク比についての対立仮説の期待値を満たす 3 通りのケースを取り上げ、条件付き法で①必要総イベント数、無条件法で③必要総被験者数を算出した結果を表 3 に示した。この表では、条件付き法による②期待必要総被験者数(①必要総イベント数から算出)、無条件法による④期待総イベント数(③必要総被験者数から算出)も示している。ここで、「期待」の語を冠している理由は、2 群の発現割合が正しく想定されていることを仮定したもとでの期待値であるためである。表 3 から、条件付き法による②期待総被験者数が無条件法による③必要総被験者数よりも大きな例数で算出されることがわかる。このことは、Loiacono (2018)の Discussion での記述、すなわち「条件付き法で算出される必要標本サイズが大きなものとなり易い」と一致している。

表 3. 条件付き法による①必要総イベント数、無条件法による③必要総被験者の算出結果

		帰無仮説値	対立仮説の期待値	条件付き法		無条件法	
				①必要総イベント数	②期待必要総被験者数	③必要総被験者数	④期待総イベント数
リスク比 (ワクチン有効性)		RR_0=0.9 (π_0=0.1)	RR_1=0.6 (π_1=0.4)	279			
発現割合の想定値	ケース 1 ワクチン群 プラセボ群	P_{v0}=0.045 P_p=0.050	P_{v1}=0.030 P_p=0.050	279	6,975	6,402	257
	ケース 2 ワクチン群 プラセボ群	P_{v0}=0.0225 P_p=0.025	P_{v1}=0.015 P_p=0.025	279	13,950	13,070	262
	ケース 3 ワクチン群 プラセボ群	P_{v0}=0.009 P_p=0.010	P_{v1}=0.006 P_p=0.010	279	34,875	33,078	265

ところで、より重要なことは、**条件付き法では①必要総イベント数が、リスク比についての帰無仮説値と対立仮説の期待値によって一意に算出され、3 通りのケースのすべてで 279 件となっていることである。**一方、無条件法による③必要総被験者数の算出では、リスク比についての帰無仮説値、対立仮説の期待値に加えて、プラセボ群の発現割合を指定することが必要であり、その発現割合の想定値が実際の値から乖離すれば、得られる③必要総被験者数が大きく変動するという難点が伴う。具体的に、3 通りのケースで③必要総被験者数が大きく相違することがわかる。同様の不確定性は、条件付き法による②期待総被験者数、無条件法による④期待総イベント数の算出においても起こり得る。

以上のことから、**必要総イベント数は、プラセボ群の発現割合の想定値の妥当性に左右されないという意味で確定的な目標値となり得る。そして、試験実施中にそれまでの時点で起こった総イベント数の、目標値に対する到達度によって試験の進捗管理を行えることが、条件付き法で必要総イベント数を算出することの意義である**といえる。

このことを表す例示として、2 つの状況についてのシミュレーション結果を示す。そこでは、表 3 におけるリスク比についての対立仮説の期待値の通りのデータが得られる状況を設定した。

1 つ目は、表 3 におけるケース 1 での発現割合の想定値（すなわちワクチン群の発現割合がP_{v1}=0.030、プラセボ群の発現割合がP_p=0.050）の通りの状況である。リスク比は 0.6 であり、条件付き法での必要総イベント数は 279 件である。このもとで 2 群合計 6,976 例（1 群あたり 3,488 例）について、イベントの有無の二項乱数を生成し、得られたデータについて、リスク比についての帰無仮説$RR \geq 0.9$、対立仮説$RR<0.9$ の片側有意水準 0.025 の検定を、Farrington & Manning(1990)のスコア検定で実施した。このことを 100,000 回反復したときの 2 群のイベント数合計の割合のヒストグラムと、検定結果が有意であった割合、すなわち検出力の折れ線グラフを図 3-a に示した。なお、グラフの横軸の 1 つの級の幅を 10 としている。例えば、総イベント数T_{class}が 270 の級は 265<総イベント数≤275 である。この図から、総イベント数が 270 以上のT_{class}に含まれる割合が大きく（正確には 80.0%）、その場合に 90%以上の検出力が得られていることがわかる。

2 つ目は、両群の発現割合がケース 1 における想定値よりも小さい状況であり、ワクチン群の発現割合がP_{v1}=0.024、プラセボ群の発現割合がP_p=0.040 である。ただし、リスク比は 0.6 で上記と同じとした。そして、図 3-a の場合と同じ 2 群合計 6,976 例（1 群あたり 3,488 例）のデータを生成した。その結果を図 3-b に示す。この図から、総イベント数が 270 以上のT_{class}に含まれる割合が相当に小さいこと（正確には 1.95%）、その場合でも 90%以上の検出力が得られていることがわかる。

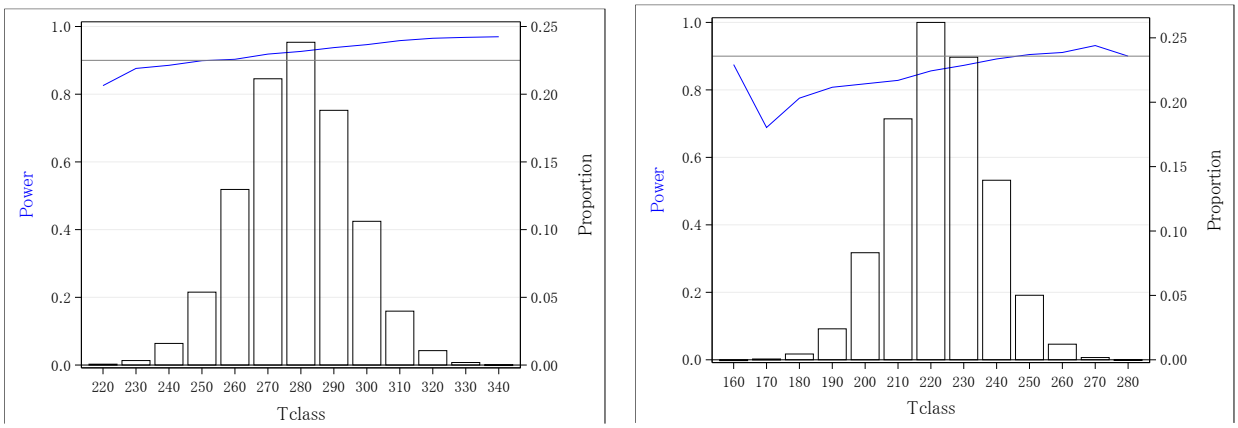


図 3. イベント数合計のヒストグラムとリスク比の検定の検出力

a. (左図)両群の発現割合が想定通りの場合：P_{v1}=0.030、P_p=0.050、リスク比＝0.6

b. (右図)両群の発現割合が想定より小さい場合：P_{v1}=0.024、P_p=0.040、リスク比＝0.6

以上より、目標総イベント数（上記シミュレーションでは 279）を目標値として試験の進捗管理を行うことによって、2 群のリスク比に関する仮説検定で目標検出力の達成可能性を高い確度で予測することが可能であることがうかがえる。

ただし、2 群の実際のリスク比が対立仮説の期待値とは相違する場合には、目標総イベント数を頼りにする試験の進捗管理は不確かなものになることには注意が要る。図 3-b で示した 2 群の発現割合が想定よりも小さい場合も含め、開鍵を伴う中間解析によって 2 群の実際の発現割合を確認し、目標総イベント数の見直しを含めた試験のアダプテーションを念頭に置くことが必要であろう。

4. ワクチン有効性の推定の観点からの必要イベント数の決定

4.1 ワクチン有効性の推定方法

2 群の総イベント数Tのうち、ワクチン群のイベント数Y_vの観測値が$y_{v,obs}$であるとき、ケース・プロポーションθの推定値は$\hat{\theta} = y_{v,obs}/T$であるが、Senn(2022)は$\theta$の 95%信頼区間を次のように与えている。

95%信頼下限値θ_{LCL}：$\Pr(Y_v \le y_{v,obs}|\theta)$ =0.025を満たすθ

95%信頼上限値θ_{UCL}：$\Pr(Y_v \ge y_{v,obs}|\theta)$ =0.025を満たすθ

この信頼区間を数値的に算出するには、θ = 0.0001,0.0002,…,0.9999 (0.0001 刻み)で、二項変数Y_v~Bin(T, θ)を用いて上記の確率を評価すればよい。$\hat{\theta}$、θ_{LCL}およびθ_{UCL}のそれぞれからワクチン有効性πの推定値$\hat{\pi}$、上限値π_{UCL}および下限値π_{LCL}を次のように算出する。

$$\hat{\pi} = (2\hat{\theta} - 1)/(\hat{\theta} - 1)$$
$$\pi_{UCL} = (2\theta_{LCL} - 1)/(\theta_{LCL} - 1)$$
$$\pi_{LCL} = (2\theta_{UCL} - 1)/(\theta_{UCL} - 1)$$

表 4. 総イベント数が$T = 3000$の場合のワクチン群のイベント数$y_{v,obs}$の各値に対応するケース・プロポーションおよびワクチン有効性の 95%信頼区間

$y_{v,obs}$	$\hat{\theta}$	θ_{LCL}	θ_{UCL}	$\hat{\pi}$	π_{LCL}	π_{UCL}	$H_0:\pi \le \pi_0$に対する p 値		
							$\pi_0 = 0$	$\pi_0 = 0.1$	$\pi_0 = 0.2$
1,278	0.4260	0.4082	0.4439	0.2578	0.2018	0.3102	0.0000	0.0000	0.0218
1,279	0.4263	0.4085	0.4443	0.2568	0.2005	0.3094	0.0000	0.0000	0.0238
1,280	0.4267	0.4089	0.4446	0.2558	0.1995	0.3082	0.0000	0.0000	0.0260
:	:	:	:	:	:	:	:	:	:
1,365	0.4550	0.4371	0.4730	0.1651	0.1025	0.2235	0.0000	0.0211	0.8813
1,366	0.4553	0.4374	0.4734	0.1640	0.1010	0.2225	0.0000	0.0230	0.8884
1,367	0.4557	0.4377	0.4737	0.1629	0.0999	0.2216	0.0000	0.0250	0.8953
:	:	:	:	:	:	:	:	:	:
1,444	0.4813	0.4633	0.4994	0.0720	0.0024	0.1368	0.0213	0.8044	1.0000
1,445	0.4817	0.4636	0.4997	0.0707	0.0012	0.1357	0.0233	0.8143	1.0000
1,446	0.4820	0.4640	0.5001	0.0695	-0.0004	0.1343	0.0254	0.8240	1.0000
:	:	:	:	:	:	:	:	:	:

$y_{v,obs}$:ワクチン群イベント数、$\hat{\theta}, \theta_{LCL}, \theta_{UCL}$:ケース・プロポーションの推定値、95%信頼下限値、上限値、$\hat{\pi}, \pi_{LCL}, \pi_{UCL}$: ワクチン有効性の推定値、95%信頼下限値、上限値

例えば、総イベント数が$T = 3{,}000$の場合のワクチン群のイベント数$y_{v,obs}$の各値に対応するケース・プロポーションおよびワクチン有効性の 95%信頼区間を表 4 に示す。この表には、ワクチン有効性に関する仮説$H_0:\pi \le \pi_0$ vs $H_1:\pi > \pi_0$について、$\pi_0 = 0$、$\pi_0 = 0.1$および$\pi_0 = 0.2$の 3 通りの検定における p 値も示している。検定の p 値が片側有意水準であるα(=0.025)未満かα以上かの識別と、ワクチン有効性の 95%信頼下限値

のπ_{LCL}が帰無仮説値であるπ_0を超えるかπ_0以下かの識別とが、寸分違わず一致していることを確認できる。したがって、Senn(2022)が提示したケース・プロポーションの 95%信頼区間は、信頼区間が具備すべき要件を備えていると判断できる。

4.2 ワクチン有効性の推定精度に基づく必要イベント数の設定方法の一案

ここでは、所定のワクチン有効性が期待される状況で、**ワクチン有効性の 95%信頼区間長を所定の限界値未満とするのに必要な総イベント数の算出**を考える。ワクチン有効性の期待値をπ_1とすると、対応するケース・プロポーションは、$\theta_1 = (1-\pi_1)/(2-\pi_1)$である。総イベント数$T$のもとで、ワクチン群のイベント数$Y_v$が二項分布 Bin$(T,\theta_1)$に従って変動し、それに伴ってワクチン有効性の 95%信頼区間も変動するため、その信頼区間を$[\pi_{LCL}(Y_v|T,\theta_1), \pi_{UCL}(Y_v|T,\theta_1)]$で表す。このとき 95%信頼区間の半分幅の期待値が目標値(hとする)未満であること、すなわち

$$\{E[\pi_{UCL}(T,\theta_1)] - E[\pi_{LCL}(T,\theta_1)]\}/2 < h \qquad (9)$$

となるTの最小値を必要な総イベント数T^*とする。T^*の算出手順は次の通りである。

① Tの初期値をT_0とする。

② 起こり得るY_vの値y_v (0,1,…, T)のそれぞれで、ケース・プロポーションの 95%信頼区間の下限値$\theta_{LCL}(y_v)$と上限値$\theta_{UCL}(y_v)$を算出する。前項の Senn(2022)の方法を用いる。

③ ケース・プロポーションの 95%信頼区間の下限値$\theta_{LCL}(Y_v)$と上限値$\theta_{UCL}(Y_v)$の各期待値を算出する。

・$E[\theta_{LCL}|T,\theta_1] = \sum_{y_v=0}^{T} \theta_{LCL}(y_v)\binom{T}{y_v}{\theta_1}^{y_v}(1-\theta_1)^{T-y_v}$

・$E[\theta_{UCL}|T,\theta_1] = \sum_{y_v=0}^{T} \theta_{UCL}(y_v)\binom{T}{y_v}{\theta_1}^{y_v}(1-\theta_1)^{T-y_v}$

④ ワクチン有効性の 95%信頼区間の下限値と上限値のそれぞれの期待値を算出する。

・$E[\pi_{LCL}|T,\theta_1] = \frac{2\times E[\theta_{UCL}|T,\theta_1]-1}{E[\theta_{UCL}|T,\theta_1]-1}$

・$E[\pi_{UCL}|T,\theta_1] = \frac{2\times E[\theta_{LCL}|T,\theta_1]-1}{E[\theta_{LCL}|T,\theta_1]-1}$

⑤ 判定:$\{E[\pi_{UCL}(T,\theta_1)] - E[\pi_{LCL}(T,\theta_1)]\}/2 < h$であれば、$T$を必要総イベント数$T^*$とし、手順を終了する。そうでなければ、$T = T+1$へ更新し、手順(2)へ戻る。

上記手順を用い、0.1 から 0.5 まで（0.1 刻み）のπ_1について、95%信頼区間の半幅の期待値を 0.1 未満とするための必要総イベント数を算出し、表 5 に示した。

表 5.ワクチン有効性の 95 %信頼区間の半幅の期待値を 0.1 未満とするための必要総イベント数

ワクチン有効性の期待値π_1	必要総イベント数T^*	ワクチン有効性の 95%信頼区間の期待値	信頼区間の半幅
0.1	1,289	[-0.005427,0.19451]	0.09997
0.2	1,032	[0.093953,0.29394]	0.09999
0.3	810	[0.19324,0.39319]	0.09998
0.4	620	[0.29234,0.49218]	0.09992
0.5	458	[0.39093,0.5908]	0.09998

5. 最後に

本稿では、ワクチン群とプラセボ群の 2 群比較試験の計画において、2 群のイベント発現総数に条件付きでワクチン群のイベント発現数が従う二項分布に基づいて、ワクチン有効性についての仮説検定で必要なイベント発現数を算出する方法（条件付き法）に関する検討結果を記述した。本稿では、条件付き法の適用が妥当であるための条件を、両群の試験期間中の発現割合が 0.3 以下であることと設定した。必要被験者数ではなく、必要イベント数に注目したのは、イベント数による試験の進捗管理が行いやすいためである。そのこ

とを、総イベント数についての条件付けを行わず必要被験者数を算出する方法との比較に基づいて示した。さらに、ワクチン有効性の推定において、Senn(2022)によるケース・プロポーションの 95%信頼区間算出法が優れた性質をもつことを示したうえで、ワクチン有効性の推定精度の観点から必要イベント数を算出するための一案を提示した。

本稿では、1 対 1 の割付比の 2 群試験を対象としたが、1 対 1 以外の割付比の 2 群試験における条件付き法の定式化が Chan & Bohidar(1988)で与えられている。また、条件付き法は、プラセボ対照試験だけではなく、実薬対照の非劣性試験でも利用可能である(江口・浦狩,2025)。具体的に、対照薬を基準とするリスク比をRR'=(被験薬での発現割合)/(対照薬での発現割合)と定義し、非劣性検定の帰無仮説を$RR' \geq r_0$、対立仮説を$RR' < r_0$と与える。ここで、r_0はリスク比の尺度上の非劣性マージンである。対照薬を基準とするワクチン有効性をπ'(=1-RR')とおけば、非劣性検定の帰無仮説が$\pi' \leq \pi'_0$、 対立仮説が$\pi' > \pi'_0$で表せる(π'_0=1-r_0はワクチン有効性の尺度上の非劣性マージン)。例えば、リスク比RR'について非劣性マージンが 2.0 と設定され、被験薬と対照薬の発現割合が等しいことを仮定して行う非劣性検定は、帰無仮説値を$r_0 = 2.0$、対立仮説の期待値を$r_1 = 1.0$とする検定となる。そして、この検定は対照薬を基準とするワクチン有効性π'について、帰無仮説値を$\pi'_0 = -1.0$、対立仮説の期待値を$\pi'_1 = 0$とする検定と同等である。この非劣性検定において、片側有意水準 0.025 で 90%の検出力を確保するのに必要な総イベント数は 275 件と算出される。

参考文献

- Chan,I.S.F. and Bohider,N.R.(1998). Exact power and sample size for vaccine efficacy studies. Communications in Statistics-Theory and Methods, 27(6),1305-1322.
- Gu, K., Ng, H.K.T., Tang, M.L., and Schucany, W. (2008) Testing the ratio of two Poisson rates. Biometrical Journal, 50(2), 283-298.
- Farrington, C. P., and Manning, G. (1990). Test statistics and sample size formulae for comparative Binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. Statistics in Medicine, 9,1447–1454.
- Loiacono,M.L.(2018). Sample size estimation and power calculations for vaccine efficacy trials for exceedingly rare diseases. (This thesis was approved by Hanna Bandos, PhD, Robert Krafty, PhD and Willem G. van Panhuis, MD PhD in University of Pittsburgh).
- Senn,S.(2022). The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy. Pharmaceutical Statistics, 21,790-807.
- Przyborowski,J. and Wilenski, H.(1940). Homogeneity of results in testing samples from Poisson series: with an application to testing clover seed for dodder. Biometrika, 31,313-323.
- 伊藤要二(2016). ポアソン分布に従うと仮定された総ケース数が固定された条件の下での rate ratio の条件付き信頼区間. SAS ユーザー総会・報告資料.
- 江口凌平・浦狩保則(2025). ワクチン開発における非劣性試験のデザインについて. SAS ユーザー総会・報告資料.
- 杉山 将(2013).確率分布間の距離推定：機械学習分野における最新動向. 日本応用数理学会論文誌, 23(3),439-452.

PMx解析に向けた統合データセット構築の実務的工夫

～SASプログラマーの視点から～

○トロンナムチャイ　ブンスィニー[1]、白武　要[1]

（[1]イーピーエス株式会社　臨床開発事業本部　データサイエンスセンター　統計解析１部）

Improvement of work content for building Integrated Dataset for PMx analysis
- From SAS programmer's perspective

Bunsinee Throngnumchai[1] and Kaname Shiratake[1]
[1]EPS Corporation

要旨

CDISC 準拠の統合データセット（IDS）を SAS で構築し、R（IQRtools）との連携による PMx 解析支援のための設計・実装の工夫を紹介します

キーワード：
Pharmacometrics Analysis、PMx、SDTM、SAS、R

1.　はじめに

Pharmacometrics（PMx）解析用データセットは、臨床試験で作成した SDTM データを統合して構築される、投与情報や被験者背景、薬物濃度など多くのパラメータを含むデータセットである。PMx 解析用データセットを使用することで、薬物動態、効果の関係をモデル化し、定量的に評価することが可能になる[1]。

本稿は、協和キリン株式会社とイーピーエス株式会社が共同で実施したトライアルを通じて、解析に必要な統合データセット（IDS）の作成プロセスに着目し、SAS プログラマーの立場からその改善事例を報告する。

本トライアルでの両社の共通目標は、臨床開発プロセス全体を通じて一貫性のあるデータ構造を確立し、解析の品質および効率性を向上させることである。本トライアルでは、臨床試験期間中にカットオフデータを用いて PMx 解析用データセットを作成し、タイムリーな評価を行っている。

また、モデルの精度や信頼性における PMx 解析用データセットの影響は大きい[1]ため、より精度が高く、信頼性の高い PMx 解析用データセットの作成が求められた。

2. PMx 解析の実施体制

本トライアルでは、協和キリン株式会社が Pharmacometrician として R パッケージ「IQRtools」を用いた解析を実施した。一方、イーピーエス株式会社は SAS プログラムを用いて統合データセット（IDS）の作成を担った。

PMx 解析の全体的な流れは次の通りである。まず、1 試験または複数試験から得られた SDTM または ADaM データセットを統合する作業と IQRtools を用いた解析に適した形式への変換作業を、SAS プログラムを用いて行い、IDS を作成した。次に、IDS と IQRtools を用いて薬物動態および薬効との関係性をモデル化することで、定量的な評価を実施した。

準備段階において、協和キリン株式会社は、IQRtools の活用を前提とした General Dataset Format の設計を行い、CDISC プログラマーにとって親和性の高い仕様およびデータ構造の定義を検討した。一方、イーピーエス株式会社は、SDTM 作成業務の延長線上で実装可能なデータ構造へとアップデートを行った。

将来的には、複数試験の SDTM または ADaM データセットを統合した PMx 解析用データセット作成を検討しているが、本トライアルでは 1 試験分の SDTM および ADaM データセットを統合して IDS を作成した。

3. IDS について

3.1 IDS の基本データ構造

SDTM の多様なデータ構造を IDS で統一的に取り扱うため、PMx 解析に必要なパラメータすべてを LB ドメインなどの Findings クラス様の縦積み構造に変換し、統一された形式で SAS データセットに格納した。1 つのレコードにつき、年齢、年齢カテゴリ、人種などの単一パラメータを表す構造としている。DM ドメイン由来の情報は、横積み構造として各変数に格納されていた情報を項目ごとにパラメータ化し、縦積み構造へと変換した。

この設計は、拡張性に優れ、新たなパラメータやイベント情報の追加が容易であるとともに、CDISC の SUPPQUAL に類似した発想に基づいており、SAS プログラマーにとっても、理解しやすい構造である。

3.2 IDS の主な変数一覧

表 1 に、IDS で作成する主な変数を示す。特に、対象変数の名称を示す NAME 変数、測定値を示す VALUE 変数、治験薬初回投与日時を示す STDTC 変数、イベント日を示す ENDTC 変数は、IQRtools での解析において必須の変数である。

導出元のドメインを示す DOMAIN 変数、NAME 変数、VALUE 変数については、参照元を明確に IDS 仕様書に記載することで、トレーサビリティを確保した。

また、STDTC 変数および ENDTC 変数を用いて算出する、経過時間を示す TIME 変数も重要な変数であり、特に EDC において日付のみが収集されている検査項目においては、時間情報の補完を行うことが不可欠である。

これらの変数設計により、R との連携を前提とした、柔軟かつ汎用性の高いデータ構造を実現した。

表 1 IDS で作成する主な変数

Variable	Type	Description
Study Information		
STUDY	String	試験名
Identification of Subject		
USUBJID	String	被験者識別子
Treatment Group Information		
TRTNAME	String	投与群
TRT	Numeric	投与群のコード値
Visit Information		
VISNAME	String	Visit 名
VISIT	Numeric	Visit Number
BASE	Numeric	ベースラインフラグ
SCREEN	Numeric	スクリーニングフラグ
Event Time Information		
STDTC	String	初回投与日時
ENDTC	String	投与・採血イベント日：EXSTDTC, PCDTC, LBDTC, VSDTC…
TIME	Numeric	初回投与からの経過時間：ENDTC-STDTC
Event Value Information		
DOMAIN	String	導出元のドメイン：DM, EX, PC, LB…
NAME	String	対象変数の名称：Age, Race, Weight, Baseline Height, Albumin…
NAMECD	String	NAME のコード値：AGE, RACE, WEIGHT, HEIGHT, ALB…
VALUE	Numeric	数値データ：0, 0.1, 1, 2, 4.2…
VALUETXT	String	カテゴリカルな値のテキスト：Null, Asian, Normal…
UNIT	String	単位：Years, Null, kg, cm, g/dL…
ULOQ	Numeric	定量上限
LLOQ	Numeric	定量下限

3.3 IQRtools を活用するための最小要件

IQRtools を有効に活用するためには、以下の 3 点が最小要件である。

1. データセットが縦積み構造であること。
2. STDTC 変数および ENDTC 変数を用いて経過時間（TIME 変数）を定義していること。
3. NAME 変数および VALUE 変数により、パラメータとその値が明確に記述されていること。

これらの要件を満たすことで、IQRtools によるセミオートマチックな SAS データセットの変換が可能となり、解析用データセットへの効率的な変換が実現された。さらに、品質診断機能や自動プロット生成などの IQRtools での解析支援機能も活用可能となる。

3.4 IDS の欠損補完

IDS における欠損補完は、PMx 解析精度の維持において極めて重要な工程である。

特に、LB などの Findings 系ドメインにおいては、EDC で開始日付および終了日付が収集されていても、時刻情報が欠落しているケースがある。このような場合、経過時間を示す TIME 変数を正確に算出することが困難となる。IQRtools は時間情報を軸に動作するため、日時情報の形式が統一されていないと、正確な処理が不可能となる。したがって、STDTC 変数や ENDTC 変数に時間情報が欠落している場合には、時刻情報として”00:00”などの補完処理を行い、TIME 変数の算出を可能にした。

本対応は、通常の CDISC 関連解析におけるデータハンドリングと同様に、前処理の品質が PMx 解析精度に直結するため、慎重かつ丁寧な対応が求められる対応であった。

3.5 IDS 仕様書の実務ポイント

IDS 仕様書の作成においては、以下の 3 点が重要な実務的留意事項である。

1. 参照元ドメインと変数を明確に定義すること。
2. 本体仕様を記載する spec シートと、パラメータ定義を記載する NAME_VALUE シートに分離すること。
3. IDS の変数定義は、SDTM 仕様に準拠した定義とすること。

1 点目については、各変数が複数の SDTM ドメインを参照するため、参照元ドメインと変数を明確に定義し、トレーサビリティを確保する必要があるため、重要である。

2 点目については、 VALUE 変数に多様なパラメータが集約されるため、仕様書の構造の複雑化を回避する目的でシートを分離した。これにより、仕様書の可読性と保守性が向上した。

3 点目については、VALUE 変数に格納されている値を NAME 変数で説明する構造を採用し、IDS と CDISC の各変数を表 2 の通りに対応するよう設計した。これにより、CDSIC プログラマーが直感的に理解しやすい構造となった。

表 2 IDS と CDISC で対応する変数一覧

IDS	SDTM
NAMECD	xxTESTCD
NAME	xxTEST
VALUE	xxSTRESN
UNIT	xxSTRESU

これらの実務的留意事項を意識した設計により、CDISC プログラマーが直感的で理解しやすい構造となり、仕様作成やプログラミング時の迷いを最小限に抑えることできたと考えられる。

4. まとめ

本トライアルでは、CDISC 準拠の統合データセット作成手順を SAS により確立し、PMx 解析に必要な高品質データを安定的に提供可能な体制を構築した。R との連携を通じて、解析の柔軟性および再現性が向上し、臨床試験中における迅速な意思決定支援にも寄与する結果となった。

SAS プログラマーの視点からは、標準化された仕様書に基づく効率的なデータ作成と、他ツールとの連携を前提とした運用設計が、今後の解析業務の高度化に向けた重要な基盤であると位置づけられる。これらの取り組みは、PMx 解析の実務において、品質保証と作業効率の両立を可能にする有効な手法であると考える。

5. 謝辞

本トライアルの遂行にあたり、SAS ユーザー総会 2025 において「Pharmacometrics 解析のための データセット作成プロセスの改善 SAS と R の Collaboration の提案」のご発表とともに、貴重なご協力を賜りました協和キリン株式会社 碇 新里様に心より感謝申し上げます。

6. 引用情報

[1] Pharmacometrics 解析のための データセット作成プロセスの改善 SAS と R の Collaboration の提案 (碇新里，Dimple Patel，永田吉範，Douglas Marsteller (協和キリン株式会社), SAS ユーザー総会 2025)

SASから呼び出すPowerShell

○伊住 寿雄

（ソーバル株式会社）

Powershell to be called from SAS

Izumi Toshio
Sobal corporation

要旨

本稿では、SAS のテキストファイル出力機能を活用し、PowerShell スクリプトを動的に生成・実行することで、SAS 単独では困難な処理を可能にする手法について述べる。従来の DDE（Dynamic Data Exchange）による外部アプリケーション制御の課題を整理し、PowerShell による代替手段の利点を示す。さらに、Word や Excel を対象とした具体的な活用事例と、応用的な比較処理の実装例を紹介する。

キーワード：

PC-SAS、PowerShell、Excel、Word、RTF

はじめに

SAS で出力した Excel ファイルや RTF ファイルを DDE（Dynamic Data Exchange）等を使用して見栄え等を整えることは一般的によく知られている。
また、昨今文字コードが utf-8 であるデータを扱うことが増えてきており、この環境で DDE を使用するといくつか問題点がある。
そのため、DDE に代わるものとして、utf-8 環境でも使用可能な Windows PowerShell（以下 PowerShell）を用いた方法を検討した。
本稿では、SAS から PowerShell を呼び出して、Excel ファイルや RTF ファイルの見た目を整える方法や、応用例として、Excel/RTF ファイルと SAS データを比較する事例の紹介をする。

Windows PowerShell とは

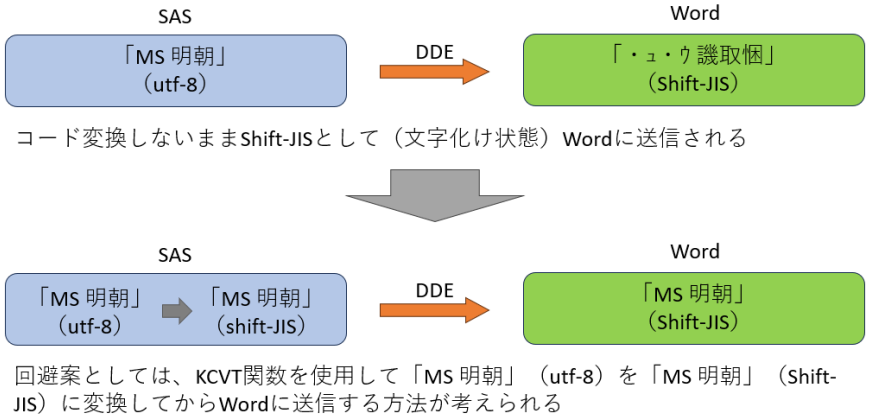
米マイクロソフト（Microsoft）社の Windows が標準で備えるコマンドライン方式のシェルおよびスクリプト実行環境の一つ。.NET Framework 上で動作する。（IT 用語辞典 e-words より引用）。
非推奨となった VBScript の代替言語として今後 PowerShell がこれまでよりも多く使われる可能性がある。

DDE を使用する際の問題点

- 文字コードの問題

  DDE は Shift-JIS で動作するため、UTF-8 で出力された文字列をそのまま送信すると文字化けが発生する。SAS の KCVT 関数によるコード変換も可能だが、大量データには不向きである。

  rtf出力後、Wordを開いてフォントを変更する例



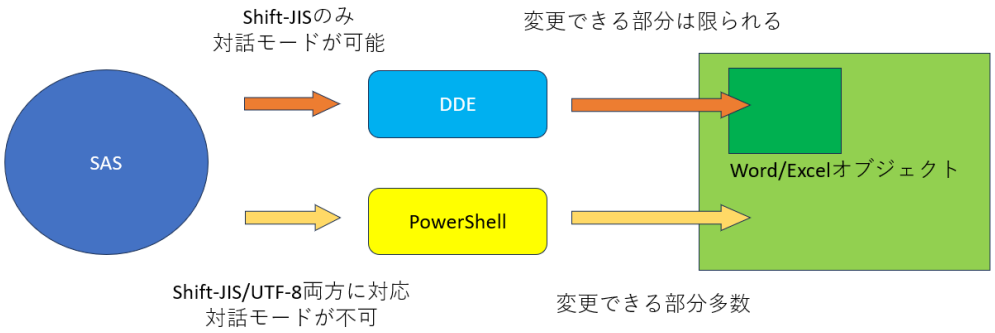
- 技術的制約

  DDE はコマンド構文がわかりにくく、特に Word DDE の引数仕様が不明瞭である。Excel DDE は Excel マクロ 4.0 に基づいているが、設定できる項目が限定的である。

代替手段の検討

以下の理由により、DDE の代替手段として PowerShell を使用することを検討する。

- Windows 標準で持っている機能
  ※余計なコストを必要としない
- SAS から起動可能
- utf-8 環境でも動作可能
- PowerShell は DDE と同様に Word/Excel の COM オブジェクトにアクセス可能である
  ※DDE と記述方法が似ているのと、記述方法の情報が DDE よりも豊富である
- DDE よりも設定可能項目が豊富



DDE と PowerShell の比較

DDE は対話モードで動作するため、Word/Excel がフォアグラウンドの状態でなければ、正常に動作しないことがあるが、PowerShell の場合はバックグラウンドで動作させることができるため、他の Word/Excel のウインドウが開いていても正常に動作する。また、PowerShell 実行中、他の作業も並行して行えるという利点もある。

DDE	PowerShell
フォアグラウンドで動作	バックグラウンドで動作
・他の Word/Excel のウインドウが開いていない状態で実行 ・実行中はほかの操作ができない	・他の Word/Excel のウインドウが開いていても実行可能 ・実行中も他の操作が可能

実装方法

実装するには主に 2 つのパートが必要である。

1. SAS のテキスト出力機能を用いて、PowerShell スクリプトファイル（ps1 ファイル）を動的に生成する
2. 生成した PowerShell スクリプトファイル（ps1 ファイル）を実行する

PowerShell スクリプトファイル（ps1 ファイル）を動的に生成する方法は、用途に応じて主に 2 種類の方法が考えられる。

- proc stream を用いた出力（固定処理向け）
- データステップ中で put ステートメントを用いた出力（条件分岐処理向け）

Word/Excel を用いて rtf/xlsx ファイルを編集する際は、PowerShell スクリプトファイルは以下の流れになるように記述する。

1. Word/Excel の COM オブジェクトを取得する
2. 取得した COM オブジェクトを用いて、編集するファイルを開く
3. 編集する箇所を選択する
4. 編集する
5. 3, 4 を繰り返してすべての編集を行う
6. 編集が終了したら上書き保存する
7. 編集したファイルを閉じる
8. Word/Excel を終了する
9. 1 で取得した Word/Excel の COM オブジェクトを解放する
10. メモリリークを防ぐため、ガベージコレクションを実行する

■ PowerShell スクリプトファイル（ps1 ファイル）生成の記述例

以下、PowerShell スクリプトファイル（ps1 ファイル）を動的に生成する SAS のサンプルコードを示す。

- proc stream を用いた出力（Word を操作する例）（固定処理向け）

```
/*PowerShell 実行ファイル名(work フォルダ内に作成)*/
%let psfile    = %sysfunc(pathname(work))¥sample.ps1;
filename psfile "&psfile." encoding = "utf-8" bom;

/*PowerShell の実行ファイルを作成*/
proc stream outfile = psfile resetdelim = "goto";
begin
<# スクリプトブロック化 #> goto NEWLINE;
%str(&) { goto NEWLINE;
<# Word オブジェクトの取得 #> goto NEWLINE;
$WordObj = New-Object -ComObject Word.Application goto NEWLINE;
$WordObj.Visible = $true goto NEWLINE;
<# 出力ファイルを開く #> goto NEWLINE;
$WordDoc = $WordObj.Documents.Open("&outputfile.") goto NEWLINE;
<# 印刷レイアウト表示にする #> goto NEWLINE;
$WordDoc.ActiveWindow.View.Type = 3 goto NEWLINE;
<# 内容の編集 #> goto NEWLINE;

ここに出力ファイルの内容を編集する PowerShell のコードを記述する

<# 上書き保存 #> goto NEWLINE;
[void] $WordDoc.save() goto NEWLINE;
<# Word 終了 #> goto NEWLINE;
```

```
[void] $WordDoc.close() goto NEWLINE;
[void] $WordObj.quit() goto NEWLINE;
[void][System.Runtime.Interopservices.Marshal]::ReleaseComObject($WordObj) goto NEWLINE;
<# ガベージコレクション #> goto NEWLINE;
} goto NEWLINE;
Get-Variable | Where-Object Value -is [__ComObject] | Clear-Variable goto NEWLINE;
[gc]::Collect() goto NEWLINE;
[gc]::WaitForPendingFinalizers() goto NEWLINE;
exit goto NEWLINE;
;;;;
```

- データステップ中で put ステートメントを用いた出力（Excel を操作する例）（条件分岐処理向け）

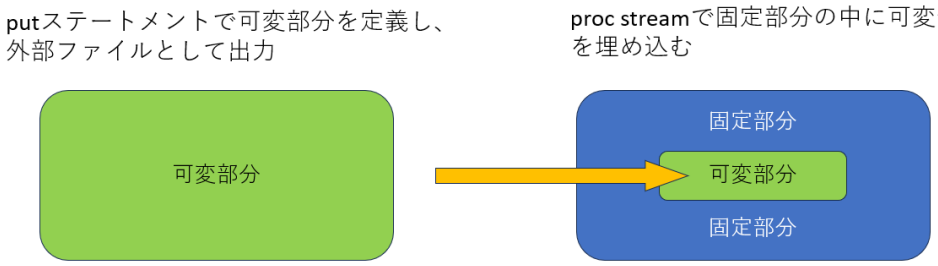
```
/*PowerShell 実行ファイル名(work フォルダ内に作成)*/
%let psfile    = %sysfunc(pathname(work))¥sample.ps1;

/*PowerShell の実行ファイルを作成*/
data _null_;
   file "&psfile." encoding = "utf-8" bom;
   /*スクリプトブロック化*/
   put '& {';
   /*Excel オブジェクトの取得*/
   put '$ExcelObj = New-Object -ComObject Excel.Application';
   put '$ExcelObj.Visible = $true';
   /*出力ファイルを開く*/
   put %unquote ( %bquote ( '$WorkBook = $ExcelObj.Workbooks.Open("&outputfile.")' ) );
   /*出力シートを選択する*/
   put %unquote ( %bquote ( '$WorkSheet = $WorkBook.Sheets("&sheetname.")' ) );
   /*内容の編集*/

ここに出力ファイルの内容を編集する PowerShell のコードを記述する

   /*上書き保存*/
   put '[void] $WorkBook.save()';
   /*Excel 終了*/
   put '[void] $WorkBook.close()';
   put '[void] $ExcelObj.quit()';
   put '[void][System.Runtime.Interopservices.Marshal]::ReleaseComObject($ExcelObj)';
   /*ガベージコレクション*/
   put '}';
   put 'Get-Variable | Where-Object Value -is [__ComObject] | Clear-Variable';
   put '[gc]::Collect()';
   put '[gc]::WaitForPendingFinalizers()';
   put 'exit';
run;
```

なお、proc stream では出力内容に別の外部ファイルの内容を記述可能であるため、可変部分を put ステートメントで一度外部ファイルに出力して、固定部分作成時に組み合わせるというハイブリッドな出力も可能である。



■ **PowerShell スクリプトファイル（ps1 ファイル）実行の記述例**

以下、生成した PowerShell スクリプトファイル（ps1 ファイル）を実行するサンプルコードを示す。

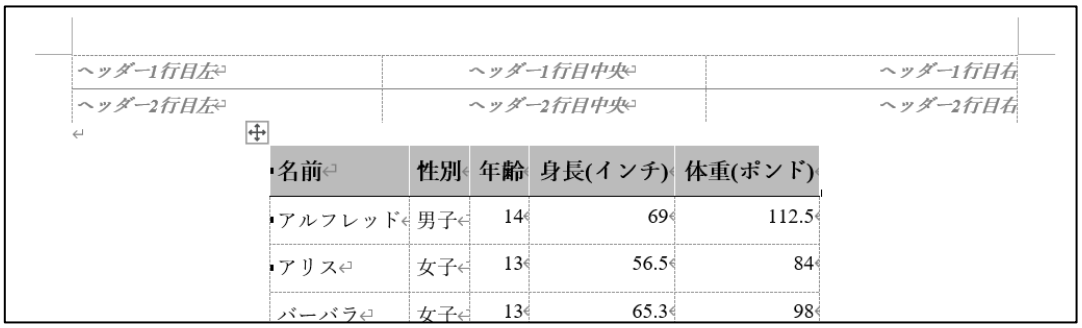
```
/*PowerShell の実行*/
data _null_;
   outdat = 'powershell -ExecutionPolicy RemoteSigned -File "' || "&psfile." || '"';
   putlog outdat;
   rc = system ( outdat );
run;
```

活用事例

活用事例として、SAS 単独では難しい見た目の調整を Word、Excel それぞれ 1 つずつ例として挙げる。

- **rtf ファイルのヘッダー部分の 1 行目と 2 行目の間に罫線を引く。**

  SAS 単独（ods rtf & proc report 等を使用）ではおそらく設定するのが難しい。



コードは以下の通り。なお、前述の proc stream を用いた出力（Word を操作する例）のコードを基に追加した部分（下線部分）のみを記す。

```
title1 height = 9pt justify = left "タイトル１行目左" justify = center "タイトル１行目中央" justify = right 'タイトル１行目右';
title2 height = 9pt justify = left "タイトル２行目左" justify = center "タイトル２行目中央" justify = right 'タイトル２行目右';
---途中省略---
proc stream outfile = psfile resetdelim = "goto";
begin
---途中省略---
<#  内容の編集  #> goto NEWLINE;

<#  ヘッダーを編集状態にする  #> goto NEWLINE;
```

```
$WordDoc.ActiveWindow.ActivePane.View.SeekView = 9 goto NEWLINE;
<# ヘッダーのすべてを選択 #> goto NEWLINE;
[void]$WordObj.Selection.WholeStory() goto NEWLINE;
<# ヘッダーの１行目に下線を引く #> goto NEWLINE;
$WordDoc.Sections(1).Headers(1).Range.Tables(1).Rows(1).Borders(-3).LineStyle = 1 goto NEWLINE;
$WordDoc.Sections(1).Headers(1).Range.Tables(1).Rows(1).Borders(-3).LineWidth = 4 goto NEWLINE;
<# ヘッダーの編集を終了する #> goto NEWLINE;
$WordDoc.ActiveWindow.ActivePane.View.SeekView = 0 goto NEWLINE;

---以降省略---
;;;;
```

- **Excel ファイルのヘッダー行の高さを調整＆中央寄せにする。**

  SAS では結合セルなどは行の高さが設定できないので、Excel 出力後調整する必要がある。

	A	B	C	D	E	F
1	名前	性別	年齢	身長(インチ)	体重(ポンド)	
2	アルフレッド	男子	14	69	112.5	
3	アリス	女子	13	56.5	84	
4	バーバラ	女子	13	65.3	98	
5	キャロル	女子	14	62.8	102.5	
6	ヘンリー	男子	14	63.5	102.5	
7	ジェームズ	男子	12	57.3	83	
8	ジェーン	女子	12	59.8	84.5	
9	ジャネット	女子	15	62.5	112.5	
10	ジェフリー	男子	13	62.5	84	
11	ジョン	男子	12	59	99.5	

コードは以下の通り。なお、前述のデータステップ中で put ステートメントを用いた出力（Excel を操作する例）のコードを基に追加した部分（下線部分）のみを記す。

```
/*PowerShell 実行ファイル名(work フォルダ内に作成)*/
%let psfile    = %sysfunc(pathname(work))¥sample.ps1;

/*PowerShell の実行ファイルを作成*/
data _null_;
   file "&psfile." encoding = "utf-8" bom;
---途中省略---
   /*内容の編集*/

   /*ヘッダー行の高さを変更する*/
   put '$WorkSheet.Range("1:1").EntireRow.RowHeight = 34';
   /*ヘッダー行を上下中央揃いにする*/
   put '$WorkSheet.Range("1:1").VerticalAlignment = -4108';
   /*ヘッダー行を中央寄せにする*/
   put '$WorkSheet.Range("1:1").HorizontalAlignment = -4108';
   /*カーソルを左上にする*/
   put '[void] $WorkSheet.Range("A1").Select';

---以降省略---
```

```
run;
```

他の用途の検討

PowerShell は COM オブジェクトを操作可能であるため、Outlook 等の MS 系アプリケーションや Acrobat などのアプリケーションとの連携も可能となる。そのため、以下のような応用も考えられる。

- SAS で出力ファイル作成後、 Outlook と連携させて出力ファイルの情報を関係者にメールする。
- SAS で PDF ファイル作成後、 Acrobat と連携させて出力した PDF ファイルに電子署名を行う

応用事例

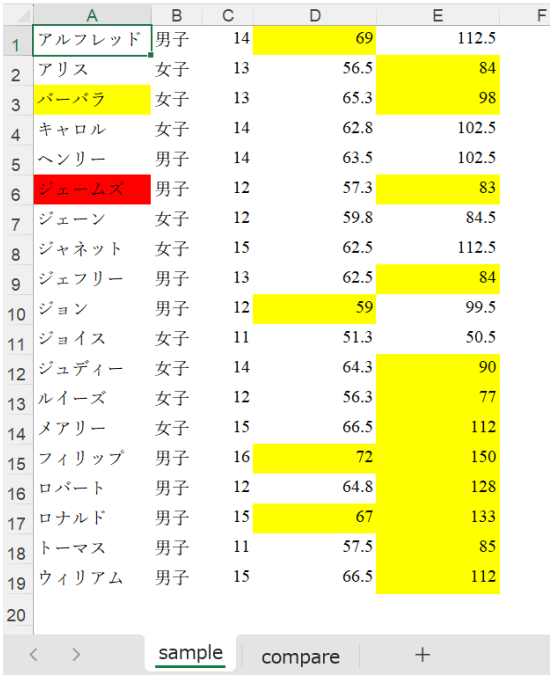
応用事例として、rtf/Excel ファイルの比較をするコードを検討する。
なお、諸事情によりソースコードは非公開とする。
以下の通りに処理が進むよう手順を検討した。

1. 比較元となる rtf ファイルを Word で開き、HTML 形式で保存しなおす。
   （比較元のファイルが Excel の場合）
   → Excel で rtf ファイルを直接読み込めないため HTML に変換する
2. 保存しなおした HTML ファイルを Excel で開く
   （比較元のファイルが Excel の場合は比較元となる Excel ファイルを開く）
   ※この際に文字列として出力された数値が数値として読み込まれてしまう
   例） 98.0 と表示されていた部分は 98 として読み込まれてしまう。
3. 開いた Excel ファイルからヘッダー領域およびフッター領域を削除する
4. 空白行を削除する
5. 比較先となる SAS データセットを Excel に出力する
6. 比較元、比較先の Excel ファイル両方を開き、比較先ファイルにシートを統合する
7. 比較元シート、比較先シート双方に以下のような条件付き書式を設定する
   ＜比較元シート：sample＞
   - =IF(ISERROR(NUMBERVALUE(A1)),TRIM(VALUETOTEXT(A1,0)),NUMBERVALUE(A1)) <> IF(ISERROR(NUMBERVALUE(compare!A1)),TRIM(VALUETOTEXT(compare!A1,0)),NUMBERVALUE(compare!A1))
   - =VALUETOTEXT(A1,0)<> VALUETOTEXT(compare!A1,0)

   ＜比較先シート：compare＞
   - =IF(ISERROR(NUMBERVALUE(A1)),TRIM(VALUETOTEXT(A1,0)),NUMBERVALUE(A1)) <> IF(ISERROR(NUMBERVALUE(sample!A1)),TRIM(VALUETOTEXT(sample!A1,0)),NUMBERVALUE(sample!A1))
   - =VALUETOTEXT(A1,0)<> VALUETOTEXT(sample!A1,0)
8. 上書き保存する

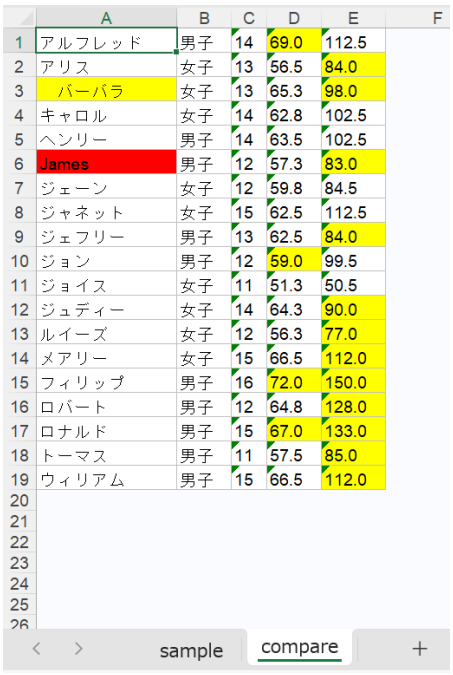
結果例は以下のようになる。比較元シート、比較先シート双方に条件付き書式を設定しているため、差分があるセルは以下のようになる。

- 文字列前後のブランク有無および数値としては一致しているが、表記は異なるセルは黄色
- 値そのものが異なるセルは赤色



	A	B	C	D	E
1	アルフレッド	男子	14	69	112.5
2	アリス	女子	13	56.5	84
3	バーバラ	女子	13	65.3	98
4	キャロル	女子	14	62.8	102.5
5	ヘンリー	男子	14	63.5	102.5
6	ジェームズ	男子	12	57.3	83
7	ジェーン	女子	12	59.8	84.5
8	ジャネット	女子	15	62.5	112.5
9	ジェフリー	男子	13	62.5	84
10	ジョン	男子	12	59	99.5
11	ジョイス	女子	11	51.3	50.5
12	ジュディー	女子	14	64.3	90
13	ルイーズ	女子	12	56.3	77
14	メアリー	女子	15	66.5	112
15	フィリップ	男子	16	72	150
16	ロバート	男子	12	64.8	128
17	ロナルド	男子	15	67	133
18	トーマス	男子	11	57.5	85
19	ウィリアム	男子	15	66.5	112
20					

sample compare +



	A	B	C	D	E
1	アルフレッド	男子	14	69.0	112.5
2	アリス	女子	13	56.5	84.0
3	バーバラ	女子	13	65.3	98.0
4	キャロル	女子	14	62.8	102.5
5	ヘンリー	男子	14	63.5	102.5
6	James	男子	12	57.3	83.0
7	ジェーン	女子	12	59.8	84.5
8	ジャネット	女子	15	62.5	112.5
9	ジェフリー	男子	13	62.5	84.0
10	ジョン	男子	12	59.0	99.5
11	ジョイス	女子	11	51.3	50.5
12	ジュディー	女子	14	64.3	90.0
13	ルイーズ	女子	12	56.3	77.0
14	メアリー	女子	15	66.5	112.0
15	フィリップ	男子	16	72.0	150.0
16	ロバート	男子	12	64.8	128.0
17	ロナルド	男子	15	67.0	133.0
18	トーマス	男子	11	57.5	85.0
19	ウィリアム	男子	15	66.5	112.0

sample compare +

まとめ

PowerShell を用いることで、SAS から出力後、出力ファイルの編集に柔軟性と拡張性を持たせることが可能となった。DDE の限界を補完し、UTF-8 環境にも対応できる点は大きな利点である。

参考資料

- 井上 貴博氏（2024） SAS ユーザー総会 2024
  文字コードに関する SAS の対応

- 関根　暁史氏（2024） SAS ユーザー総会 2024
  RTF に使える DDE

- 山野辺浩己氏（2024） SAS ユーザー総会 2024
  解析結果 RTF ファイル検証手法の検討と報告

- Office VBA リファレンス
  https://learn.microsoft.com/ja-jp/office/vba/api/overview/
  (参照日：2025/08/19)

ワクチン開発における非劣性試験のデザインについて

○江口凌平[1]、浦狩保則[1]

([1]シミック(株)データサイエンス事業本部)

Design of Non-Inferiority Clinical Trials in Vaccine Development

Ryohei Eguchi, Yasunori Uragari
Data Science Div., CMIC Co.,Ltd.

要旨

新規ワクチン群と対照薬群の非劣性試験の計画において、リスク比を治療効果の尺度とした場合の非劣性試験のデザインについて検討した。非劣性試験では、適切な非劣性マージンを担保した上で、試験薬のワクチン有効性を評価しなければならない。非劣性マージンを設定する上で、FDA ガイドライン(2016)で述べられている手法に従うことが最良であると考えた。リスク比についての非劣性マージンの設定のからくり(対数尺度での解釈、リスク比が望小特性であることなど)を図解した。

非劣性マージンの設定事例として、Senn(2022)による複数の過去試験のメタアナリシスを行った。そこでは、恒常性(constancy assumption)を考慮して、変量効果モデルの当てはめを行い、非劣性マージンの導出を行った。算出した非劣性マージンをもとに、必要イベント数・必要症例数のシミュレーションを行い、イベント発現率が大きい場合、小さい場合に算出される必要症例数についても考察を行った。

キーワード：ワクチン試験、非劣性試験、非劣性マージン、必要イベント数、ワクチン有効性、リスク比

1. ワクチン試験について

パンデミック下では、新型コロナウイルスワクチンのプラセボ対照試験において、プラセボ投与群の被験者には実薬が投与されないため、被験者が最善の医療を得られるような倫理的な配慮が必要とされた。このようなプラセボ投与群を設定することが困難な場合に、すでにコロナワクチンとして承認されている実薬を比較対象として臨床試験を実施することが考えられる。そこで、すでに有効性が判明しているワクチンを対照薬とおいて、免疫原性の指標、主に中和抗体価をエンドポイントとした、ランダム化実薬対照非劣性試験、又は優越性試験を代替とした開発が進められている。2021 年 6 月 24 日に開催された国際薬事規制当局連携組織(International Coalition of Medicines Regulatory Authorities: ICMRA)のワークショップでの議論によれば、このような試験では、ワクチンの免疫原性とその発症予防効果との間の高い相関性を示す必要がある。そのためには、免疫原性(抗体価の上昇)がサロゲートマーカーとして使用可能であること(中和抗体価の測定方法の信頼性確保、実薬と新たなワクチンとの間の細胞性免疫の反応の違いの考慮など)、国際的な評価基準(EMA, FDA など)に基づいていること、および複数の臨床試験で一貫した結果が得られていることなどの条件が必要

である (原 他(2013), 渡邉 他(2018))。

パンデミック後の現在は、新型コロナウイルス感染症の位置づけが「5 類感染症」になり、社会の向き合い方も大きく変化した。ワクチン試験の実施についても、発症予防効果をエンドポイントとした臨床試験の実施の動きがある。2 群試験に限れば、プラセボ対照優越性試験か実薬対照非劣性試験の選択がある。なお、上記の ICMRA ワークショップでも、発症予防効果をエンドポイントとしたプラセボ対照試験が依然として重要であることが結論付けられている。ただし、冒頭で述べた倫理的問題を考慮すれば、発症予防効果をエンドポイントとする実薬対照非劣性試験の計画・実施の可能性は考えられる。

本稿では、実薬対照非劣性試験における必要な適切な非劣性マージンの設定、適切な必要イベント数の決定について検討する。

2. リスク比を評価指標とする場合の非劣性マージン設定の手順

一般に非劣性試験とは、試験薬 (T)が対照薬 (C)に比べて非劣性マージン (δ)以上に劣ることはないことを示す試験のことである。帰無仮説 H_0 と対立仮説 H_1 とすると、H_0: C – T ≥ δ 、H_1: C – T < δ とそれぞれ表すことができる。ワクチン試験において、有効性の評価指標としてワクチン有効性が用いられる。ワクチン有効性は、ワクチンにおける発現率のプラセボにおける発現率に対するリスク比(Risk Ratio)を 1 から減じた値、すなわち 1−リスク比で定義されている。今回はリスク比を治療効果の尺度として使用することを想定する。

ここでは、上で定義されるプラセボに対するリスク比ではなく、対照薬群に対するリスク比、すなわち(試験薬群のイベント発現率 : P_T) / (実対照薬群のイベント発現率: P_C)を用いる。非劣性仮説は以下の通り示すことができる。(ここでの発現率は発現割合の意味合いで用いている。)

$$H_0: \frac{P_T}{P_C} \geq M \text{ vs } H_1: \frac{P_T}{P_C} < M \qquad (1)$$

また、上式を対数変換すると、同様に以下の通り示すことができる。

$$H_0: logP_T - logP_C \geq \log M \text{ vs } H_1: logP_T - logP_C < log\ M \qquad (2)$$

非劣性試験では、非劣性マージン M の設定が重要である。非劣性マージンは、対照薬 vs. プラセボの既存の試験で観測された群間差をもとに、当該試験において、プラセボ以上の効果が間接的に示されるような値に設定すべきことが、FDA ガイドラインでも明確に述べられている。具体的には、過去に行われた対照薬 vs. プラセボの優越性試験の差を M_1、M_1 の何割が維持されていれば試験薬が受け入れ可能かの臨床的判断を反映するための M_2 という指標を用いて非劣性マージンを算出する。今回は固定マージン法による M_1 と M_2 の選定に着目する。

試験薬 vs. 対照薬における M_1 と M_2 の選定手順について、図 1 を用いて説明する。図 1 中に記載の括弧書きの数値については、後ほど説明する。試験薬 vs. 対照薬の試験を検討する上で、①対照薬 vs. プラセボ、②プラセボ vs. 対照薬、③試験薬 vs. 対照薬の 3 つの段階に分けて考える。

① 対照薬 vs. プラセボ : 対照薬群のイベント発現率をP_c、プラセボ群のイベント発現率をP_pとすると、対照薬群のプラセボ群に対するリスク比は、P_c/P_pとなる。以降、リスク比が望小特性であることに注意する。このリスク比をもとに非劣性マージンの算出を行うが、対数尺度上で解釈することが重要である。右図の①はリスク比を対数変換した $\log P_c - \log P_p$ を用いた場合の図である。

② プラセボ vs. 対照薬 : ②は、プラセボ群と対照薬群の関係を①とは逆にしており、最後の③における試

験薬 vs. 対照薬の前段階の図である。FDA ガイドラインでは、左図の②にあるP_p/P_cの 95%信頼区間の下限値を非劣性マージンM_1と置いている。右図に示すように、対数尺度上では、①での 0 から $\log P_c - \log P_p$ の 95%信頼区間の上限値までの距離と②での $\log P_p - \log P_c$ の 95%信頼区間の下限値から 0 までの距離が等しい、すなわち①と②の赤い線分の長さが等しいことに留意したい。

③ 試験薬 vs. 対照薬：③は、②でのプラセボ群を試験薬群に置き換えた形になっている。左図の③に示すP_T/P_cのリスク比について、その信頼区間の上限< M_1 であることを、試験薬がプラセボよりも大きい効果を持つこと、つまり有効性の証明として位置づけることができる。ただし、臨床的判断のもとでは、その信頼区間の上限がM_1よりも小さいだけでは十分ではなく、$M_2 < M_1$ となるような M_2 を非劣性マージンとすることが必要であるとされる。通常、M_2 を M_1 の 50%とすることが多いが、この計算は対数変換尺度上で行われていることに留意したい。実際に、右図の③で、$\log M_2 = \frac{1}{2}\log M_1$として M_2 の値が求められる。

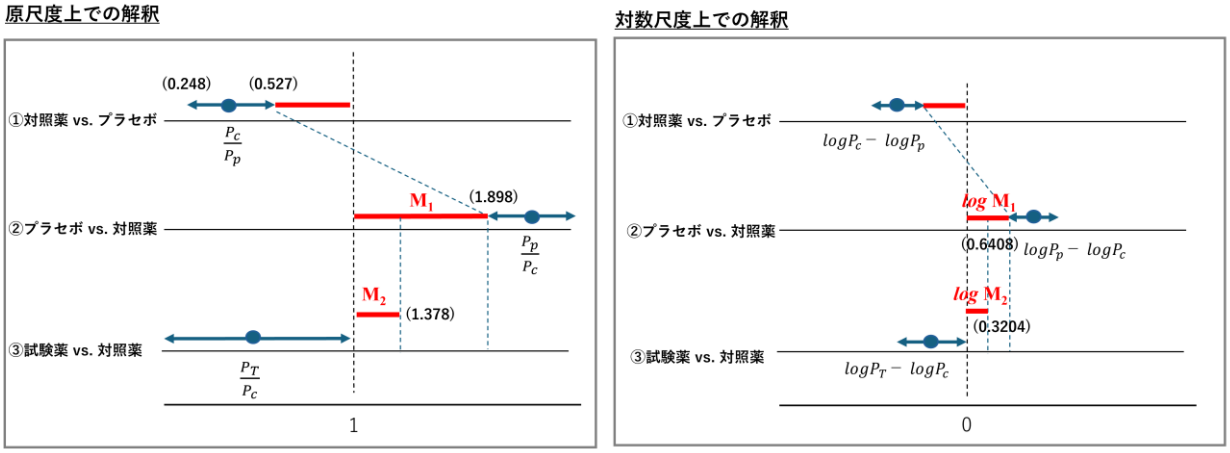


図 1．リスク比を指標とする場合の非劣性マージン M_1 と M_2 の選定手順

M_1 と M_2 を算出する事例が、FDA ガイドライン(2016)で示されている。その事例では、新しい抗凝固剤 ximelagatran と対照薬ワーファリンを用いた非劣性試験(SPORTIF V)を取り上げている。まずは、固定マージン法を用いた非劣性マージン M_1 の算出を記述している。M_1 を算出するために、変量効果メタアナリシスを用いて、6 つのプラセボ対照試験のそれぞれのリスク比を統合し、リスク比の点推定値 0.361 とその 95%信頼区間(0.248,0.527)を得ている。この変量効果メタアナリシスの手法については、次の 3 節で詳細に記載する。95%信頼区間の上限値 0.527 は、図 1 の左図の①に示す対照薬のプラセボに対するリスク比P_c/P_pの 95%信頼区間の上限値に該当する。その逆数の 1/0.527＝1.898 は、左図の②に示すプラセボの対照薬に対するリスク比P_p/P_cの 95%信頼区間の下限値に該当する。ガイドラインでは、もし試験薬とプラセボの効果が同じであれば何が起こるかという観点から、プラセボであれば対照薬に対して約 90%のリスクを増大させる(すなわち、リスク比が 1.898 となる)と説明している。つまり、左図の③で、試験薬の対照薬に対するリスク比P_T/P_cが 1.898 よりも小さいことが、試験薬がプラセボよりも優れた効果をもつことを意味している。

続いて、その事例では M_2 の算出を記述している。具体的に、試験薬が最大限に許容できる臨床的マージン M_2 を M_1 の 50%とするとの設定のもとで、図 1 の右図の③の通り、$\log M_1 = \log 1.898 = 0.6408$ の 50%は 0.3204 と算出される。これを原尺度上での値に変換すると、$M_2 = \exp(0.3204) = 1.378$ となる。なお、SPORTIF V 試験では、試験薬の対照薬に対するリスク比の点推定値が 1.39 であり、その両側 95%信頼区間が(0.91,2.12)であった。上限値(2.12)が M_2(1.378)よりも大きいため、この事例において ximelegatran のワーファ

リンに対する非劣性は示されなかったと説明されている。続く 3 節では、2 節で紹介した手順に従い、文献で公表されているデータを用いながら非劣性マージンの設定について紹介する。

3. 実データに基づく非劣性マージンの設定と試験の選択

非劣性試験のデザインでは、分析感度の検討が必須である。分析感度は治療間の差を識別する能力(ability)であり、仮に今の非劣性試験にプラセボ群を含めていたと想定すると、分析感度として、対照薬とプラセボの間に少なくとも M_1 の差が存在したはずであることを意味している。このような M_1 の選択や試験が分析感度を持っているかどうかを判断する上で、歴史的エビデンス(HESDE：Historical evidence of sensitivity to drug effects)を用いた変量効果メタアナリシスの手法が FDA ガイドラインでは推奨されている。HESDE は、過去に適切にデザインされ実施された試験で、特定の実治療を用いて、その治療がプラセボ群あるいは他の治療群よりも優れていることを示した試験を指す。過去の複数の試験で一貫して見られていることをもって、プラセボ群あるいは他の治療群に対してその治療が優れていると解釈する。ここで、上記の HESDE の情報を使用するためには、今計画している非劣性試験と過去の試験の類似性が担保されていなければならない。この仮定を、「恒常性の仮定(constancy assumption)」と呼ぶ。

我々は Senn (2022)の 5 試験のデータから変量効果メタアナリシスの手法を用いて、一つの包括的な統計量を算出する。この 5 試験は全て大規模なランダム化プラセボ対照試験であり、表 1 に過去 5 試験のイベント発現状況についてまとめている。

表 1. 過去 5 試験のイベント発現状況

Study	Subjects		Cases		追跡期間
	Vaccine	Placebo	Vaccine (%)	Placebo (%)	
A	20,712	21,096	77 (0.37)	850 (4.03)	6 ヵ月
B	17,662	8550	73 (0.41)	130 (1.52)	2 ヵ月
C	14,134	14,073	11 (0.08)	185 (1.31)	2 ヵ月
D	7020	7019	10 (0.14)	96 (1.37)	3 ヵ月
E	19,630	19,691	116 (0.59)	348 (1.77)	2 ヵ月

変量効果メタアナリシスのモデルとして、以下に記載の Bakbergenuly et al.(2019)の二項一般化線形混合モデルを用いる。

$$log(P_{ji}) = \alpha_j + (\theta + b_j)x_i\,,\ b_j \sim N(0,\tau^2)\quad j = 1,\cdots,K;\ \ i = C,T \qquad (3)$$

ここで、j は 1~K 番目の試験、i は対照薬群(C)、試験薬群(T)をそれぞれ表し、$x_c = 0,\ x_T = 1$とする。そして、P_{ji}は試験 j における群 i の発現率、α_jは試験 j におけるプラセボ群の対数発現率を表す固定効果、および$(\theta + b_j)$は試験 j におけるワクチン群のプラセボ群に対する対数リスク比である。ここで、θは全体の対数リスク比、$b_j \sim N(0,\tau^2)$は変量効果、およびτ^2は試験間分散を指している。したがって、このモデルのあてはめにおいて、α_j、θおよびτ^2を推定する。上記のモデルに対応する SAS プログラムを以下に記載する。

```
proc glimmix data=targ;
    class study treatment;
    model event/n = study treatment / noint link=log solution cl;
    random treatment / subject=study;
    lsmeans treatment / diff=all cl ilink;
    covtest 'common variance' ZEROG;
    output out=out2 pred(noilink)=predMu resid(noilink)=residMu
    LCL(noilink)=lower UCL(noilink)=upper;
    ods output diffs=diffsall;
run;
```

上記のように、GLIMMIX プロシジャ(SAS Ver.9.4)を用いて、Bakbergenuly et al.(2019)の二項一般化線形混合モデルをあてはめた。以下にプログラムの要点を記載する。

- 非劣性マージンを設定する上で、恒常性の仮定を満たすことは重要であるため、プラセボ群に対する対照薬群の処理効果が試験間でどれほど変動するかに関心がある。そこで、処理効果の部分に変量効果を加え、random ステートメントには treatment を記載している。一方、プラセボ効果については、固定効果とした。(なお、Bakbergenuly et al.(2019)にはプラセボ効果も変量効果とするモデルも示されているが、2 つの変量間の相関も考慮することが必要である。)
- この解析では、対数リスク比にてワクチン有効性を推定することから、link 関数を log に設定している。
- 対数リスク比が対数変換尺度上では引き算で求めることができるため、diff オプションを指定している。
- covtest 'common variance' ZEROG; にて、変量効果の分散＝0 の仮定が妥当かどうか、すなわち恒常性を仮定することの妥当性を評価している。

さらに、以下の 2 点を補足しておく。

- 一般的にリスク比のメタアナリシスで二項一般化線形混合モデルを用いる場合、各試験のイベント発現率が区間(0,1)に存在するという自然な制約が守られず、試験間変動の分散や対数リスク比の推定においてバイアスが招来され得ること、そして対照群の発現率が中程度かつ/あるいは対数リスク比が大きな値である場合に、バイアスが問題であることが報告されている(Bakbergenuly et al.,2019)。しかし、ワクチン試験では、イベント発現率が表 1 でも示される通り、0.08%～0.59%とかなり小さいため、そのバイアスは起こらないとしても良い。(起こらないことを確認している。)
- 表 1 に示したように、各試験の追跡期間は一様ではなかった。実際に、プラセボ群の発現率は、追跡期間が最長の試験 A が他の 4 試験よりも高値となっており、これらの試験の解析では追跡期間の長さも考慮することが必要であると考えられた。そのため、model ステートメントの右辺に、追跡期間を表す変数 followup を追加した場合も実行したが、変数 followup の回帰係数は 0 で推定された。その理由は、各試験のプラセボ効果を表す変数 study と変数 followup が別名(alias)の関係になっており、変数 study の効果の推定値に followup の効果が込みにして含められているためであると考えた。

上記モデルを用いて、各試験および全体の対数リスク比とその 95%信頼区間を算出したところ、図 2 のような結果が得られた。

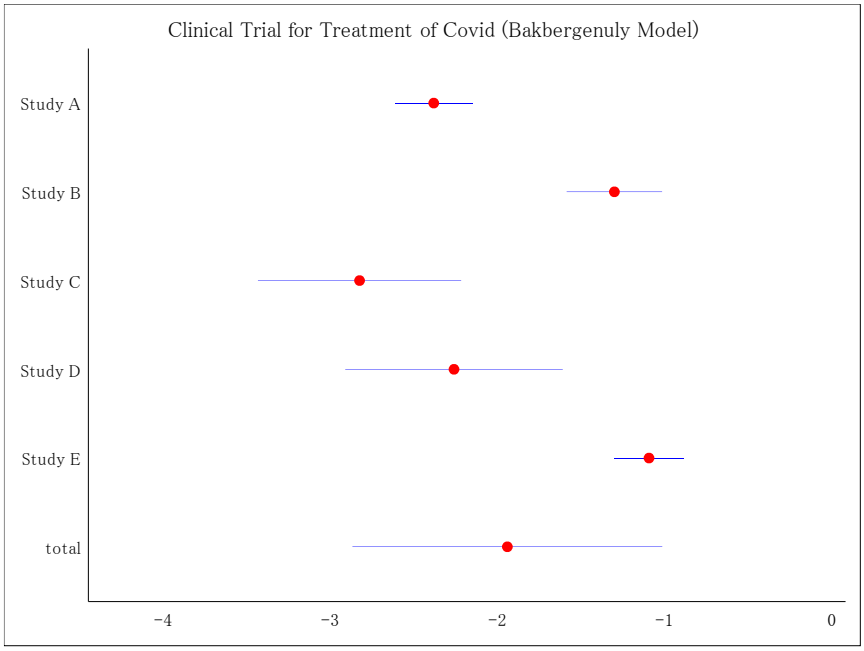


図 2. 過去 5 試験および全体の対数リスク比とその 95%信頼区間

Total がメタアナリシスを用いた全体の対数リスク比の値であり、−1.9428 と算出された。

また、上図の結果を見ると、他試験と比べて、Study B と Study E は明らかに対数リスク比の値が大きいことが分かった。これらの試験を含めた解析を行うことは、恒常性の仮定から外れた解析になっている可能性があるため、Study B と Study E は除いた解析も行った。

Study B と Study E は除いた上で、各試験および全体の対数リスク比とその 95%信頼区間を算出したところ、図 3 のような結果が得られた。

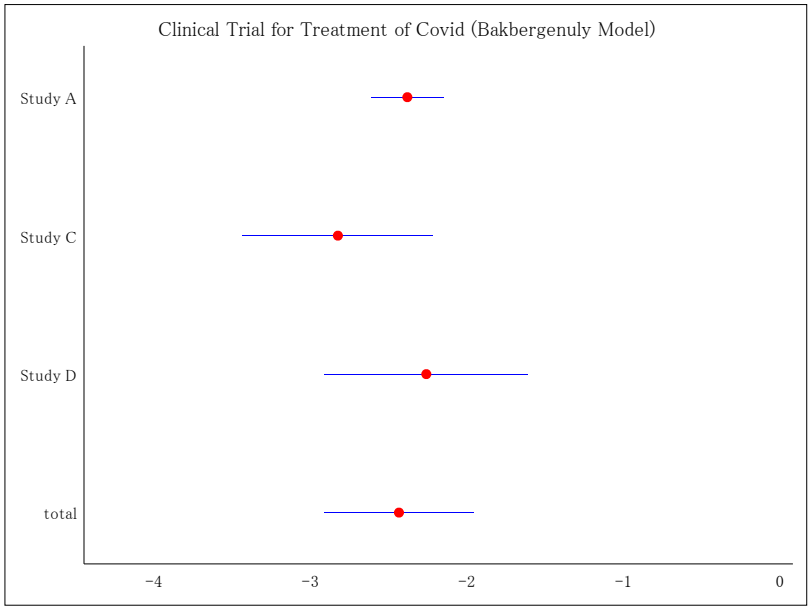


図 3. 過去 3 試験および全体の対数リスク比とその 95%信頼区間

ここで、先ほどの 5 試験全てを含めた解析と結果を比較したところ、表 2、表 3 の結果が得られた。

表 2．対数リスク比の推定値

Study	5 試験全てを含む		3 試験のみを含む	
	Estimate	95%CI	Estimate	95%CI
A	−2.3831	(−2.6155, −2.1506)	−2.3831	(−2.6155, −2.1506)
B	−1.3026	(−1.5881, −1.0171)	-	-
C	−2.8268	(−3.4346, −2.2190)	−2.8268	(−3.4346, −2.2190)
D	−2.2619	(−2.9123, −1.6115)	−2.2619	(−2.9123, −1.6115)
E	−1.0955	(−1.3047, −0.8863)	-	-
Total	−1.9428	(−2.8695, −1.0160)	−2.4366	(−2.9144, −1.9588)

表 3．COVTEST 結果

5 試験全てを含む			3 試験のみを含む		
共分散パラメータ推定値τ^2	標準誤差	P-value	共分散パラメータ推定値τ^2	標準誤差	P-value
0.2561	0.195	<0.0001	0.001	0.027	0.4828

表 2 において推定値の結果を比較したところ、先ほどの 5 試験全てを含めた解析と比べ、全体の対数リスク比は小さくなり、−2.4366 と算出された。

また、COVTEST の結果を比較したところ、5 試験全てを含めた場合の COVTEST の検定結果は p<0.0001 と有意な結果になっているのに対し、3 試験のみを含めた場合には、COVTEST の検定結果は p=0.4828 となり、有意ではない結果となっている。この結果から、対数リスク比の値が大きい試験を除いた上での解析については、非劣性試験に必要である恒常性の仮定が満たされたと判断し、2 節で記述した考え方を元に、M_1 と M_2 の非劣性マージンを算出する。

上述で求められた Total の推定値の信頼区間の上限値に該当する−1.9588 の値を用いる。−1.9588 は対照薬 vs. プラセボでの値であるため、2 節で述べた通り、プラセボ vs. 対照薬を考慮し、$log\ M_1$ =1.9588 となる。臨床的にはその半分が維持されるべきであると考えると、$log\ M_2 = (0.5) log\ M_1$ =0.9794 となる。原尺度へ戻すと、M_2 =exp(0.9794)=2.6629 が求められる。すなわち、新たなワクチンを開発する上で、試験薬の対照薬に対するリスク比の信頼区間の上限値が、算出した非劣性マージン値の 2.6629 よりも低い場合に非劣性が証明されることになる。

ところで、Fleming et al.(2021)は、類似の方法を用いて、非劣性マージンを算出している。具体的に、対照薬のプラセボに対するワクチン有効性が 0.9、すなわちリスク比が 0.1 の場合(これは、表 1 の試験 A および試験 C におけるワクチン有効性に対応する)、そのプラセボ対照試験でのイベント数が 175 例および 350 例であった場合について、新規ワクチンの対照薬に対する非劣性マージンをそれぞれ 2.561 および 2.724 とすることが示されている。これらの値は、我々が算出した非劣性マージンに酷似している。また、Fleming et al.(2021)は非劣性マージンを 3 よりも大きくすべきでないとのステークホルダの見方を考慮して、データ解析で得られる非劣性マージンをMとして、試験デザインに用いる非劣性マージンを min(M,3)で与えることを提示している。この考え方は、ICH-E10 (2001)の「非劣性試験における限界値の決定は、統計的推論と臨床判断の双方

に基づくものであり、その選択の基となった証拠の不確実性を反映するとともに、適度に保守的であるべきである。」に基づいている。

4. 条件付き法による必要イベント数・必要被験者数の決定

3 節にて算出した非劣性マージンを用いて、必要イベント数・必要被験者数を決定した。必要イベント数は、浦狩(2025)で述べられている方法を元にして算出した。すなわち、ワクチン群とプラセボ群の 2 群比較試験の計画(割付け比は 1 対 1)において、2 群のイベント発現数合計に条件付きでワクチン群のイベント発現数が従う二項分布に基づいて、ワクチン有効性についての仮説検定で必要なイベント発現数を算出する方法(条件付き法)を基本とした。その際、2 群の割付け比を 1 対 1、片側有意水準を 0.025、検出力を 0.9 とした。優越性試験では、試験ワクチンの有効性をπとして、検定の帰無仮説と対立仮説を

$$H_0: \pi \leq \pi_0 \ \text{ vs } \ H_1: \pi > \pi_0 \qquad (4)$$

とし、帰無仮説値π_0、対立仮説の期待値π_1(>π_0)を指定すれば必要イベント数(2 群合計)を算出できる。

本稿で対象としている非劣性試験では、検定の帰無仮説と対立仮説が

$$H_0: RR' \geq r_0 \ \text{ vs } \ H_1: RR' < r_0 \qquad (5)$$

で表される。ここで、RR'は試験ワクチンの対照ワクチンに対するリスク比、r_0は非劣性マージン((1)式のMと同一)であり、これを帰無仮説値とおく。2 つのワクチンの有効性が等しいとの仮定のもとで対立仮説の期待値をr_1=1 とおくのが一般的である。この検定仮説を、(プラセボではなく)対照ワクチンを基準とするワクチン有効性$\pi' = 1 - RR'$を用いて書き換えると、上記の非劣性検定の帰無仮説と対立仮説は

$$H_0: \pi' \leq \pi'_0 \ \text{ vs } \ H_1: \pi' > \pi'_0 \qquad (6)$$

で表される。ここで、帰無仮説値は$\pi'_0 = 1 - r_0$であり、対立仮説の期待値を$\pi'_1 = 1 - r_1 = 0$ とおく。これらの設定のもとで条件付き法を用いれば、非劣性試験での必要イベント数を算出することができる。なお、非劣性マージンがr_0＞1 であるため、帰無仮説値はπ'_0 <0となるが対応するケース・プロポーションθ(総発現イベント数に条件付きの二項分布のパラメータ)について$0 < \theta_0 = (1 - \pi'_0)/(2 - \pi'_0)$ <1 である。また、対立仮説の期待値π'_1に対応するケース・プロポーションは$\theta_1 = 0.5$である。したがって、帰無仮説と対立仮説のそれぞれについて、条件付き二項分布の確率計算が問題なく実行可能である。詳細は、浦狩(2025)を参照。

3 節で算出した非劣性マージン$r_0 = \mathrm{M}_2$＝2.6629 から$\pi'_0 = 1 - r_0$＝－1.6629 および$\pi'_1 = 0$とし、片側有意水準 0.025 で 0.9 以上の検出力を得るために必要な総イベント数(T_{req})が 53 イベント、棄却限界値(Y_c：非劣性が認められる試験ワクチン群のイベント数の最大値)が 31 イベントと算出される(表 4)。なお、Fleming et al.(2021)は、非劣性マージンが 2.561 および 2.724 に対応する必要な総イベント数がそれぞれ 54 例および 48 例であることを示している。したがって、表 4 での結果は Fleming et al.(2021)で示された結果と同等のものが示されたことが分かった。

表 4. 非劣性試験で必要な総イベント数(2 群合計)

ワクチン有効性による記述		必要総イベント数：T_{req}	棄却限界値：Y_c
π'_0	π'_1		
－1.6629	0	53	31

必要被験者数の算出は、2 つのワクチンのもとでのイベント発現率Λ_v(件/人年)が同じであることを仮定し、浦狩(2025)の(8)式のプラセボの発現率Λ_pをΛ_vで置き換えることで行える。

ところで、Fleming et al.(2021)には、ワクチンの非劣性試験はプラセボ対照試験よりも試験規模が 2~3 倍に

なることを示す次の記載があり、荒木(2021)、厚生労働省(2021)でも引用されている。

“… the 34-event trial comparing two vaccines having approximate 95% vaccine efficacy and the 48-event trial comparing two vaccines having approximate 90% efficacy would require approximately two to three-fold person years of follow-up relative to a frequently used design of a 150-event placebo-controlled trial of a vaccine having 60% vaccine efficacy, assuming these trials were conducted in settings having similar attack rates.”

ここでは、上記で算出した非劣性試験での必要イベント数 53 に対応する被験者数の算出の傍らで、この記載内容を定量的に確認することとした。その際、以下の条件設定を行った。

・プラセボ群のイベント発現率(件/人年)：0.01、0.05、0.1。ここでの 0.1(件/人年)はパンデミック時の発現率に対応している(表 1 を参照)。
・ワクチン群のイベント発現率(件/人年)：プラセボ群のイベント発現率の 10 分の 1。3 節で算出した対照ワクチンの対プラセボのリスク比の推定値 0.0875(=exp(－2.4366))であることに基づく。この設定は、“two vaccines having approximate 90% efficacy”の状況に近いものである。
・プラセボ対照優越性試験：“150-event placebo-controlled trial of a vaccine having 60% vaccine efficacy”については、(4)式において、対立仮説の期待値をπ_1=0.6 として、必要イベント数が 150 となる帰無仮説値π_0を算出した。この条件を満たすπ_0は 0.285(必要イベント数は 150)で算出された(優越性試験①)。別に、π_1=0.6、π_0＝0.3 とおくと必要イベント数は 160 になる(優越性試験②)。これら 2 つの優越性試験①、②を検討対象とした。
・非劣性試験：3 節で算出した非劣性マージンのもとでの非劣性試験の必要イベント数 53 の非劣性試験を“the 48-event trial comparing two vaccines having approximate 90% efficacy”に近い試験であると位置づけた。
・試験期間：0.5 年、1 年の 2 通りで必要被験者数を算出した。

上記の条件設定のもと優越性試験と非劣性試験における必要被験者数を算出し、結果を表 5 に示す。

表 5. 優越性試験と非劣性試験の必要被験者数

プラセボ群の発現率(件/人年)	ワクチン群の発現率(件/人年)		必要イベント数	必要被験者数(2 群合計)	
				試験期間 0.5 年	試験期間 1 年
0.01	0.001	優越性試験①	150	54,546	27,274
		優越性試験②	160	58,182	29,092
		非劣性試験	53	106,000	53,000
0.05	0.005	優越性試験①	150	10,910	5,456
		優越性試験②	160	11,638	5,820
		非劣性試験	53	21,200	10,600
0.1	0.01	優越性試験①	150	5,456	2,728
		優越性試験②	160	5,820	2,910
		非劣性試験	53	10,600	5,300

優越性試験①：帰無仮説値π_0＝0.285、対立仮説の期待値π_1=0.6
優越性試験②：帰無仮説値π_0 =0.3、対立仮説の期待値π_1=0.6
非劣性試験：非劣性マージン$r_0 = M_2$＝2.6629、帰無仮説値$\pi_0' = -1.6629$、対立仮説の期待値$\pi_1' = 0$

表 5 の結果が示すように、試験期間が 0.5 年(1 年)の場合、非劣性試験では 53 イベントを集積するために、プラセボ群の発現率(件/人年)が 0.01、0.05 および 0.1 のとき、必要な被験者数はそれぞれ 106,000 例(53,000 例)、

21,200 例(10,600 例)および 10,600 例(5,300 例)と算出された。これらの必要被験者数は、優越性試験のおよそ 2 倍であることが読み取れる。すなわち、ワクチン有効性が高い(例えば 90%)状況では、非劣性試験ではプラセボよりもイベント発現率が顕著に低い(例えば、プラセボの 10 分の 1)2 つのワクチンを用いるため、必要イベント数に到達するまでに多くの観察人年が必要であると考察した。

5. まとめ

本稿では、新規ワクチン群と対照薬群の非劣性試験の計画において、リスク比を治療効果の評価指標とした場合の非劣性試験のデザインについて検討した。

具体的には、恒常性の仮定を考慮し、文献で公表されている複数の過去試験のメタアナリシスにおいて、変量効果モデルの当てはめを行い、非劣性マージンを算出した。本稿では、変量効果モデルとして、GLIMMIX プロシジャ(SAS Ver.9.4)を用いて、Bakbergenuly et al.(2019)の二項一般化線形混合モデルで最も単純なモデルを適用した。Bakbergenuly et al.(2019) においては、他にもいくつかの変量効果モデルが提案されているが、モデルが複雑になるにつれパラメータ数が増加し、推定が困難になる場合が考えられる。そのため、今回はパラメータ数が少なく、最も単純な二項一般化線形混合モデルを適用した。

また、上記で算出した非劣性マージンの設定を行う上で、我々は、①対照薬 vs. プラセボ、②プラセボ vs. 対照薬、③試験薬 vs. 対照薬の 3 つの段階に分け、原尺度上、対数尺度上の双方での解釈を行った。FDA ガイドライン(2016)においては、数式を用いない文言のみの表記で非劣性マージンの算出が説明されているため、解釈しづらい箇所が散見される。実際に非劣性マージンを導出する際は、本稿における図解の通り、算出された値がどの尺度であり、①、②、③のどの段階の値を指しているのかを理解しておく必要がある。
また、算出した非劣性マージンを用いて、必要イベント数・必要被験者数を決定する際は、浦狩(2025)で述べられている条件付き法の手法を適用した。

ここでは、ワクチン開発における非劣性試験を取り上げたが、リスク比の尺度を用いた試験全般に本稿で示した手順は適用が可能である。FDA ガイドラインには、そのような薬剤として、抗凝固剤と急性心筋梗塞の治療薬が挙げられているが、他にも該当する薬剤が存在すると考えている。

参考文献

- Senn S.(2022). The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy. Pharmaceutical Statistics, 21,790-807.
- FDA Guidance Document (2016). Non-Inferiority Clinical Trials to Establish Effectiveness. https://www.fda.gov/media/78504/download
- Fleming TR, Krause PR, Nason M, Longini IM, Henao-Restrepo AM. (2021). COVID-19 vaccine trials: The use of active controls and non-inferiority studies. Clin Trials, 18(3): 335-342.
- Bakbergenuly I, Hoaglin DC, Kulinskaya E, et al. (2017). Pitfalls of using the risk ratio in meta-analysis. Res Synth Methods, 10(3): 398-419.
- 浦狩保則(2025). ワクチン試験での必要イベント数の設定について. SAS ユーザー総会.
- 原めぐみ, 大藤さとこ, 福島若葉, 廣田良夫(2013). ワクチン疫学研究の原理と方法：新型インフルエンザワクチンの免疫原性と有効性の評価. 日本衛生学会
- 渡邉直昭, 大島信治, 永井英明, 加藤貴史, 齋藤美奈子, 五十嵐彩夏, 鈴川真穂, 川島正裕, 浅里功, 松井弘稔(2018). 成人における 4 価インフルエンザ不活化スプリットワクチンの免疫原性と安全性の検討, 日本呼吸器学会誌, 7 (4), 197-203.

・ICH-E10(2001). 臨床試験における対照群の選択とそれに関連する諸問題.
https://www.pmda.go.jp/files/000156634.pdf
・荒木康弘(2021). これからの新型コロナウイルスワクチンの開発・評価について. 学術の動向, 第 26 巻第 10 号, 51-57.
・厚生労働省(2021). ICMRA 活動と今後の開発・評価について. ワクチンに関する説明会資料(令和 3 年 7 月 15 日).

層別 Newcombe 信頼区間による割合の群間差の非劣性検定における欠測データに対する多重代入法について

○ 多田 由希夫，長谷川 貴大
（塩野義製薬株式会社）

Multiple imputation for missing data in testing non-inferiority of group differences between binomial proportions by stratified Newcombe confidence intervals

Yukio Tada, Takahiro Hasegawa
Shionogi & Co., Ltd.

要旨

欠測がある二値データの結果変数に対して，Mantel-Haenszel 法による割合の群間差と層別 Newcombe 信頼区間を用いて非劣性検定を行う場合を想定して，完全ケースの解析と Missing at Random (MAR) を仮定した多重代入法による有効標本サイズに基づく方法 (Li, Mehrotra, and Barnard (2006)) の第 I 種の過誤確率と検出力をシミュレーションにより確認し，性能評価を行う．

キーワード：Mantel-Haenszel 法，層別 Newcombe 信頼区間，非劣性検定，多重代入法，有効標本サイズ

1. 背景

臨床試験では，割合の比較が行われることがある．たとえば，「新型コロナウイルス（SARS-CoV-2）ワクチンの評価に関する考え方 (補遺４)」 [1]では，変異株ワクチン及び新たな有効成分を用いた追加接種用ワクチンの免疫原性を評価する臨床試験において，中和抗体価の幾何平均値 (GMT) 及び免疫応答割合を主要評価項目として，既承認の SARS-CoV-2 ワクチンに対する非劣性を検証することが推奨されている．免疫応答割合は，接種前後で中和抗体価が一定以上の倍率となった者の割合で示すことが一般的であるが，その定義を事前に設定することが求められている．これらの臨床試験では，治験薬群間で結果に強い影響を与える要因のバランスをとるため，層別無作為化を行う場合がある．このとき，抗体応答割合に対する効果の要約指標として，Mantel-Haenszel 法による割合の群間差を算出して，その信頼区間を層別 Newcombe 法で算出することが選択肢として考えられる [2].

「臨床試験のための統計的原則 補遺 臨床試験における estimand と感度分析」 [3] (以下，ICH E9 (R1)) では，治療開始後に発現し，関心のある臨床的疑問に関連する測定値の解釈や測定値の有無に影響を及ぼす事象を中間事象と定義している．例えば，感染が起きると中和抗体価が上昇するため，治験薬接種後の中和抗

体価を評価する主要評価時点よりも前の感染は中間事象となりうる．本検討では，中間事象がなかった状況を想定 (仮想ストラテジー) した効果を推定するため，中間事象以降の中和抗体価を意味のないデータとして解析に用いない場合を考える．なお，ICH E9 (R1) では，中間事象の発現により意味があると見なせないデータと欠測を区別すべきとしているが，本検討では，便宜上，中間事象以降の解析に用いないデータについても欠測と表現することとする．

このような欠測が存在する場合の解析方法として，結果変数が観察されている参加者だけを対象とする完全ケースの解析が考えられる．完全ケースの解析は既存の解析方法をそのまま適用でき，実施が容易である一方で，解析が妥当である前提として，欠測メカニズムに Missing Completely at Random (MCAR) を仮定している．ICH E9 (R1) は，このような統計モデルで用いる仮定からのずれに対して，主たる推定量によって算出された推定値の安定性を感度解析によって評価することを推奨している．

本検討では，感度解析として，欠測メカニズムに Missing at Random (MAR) を仮定して，多重代入法により欠測を補完した解析を考える．多重代入法は補完値の生成，疑似的な完全データの解析，解析結果の統合の 3 つのステップで構成され，解析結果の統合では，いわゆる Rubin のルールを適用する [4]．Rubin のルールはワルド型の検定統計量を前提としており，層別 Newcombe 法などスコア法に基づく信頼区間に対して，Rubin のルールを直接的に適用できない課題がある．このため，スコア法に基づく信頼区間への対応について，複数の方法が提案されている．そのうち，比較的容易に適用できる方法として，多重代入法の有効標本サイズに基づく方法 (Multiple Imputation based on effective sample size proposed by Li, Mehrotra and Barnard; MI-LMB) が提案されている [5]．本検討では，Mantel-Haenszel 法による割合の群間差と層別 Newcombe 信頼区間で非劣性検定を行う場合の欠測への対応法として, 完全ケースの解析と MI-LMB で適用する方法を紹介し，それぞれの第 I 種の過誤確率，検出力をシミュレーションにより確認し，性能評価を行う．

第 2 章は Mantel-Haenszel 法による割合の群間差と層別 Newcombe 信頼区間，多重代入法と MI-LMB の説明を行う．また，SAS での実装例も紹介する．第 3 章はシミュレーションの想定について説明する．第 4 章はシミュレーション結果について説明する．第 5 章はまとめと今後の課題について述べる．

2. 層別解析による非劣性検定と欠測への対応

本検討では，層別無作為化臨床試験で，抗体応答割合を評価項目として，対照薬に対する被験薬の非劣性検定を行う場合を想定する．非劣性の基準として，抗体応答割合の群間差の 95%信頼区間の下限が非劣性マージンを上回るときとする．このとき，効果の要約指標として Mantel-Haenszel 法による割合の群間差，及びその 95%信頼区間として層別 Newcombe 信頼区間に注目する．

2.1. Mantel-Haenszel 法による割合の群間差

層$i(=1,\cdots,I)$，投与群$j(=1$: 被験薬群, 2: 対照薬群)，参加者$k(=1,\cdots,n_{ij})$の抗体応答有無を$Y_{ijk}(=1$:有, 0: 無)とする．層iの Mantel-Haenszel 型の重みを$w_i = n_{i1}n_{i2}/(n_{i1}+n_{i2})$として，Mantel-Haenszel 法による割合の群間差は$\hat{d}_{MH} = \sum_{i=1}^{I} w_i\hat{d}_i / \sum_{i=1}^{I} w_i$で算出される．ここで，$\hat{d}_i$は層$i$，投与群$j$の割合の推定値$\hat{p}_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk}/n_{ij}$の群間差として，$\hat{d}_i = \hat{p}_{i1} - \hat{p}_{i2}$で算出される．

2.2. 層別 Newcombe 信頼区間

割合に対する Wilson 信頼区間に基づいて算出する方法として, Newcombe 信頼区間は提案された. その後, Yan and Su (2010) [6]では，層別解析に適用するため，割合に対する層別 Wilson 信頼区間に基づいて，層別 Newcombe 信頼区間を提案した．層別 Newcombe 信頼区間の$100(1-\alpha)\%$ 信頼下限と信頼上限(L, U)は，$z_{\alpha/2}$を標準正規分布の$100(1-\alpha/2)\%$点として，

$$\begin{aligned} L &= \hat{d}_{MH} - z_{\alpha/2}\sqrt{\lambda_1 L_1(1-L_1) + \lambda_2 U_2(1-U_2)}, \\ U &= \hat{d}_{MH} + z_{\alpha/2}\sqrt{\lambda_1 U_1(1-U_1) + \lambda_2 L_2(1-L_2)} \end{aligned} \tag{1}$$

である.

ここで，λ_1，λ_2，L_1，U_1，L_2，U_2は

$$\begin{aligned} \lambda_1 = \sum_{i=1}^{I} w_i^2 \Big/ n_{i1}\,, \lambda_2 = \sum_{i=1}^{I} w_i^2 \Big/ n_{i2}, \\ L_1 = \sum_{i=1}^{I} w_i L_{i1}\,, U_1 = \sum_{i=1}^{I} w_i U_{i1}, \\ L_2 = \sum_{i=1}^{I} w_i L_{i2}\,, U_2 = \sum_{i=1}^{I} w_i U_{i2} \end{aligned} \tag{2}$$

であり，L_{i1}とU_{i1}はp_{i1}に関する方程式

$$(\hat{p}_{i1} - p_{i1})^2 = z_{\gamma 1}^2 \frac{p_{i1}(1-p_{i1})}{n_{i1}}, z_{\gamma 1} = \frac{\sqrt{\sum_{i=1}^{I} w_i^2 \frac{\hat{p}_{i1}(1-\hat{p}_{i1})}{n_{i1}}}}{\sum_{i=1}^{I} w_i \sqrt{\frac{\hat{p}_{i1}(1-\hat{p}_{i1})}{n_{i1}}}} z_{\alpha/2} \tag{3}$$

の解であり，L_{i2}とU_{i2}はp_{i2}に関する方程式

$$(\hat{p}_{i2} - p_{i2})^2 = z_{\gamma 2}^2 \frac{p_{i2}(1-p_{i2})}{n_{i2}}, z_{\gamma 2} = \frac{\sqrt{\sum_{i=1}^{I} w_i^2 \frac{\hat{p}_{i2}(1-\hat{p}_{i2})}{n_{i2}}}}{\sum_{i=1}^{I} w_i \sqrt{\frac{\hat{p}_{i2}(1-\hat{p}_{i2})}{n_{i2}}}} z_{\alpha/2} \tag{4}$$

の解である．なお，$L_{i1} < U_{i1}, L_{i2} < U_{i2}$とする.

佐々木，舟尾 (2017) [7]は，シミュレーションによる性能評価の結果として，Mantel-Haenszel 法による割合の群間差に対する層別 Newcombe 信頼区間の被覆確率が名目の水準に維持されることを報告した.

2.3. 不完全データの統計解析

一部参加者の結果変数が欠測 (評価不能として欠測扱いとする場合を含む) であるとき，結果変数が欠測でない参加者のみを対象とする完全ケースの解析，欠測データを補完値で代入して解析を行う方法の適用が考えられる．完全ケースの解析は適用が容易であるものの，欠測が MCAR でない場合，推定値にバイアスが生じる可能性がある．また，結果変数の予測に有用な補助的な変数が存在する場合，完全ケースの解析は推定精度の改善に役立つ可能性のあるこれらの情報を利用しないため, 効率性の点でも課題がある. このため，本検討では欠測メカニズムに MAR を仮定した多重代入法を考える．この中でも，スコア法に基づく信頼区間への対応が可能な有効標本サイズに基づく方法 (MI-LMB) に注目する.

2.3.1. 多重代入法による欠測値の補完に基づく解析

多重代入法は，単一の補完値でなく，複数の補完値を代入することで，生成された補完値の予測誤差に基

づく不確実性を適切に考慮した推測を行うための方法である．その目的は，バイアスのない推定量を得るとともに，関連する仮説検定の第 I 種の過誤確率や信頼区間の被覆確率が名目の水準を維持することである．多重代入法は，MAR の欠測メカニズムのもとで妥当性が担保される [4].

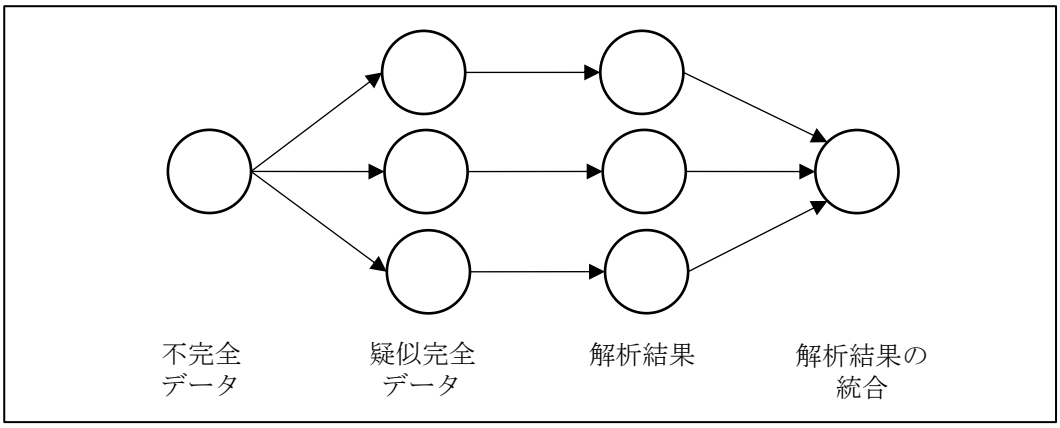


図 1　多重代入法の手順

多重代入法のアルゴリズムを図 1 に示した．まず，不完全データに対して，補助的な変数を説明変数として含むなんらかの$m(>1)$個の疑似的な完全データを作成する (補完値の生成)．次に，それぞれのデータセットに対して，解析を行う (疑似的な完全データに対する解析)．最後に，m個の結果を統合して，最終的な推定値と標準誤差を得る (解析結果の統合).

2.3.2. 多重代入法の有効標本サイズに基づく方法 (MI-LMB)

Li, Mehrotra, and Barnard (2006) [5]は，多重代入法の中で疑似完全データから算出されたスコア法に基づく信頼区間を統合するため，多重代入法の有効標本サイズに基づく算出方法を提案した．以下，層別無作為化を想定して，各層の各群に対する多重代入法の有効標本サイズの導出を示す.

m個の疑似的な完全データでの層i，投与群jの割合の推定値を$\hat{p}_{ij}^{(d)}(d=1,\dots,m)$としたとき，Rubin のルールより，統合された推定値$\bar{p}_{ij}$，代入値内分散の平均$\bar{W}_{ij}$，代入値間分散の推定値B_{ij}，推定量の全分散T_{ij}は，

$$\bar{p}_{ij}=\frac{1}{m}\sum_{d=1}^{m}\hat{p}_{ij}^{(d)}$$

$$\bar{W}_{ij}=\frac{1}{m}\sum_{d=1}^{m}\frac{\hat{p}_{ij}^{(d)}(1-\hat{p}_{ij}^{(d)})}{n_{ij}-1}$$

$$B_{ij}=\frac{1}{m-1}\sum_{d=1}^{m}(\hat{p}_{ij}^{(d)}-\bar{p}_{ij})^2$$

$$T_{ij}=\bar{W}_{ij}+\left(1+\frac{1}{m}\right)B_{ij}$$

となり，層iの投与群jにおける多重代入法の有効標本サイズ$n_{ij,MI}$は$\bar{p}_{ij}$，T_{ij}を用いて，

$$n_{ij,MI}=\frac{\bar{p}_{ij}(1-\bar{p}_{ij})}{T_{ij}}$$

として得られる.

MI-LMB を層別解析に適用した先行研究として，Lu and Guo [8]は，MI-LMB 法を適用して，Mantel-Haenszel 法による割合の群間差と層別 Miettinen-Nurminen 法による信頼区間を計算した．本検討では，彼らと同様，群と層の各組み合わせに対する標本サイズと割合として，$\{(n_{ij,MI}, \bar{p}_{ij}): i = 1, \dots, I, j = 1,2\}$を用いて，MI-LMB 法で統合された Mantel-Haenszel 法による割合の群間差，層別 Newcombe 信頼区間を計算する．すなわち，層iの Mantel-Haenszel 型の重みを$w_{i,MI} = n_{i1,MI}n_{i2,MI}/(n_{i1,MI} + n_{i2,MI})$として，Mantel-Haenszel 法による割合の群間差を$\bar{d}_{MH} = \sum_{i=1}^{I} w_{i,MI}\bar{d}_i / \sum_{i=1}^{I} w_{i,MI}$で算出する．ここで，$\bar{d}_i$は層$i$，投与群$j$の割合の推定値$\bar{p}_{ij}$の群間差として，$\bar{d}_i = \bar{p}_{i1} - \bar{p}_{i2}$で算出する．また，式(2)，(3)，(4)の$w_i$を$w_{i,MI}$，$\hat{p}_{ij}$を$\bar{p}_{ij}$でそれぞれ置き換えて計算した$\lambda_1$，$\lambda_2$，$L_1$，$U_1$，$L_2$，$U_2$を用いて，層別 Newcombe 信頼区間$(L, U)$は

$$L = \bar{d}_{MH} - z_{\alpha/2}\sqrt{\lambda_1 L_1(1 - L_1) + \lambda_2 U_2(1 - U_2)},$$
$$U = \bar{d}_{MH} + z_{\alpha/2}\sqrt{\lambda_1 U_1(1 - U_1) + \lambda_2 L_2(1 - L_2)}$$

で算出する．

2.3.3. MI-LMB の SAS での実装例

ここで，仮想のデータを用いて，MI-LMB の SAS での実装例を示す．

表 1　不完全データ

SUBJID	STRATA	TREAT	X	Y	CRITFN
101	1	1	1.56	0.96	1
102	1	2	0.76	0.49	1
103	1	2	0.07	-0.64	0
104	2	1	0.92	欠測	欠測
…	…	…	…	…	…

表 1 のデータは，層 $(STRATA)$，治験薬群 $(TREAT)$，補助変数 (X)，結果変数 (Y)，Yが 0 以上または 0 未満によって反応の有無を示す二値データ $(CRITFN)$ で構成され，変数$STRATA$と変数$TREAT$は欠測がなく，変数X，変数Yの欠測パターンは単調とする．つまり，変数Xが欠測のときは変数Yも欠測とする．この不完全データに対して，疑似完全データの生成から各群・各層の統合された二値結果変数の割合と有効標本サイズの計算および層別解析は以下の SAS コードで実行可能である．

Step 1. 変数X, Yの欠測データに対する回帰モデルを用いた疑似完全データの生成

```
proc mi data = SAMPLE seed = 12345 nimpute = 100 out = IMP_SAMP;
  by TREAT;
  class STRATA;
  monotone reg(X = STRATA);
  monotone reg(Y = STRATA X);
  var STRATA X Y;
run;
```

Step 2. 疑似完全データの整備 (欠測代入した変数*Y*の二値データ化)

```
data IMP_SAMP2;
  set IMP_SAMP;
  if Y >= 0 then CRITFN = 1; else CRITFN = 0;
  TOTAL = 1;
run;
```

Step 3. 疑似完全データの解析 (各群・各層の二値結果変数の割合とその標準誤差の算出)

```
proc summary data = IMP_SAMP2 nway;
  class STRATA TREAT _IMPUTATION;
  var CRITFN TOTAL;
  output out = RES_SE0 sum =;
run;
data RES_SE;
  set RES_SE0;
  ESTIMATE = CRITFN/TOTAL;
  STDERR = sqrt(ESTIMATE * (1-ESTIMATE) / (TOTAL-1));
run;
```

Step 4. 解析結果の統合 (各群・各層の二値結果変数の割合および標準誤差の統合)

```
proc mianalyze data=RES_SE;
  by STRATA TREAT;
  modeleffects ESTIMATE;
  stderr STDERR;
  ods output ParameterEstimates = RES_SE_MI;
run;
```

Step 5. 有効標本サイズの算出

```
data N_MI;
  set RES_SE_MI;
  N_MI = (ESTIMATE * (1-ESTIMATE)) / (STDERR ** 2);
run;
```

Step 6. 統合した割合と有効標本サイズに対する分割表データの生成

```
data SAMP_FRQ_MI;
  set N_MI;
  CRITFN = 1;
  COUNT = N_MI * ESTIMATE; output;
```

```
  CRITFN = 0;
  COUNT = N_MI * (1-ESTIMATE); output;
run;
```

Step 7. 統合した割合と有効標本サイズに対する層別解析

```
proc freq data = SAMP_FRQ_MI;
  tables STRATA * TREAT * CRITFN / riskdiff(common cl=newcombe column = 2);
  weight COUNT;
run;
```

3. シミュレーションの想定

被験薬群と対照薬群へ 1:1 で割り付ける層別無作為化臨床試験で，二値結果変数の応答ありの割合 (以下，応答割合) に対して，Mantel-Haenszel 法による群間差と層別 Newcombe 法による 95%信頼区間を用いて，非劣性マージンを-0.1 として非劣性検定を行うことを想定した．層の数は 2 として，各層の割合は 1:1，1:2 とした．応答割合は治験薬投与後の時点 1 と時点 2 の 2 時点で観察されるものとして，時点 2 の応答割合を主要評価とした．時点 2 の応答割合について，対照薬群は 0.8 または 0.9 として，対立仮説下では被験薬群と対照薬群は同じ値，帰無仮説下では被験薬群は対照薬群より 0.1 を減じた 0.7 または 0.8 とした．各層の割合が 1:1 のとき，層ごとの応答割合は，層 1 では群全体よりも 0.02 小さい値，層 2 では群全体よりも 0.02 大きい値として，各層の割合が 1:2 のとき，層ごとの応答割合は，層 1 では群全体よりも 0.02 小さい値，層 2 では群全体よりも 0.01 大きい値とした．また，層の影響が大きい状況として，対照薬群の応答割合を 0.8 として，各層の割合が 1:1 のとき，層 1 では群全体よりも 0.12 小さい値，層 2 では群全体よりも 0.12 大きい値として，各層の割合が 1:2 のとき，層 1 では群全体よりも 0.12 小さい値，層 2 では群全体よりも 0.06 大きい値の場合も検討した．時点 1 の応答割合は，時点 2 よりも 0.05 大きい値とした．標本サイズは時点 2 の応答割合を 0.8 または 0.9 としたときに，検出力 90%で非劣性を検証するために必要な参加者数に 10%の欠測による解析除外を考慮して，それぞれ 1 群あたり 228 または 378 とした．なお，層i，投与群j，時点tの応答割合を$p_{t,ij}$とするとき，$(\Phi^{-1}(p_{1,ij}), \Phi^{-1}(p_{2,ij}))$を平均値，$\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$を共分散行列とする多変量標準正規分布に従う連続量の変数(X_{ij}, Y_{ij})が存在すると想定する．ここで，Φは標準正規分布の分布関数とする．これは，二値の抗体応答に対する連続量の抗体価との関係に相当する．このとき，$X_{ij} \geq 0$，$Y_{ij} \geq 0$がそれぞれ時点 1，2 の応答ありに対応すると考えることができ，$\Pr(X_{ij} \geq 0)$，$\Pr(Y_{ij} \geq 0)$がそれぞれ時点 1，2 の応答割合に対応する．

次に，感染以降の抗体価を欠測として取り扱うため，各時点の抗体価の観察有無 (1:観察，0:欠測) を識別する変数(R_X, R_Y)を導入する．ここでは，感染による欠測以外の欠測は考えず，単調な欠測パターンである(0,0)または(1,0)のみを考える．(0,0)の欠測はベースラインと時点 1 の間の感染に相当する．治験薬投与直後の抗体価は感染予防に十分ではなく，当該期間の感染による(0,0)の欠測はランダムと仮定して，(0,0)の発現確率$p_{(0,0)}$を$p_{(0,0)} = 0.03$，0.06，0.09と想定した．次に，(1,0)の欠測は時点 1 と時点 2 の間の感染に相当するが，この期間の感染は時点 1 の抗体価Xに依存する場合を考慮して，$\eta = 0$，0.3，0.6，0.9として，(1,0)の発現確

率$p_{(1,0)}$を$p_{(1,0)} = p_{(0,0)} * (1 - \eta * (X \geq 0))$と想定した．$\eta > 0$のとき，観察された$X$を条件付けることで時点 2 の抗体価$Y$の欠測はランダムと考えられ MAR に対応する．また，$\eta = 0$のとき，時点 2 の抗体価Yの欠測は観察されたXに依存せず MCAR に対応する．シミュレーションのデータ生成の概要を図 2，パラメータの想定を表 2 にそれぞれ示した．なお，シミュレーションでは，いずれの群でも 1 件以上の欠測がある生成データを検討対象とした．欠測がない群と層の組み合わせが生じうるが，その場合，当該群と層の組み合わせの標本サイズと割合はそれぞれ観察された値を用いた．また，欠測があった群と層の組み合わせのうち，多重代入法の結果，統合した割合が 0 または 1 であったとき，有効標本サイズが算出できない．この場合，便宜的に，当該組み合わせについて，欠測がなかったときの参加者数から 1 を引いた値を標本サイズとして用いた．

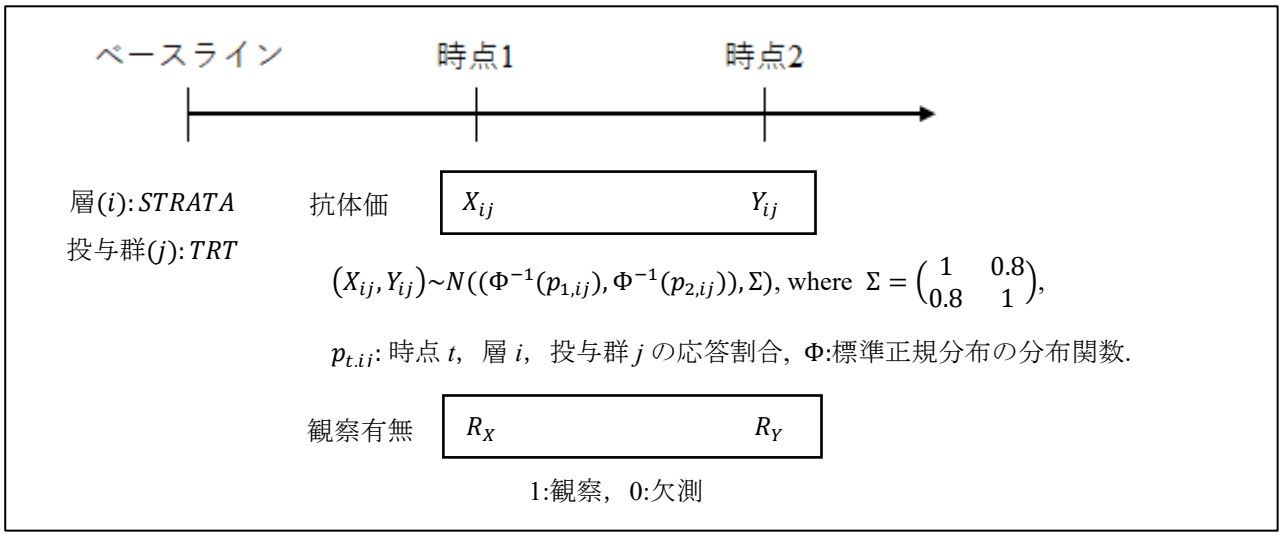


図 2　シミュレーションのデータ生成の概要

欠測に対する解析方法として，完全ケースの解析と MI-LMB を適用した．多重代入法の補完値の生成は，群ごとに，逐次的に回帰モデルによる予測モデルを構築して行った．すなわち，層の変数を$STRATA$として，時点 1 の抗体価Xの欠測データは，回帰モデル$X|STRATA, \boldsymbol{\beta} \sim N(\beta_0 + \beta_1 STRATA, \sigma^2)$で補完値を生成したのち，時点 2 の抗体価$Y$の欠測データは，回帰モデル$Y|STRATA, X, \boldsymbol{\beta} \sim N(\beta_0 + \beta_1 STRATA + \beta_2 X, \sigma^2)$で補完値の生成を行った．補完値の組の数はシミュレーションの実行時間を考慮して$m = 50$ とした．

各解析方法の性能評価の指標として，群間差の推定値に関するバイアス，第 I 種の過誤確率，検出力をそれぞれ算出した．ここで，バイアスはシミュレーションによる群間差の平均値と想定したパラメータに基づく真の群間差との差として，第 I 種の過誤確率は帰無仮説下で両側 95%信頼区間の下限が非劣性マージンを超えた割合として，検出力はシミュレーションで得られた対立仮説下で両側 95%信頼区間の下限が非劣性マージンを超えた割合として，それぞれ算出した．シミュレーションの反復回数は 10,000 回とした．

表 2　シミュレーションのパラメータ

1	割付比	1:1	
2	各層の参加者数の比	1:1 または 1:2	
3	時点 2 の抗体応答割合 $[p_{2,ij}]$ $(p_{2,11}, p_{2,21})$ vs $(p_{2,12}, p_{2,22})$ $p_{2,11}$: 被験薬群の層 1 の真値 $p_{2,21}$: 被験薬群の層 2 の真値 $p_{2,12}$: 対照薬群の層 1 の真値 $p_{2,22}$: 対照薬群の層 2 の真値	各層の割合が(1/2, 1/2)のとき 帰無仮説 $(p_{2,i1} = p_{2,i2} - 0.1)$ ①　(0.78, 0.82) vs (0.88, 0.92) ②　(0.68, 0.72) vs (0.78, 0.82) ③　(0.58, 0.82) vs (0.68, 0.92) 対立仮説 $(p_{2,i1} = p_{2,i2})$ ④　(0.88, 0.92) vs (0.88, 0.92) ⑤　(0.78, 0.82) vs (0.78, 0.82) ⑥　(0.68, 0.92) vs (0.68, 0.92)	各層の割合が(1/3, 2/3)のとき 帰無仮説 $(p_{2,i1} = p_{2,i2} - 0.1)$ ①　(0.78, 0.81) vs (0.88, 0.91) ②　(0.68, 0.71) vs (0.78, 0.81) ③　(0.58, 0.76) vs (0.68, 0.86) 対立仮説 $(p_{2,i1} = p_{2,i2})$ ④　(0.88, 0.91) vs (0.88, 0.91) ⑤　(0.78, 0.81) vs (0.78, 0.81) ⑥　(0.68, 0.86) vs (0.68, 0.86)
4	時点 1 の抗体応答割合 $[p_{1,ij}]$	$p_{1,ij} = p_{2,ij}$+0.05	
5	時点間の抗体価の相関係数 $[\rho]$	0.8	
6	各群の標本サイズ [N]	228 (①, ④), 378 (②, ③, ⑤, ⑥)	
7	欠測パターンとその発現確率 (0, 0): 時点 1 と 2 いずれも欠測 (1, 0): 時点 2 のみ欠測	(0, 0)の発現確率: $p_{(0,0)}$ =0.03, 0.06, 0.09 (1, 0)の発現確率: $p_{(1,0)} = p_{(0,0)} * (1 - \eta * (X \geq 0))$, ここで，$\eta = 0, 0.3, 0.6, 0.9$で，$X$は時点 1 の抗体価. $\eta = 0$のとき MCAR，$\eta > 0$のとき MAR に相当.	
8	シミュレーション回数	10,000	

4. シミュレーション結果

帰無仮説下における完全ケースの解析と MI-LMB による第 I 種の過誤確率を表 3，表 4，表 5，表 6，表 7，表 8 に示した．第 I 種の過誤確率が，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えた場合，数値を太字で表示した．完全ケースの解析においては，欠測が治験薬投与後の時点 1 の観測値に依存する場合，第 I 種の過誤確率が 0.028 を超える傾向があった．MI-LMB においては，第 I 種の過誤確率はいずれの設定でも 0.028 を超えなかった．しかしながら，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での下側 2.5%点に対応する 0.022 以下の第 I 種の過誤確率が殆どであった．このため，MI-LMB 法は保守的と考えられた．一方，欠測が治験薬投与後の時点 1 の観測値に依存する場合，完全ケースの解析において，MI-LMB に比べて，群間差のバイアスが大きかった．

対立仮説下における完全ケースの解析と MI-LMB による検出力を表 9，表 10，表 11，表 12，表 13，表 14 に示した．MI-LMB において，完全ケースの解析に比べて，概ね検出力が大きかった．群間差のバイアスについて，解析方法による違いはなかった．

表 3　第 I 種の過誤確率 (層の参加者数の比=1:1, 対照群の応答割合=(0.88, 0.92))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	第 I 種の過誤確率	バイアス	第 I 種の過誤確率
228	MCAR	0.03	0.060	0.060	-0.0003	0.0226	-0.0002	0.0207
		0.06	0.120	0.120	-0.0004	0.0210	-0.0003	0.0177
		0.09	0.180	0.180	-0.0003	0.0220	-0.0002	0.0163
	MAR ($\eta = 0.3$)	0.03	0.052	0.051	0.0001	0.0239	-0.0003	0.0200
		0.06	0.105	0.103	0.0006	0.0228	-0.0003	0.0180
		0.09	0.157	0.154	0.0012	0.0241	-0.0002	0.0165
	MAR ($\eta = 0.6$)	0.03	0.045	0.043	0.0006	0.0244	-0.0002	0.0204
		0.06	0.089	0.086	0.0015	0.0255	-0.0003	0.0176
		0.09	0.134	0.129	0.0026	0.0266	-0.0002	0.0169
	MAR ($\eta = 0.9$)	0.03	0.037	0.034	0.0010	0.0247	-0.0002	0.0210
		0.06	0.074	0.069	0.0025	**0.0283**	-0.0002	0.0205
		0.09	0.111	0.103	0.0040	**0.0283**	-0.0002	0.0182

太字は，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えることを示す.

表 4　第 I 種の過誤確率 (層の参加者数の比=1:2, 対照群の応答割合=(0.88, 0.91))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	第 I 種の過誤確率	バイアス	第 I 種の過誤確率
228	MCAR	0.03	0.060	0.060	-0.0002	0.0246	-0.0001	0.0210
		0.06	0.120	0.120	-0.0002	0.0224	-0.0002	0.0174
		0.09	0.180	0.180	-0.0002	0.0212	-0.0001	0.0163
	MAR ($\eta = 0.3$)	0.03	0.052	0.051	0.0003	0.0248	-0.0001	0.0213
		0.06	0.105	0.103	0.0007	0.0242	-0.0002	0.0186
		0.09	0.157	0.154	0.0013	0.0246	-0.0001	0.0158
	MAR ($\eta = 0.6$)	0.03	0.045	0.043	0.0007	0.0253	-0.0001	0.0213
		0.06	0.090	0.086	0.0016	0.0265	-0.0002	0.0191
		0.09	0.134	0.129	0.0028	0.0273	-0.0001	0.0175
	MAR ($\eta = 0.9$)	0.03	0.037	0.034	0.0011	0.0264	-0.0001	0.0221
		0.06	0.074	0.069	0.0026	**0.0290**	-0.0001	0.0212
		0.09	0.111	0.103	0.0041	**0.0309**	-0.0001	0.0195

太字は，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えることを示す.

表 5　第 I 種の過誤確率 (層の参加者数の比=1:1, 対照群の応答割合=(0.78, 0.82))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	第 I 種の過誤確率	バイアス	第 I 種の過誤確率
378	MCAR	0.03	0.060	0.060	0.0001	0.0253	0.0002	0.0234
		0.06	0.120	0.120	0.0002	0.0242	0.0002	0.0197
		0.09	0.180	0.180	0.0003	0.0264	0.0003	0.0188
	MAR ($\eta = 0.3$)	0.03	0.053	0.052	0.0005	0.0258	0.0003	0.0235
		0.06	0.107	0.105	0.0009	0.0261	0.0002	0.0204
		0.09	0.160	0.157	0.0014	**0.0290**	0.0003	0.0196
	MAR ($\eta = 0.6$)	0.03	0.046	0.045	0.0009	0.0267	0.0003	0.0239
		0.06	0.093	0.089	0.0016	0.0270	0.0002	0.0208
		0.09	0.140	0.134	0.0025	**0.0320**	0.0003	0.0205
	MAR ($\eta = 0.9$)	0.03	0.040	0.037	0.0012	0.0277	0.0003	0.0240
		0.06	0.080	0.074	0.0023	**0.0294**	0.0001	0.0219
		0.09	0.120	0.111	0.0035	**0.0337**	0.0002	0.0206

太字は，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えることを示す.

表 6　第 I 種の過誤確率 (層の参加者数の比=1:1, 対照群の応答割合=(0.68, 0.92))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	第 I 種の過誤確率	バイアス	第 I 種の過誤確率
378	MCAR	0.03	0.060	0.060	0.0001	0.0212	0.0002	0.0195
		0.06	0.120	0.120	0.0001	0.0208	0.0001	0.0167
		0.09	0.180	0.180	0.0001	0.0198	0.0001	0.0165
	MAR ($\eta = 0.3$)	0.03	0.053	0.052	0.0005	0.0220	0.0002	0.0189
		0.06	0.107	0.105	0.0008	0.0226	0.0001	0.0170
		0.09	0.160	0.157	0.0012	0.0247	0.0001	0.0167
	MAR ($\eta = 0.6$)	0.03	0.047	0.045	0.0008	0.0226	0.0002	0.0199
		0.06	0.093	0.089	0.0015	0.0242	0.0001	0.0177
		0.09	0.140	0.134	0.0023	0.0264	0.0001	0.0160
	MAR ($\eta = 0.9$)	0.03	0.040	0.037	0.0012	0.0233	0.0002	0.0199
		0.06	0.080	0.074	0.0022	0.0261	0.0001	0.0180
		0.09	0.120	0.111	0.0034	**0.0283**	0.0001	0.0166

太字は，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えることを示す.

表 7　第 I 種の過誤確率 (層の参加者数の比=1:2, 対照群の応答割合=(0.78, 0.81))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	第 I 種の過誤確率	バイアス	第 I 種の過誤確率
378	MCAR	0.03	0.060	0.060	0.0001	0.0247	0.0002	0.0241
		0.06	0.120	0.120	0.0001	0.0248	0.0002	0.0208
		0.09	0.180	0.180	0.0002	0.0267	0.0002	0.0190
	MAR ($\eta = 0.3$)	0.03	0.053	0.052	0.0005	0.0263	0.0002	0.0245
		0.06	0.107	0.105	0.0009	0.0271	0.0002	0.0203
		0.09	0.160	0.157	0.0014	**0.0291**	0.0002	0.0201
	MAR ($\eta = 0.6$)	0.03	0.046	0.045	0.0009	**0.0284**	0.0002	0.0252
		0.06	0.093	0.089	0.0016	**0.0286**	0.0002	0.0218
		0.09	0.140	0.134	0.0025	**0.0312**	0.0002	0.0199
	MAR ($\eta = 0.9$)	0.03	0.040	0.037	0.0012	0.0278	0.0003	0.0256
		0.06	0.080	0.074	0.0023	**0.0309**	0.0002	0.0212
		0.09	0.120	0.111	0.0035	**0.0332**	0.0002	0.0206

太字は，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えることを示す.

表 8　第 I 種の過誤確率 (層の参加者数の比=1:2, 対照群の応答割合=(0.68, 0.86))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	第 I 種の過誤確率	バイアス	第 I 種の過誤確率
378	MCAR	0.03	0.060	0.060	0.0003	0.0244	0.0004	0.0217
		0.06	0.120	0.120	0.0002	0.0228	0.0003	0.0188
		0.09	0.180	0.180	0.0003	0.0235	0.0003	0.0183
	MAR ($\eta = 0.3$)	0.03	0.053	0.052	0.0006	0.0245	0.0004	0.0219
		0.06	0.107	0.105	0.0010	0.0250	0.0003	0.0197
		0.09	0.160	0.157	0.0014	0.0261	0.0003	0.0186
	MAR ($\eta = 0.6$)	0.03	0.047	0.045	0.0010	0.0251	0.0004	0.0217
		0.06	0.093	0.089	0.0017	0.0263	0.0003	0.0195
		0.09	0.140	0.134	0.0025	**0.0286**	0.0003	0.0189
	MAR ($\eta = 0.9$)	0.03	0.040	0.037	0.0014	0.0252	0.0004	0.0221
		0.06	0.080	0.074	0.0024	0.0273	0.0003	0.0212
		0.09	0.120	0.111	0.0035	**0.0317**	0.0002	0.0197

太字は，反復回数 10,000 回で 2.5%の生起確率の二項分布の下での上側 2.5%点に対応する 0.028 を超えることを示す.

表 9　検出力 (層の参加者数の比=1:1, 対照群の応答割合=(0.88, 0.92))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	検出力	バイアス	検出力
228	MCAR	0.03	0.060	0.060	-0.0004	0.9178	-0.0004	0.9255
		0.06	0.120	0.120	-0.0004	0.9026	-0.0005	0.9154
		0.09	0.180	0.180	-0.0004	0.8835	-0.0004	0.9067
	MAR ($\eta = 0.3$)	0.03	0.052	0.051	-0.0004	0.9225	-0.0004	0.9258
		0.06	0.103	0.103	-0.0004	0.9090	-0.0004	0.9171
		0.09	0.155	0.154	-0.0004	0.8937	-0.0004	0.9089
	MAR ($\eta = 0.6$)	0.03	0.043	0.043	-0.0004	0.9256	-0.0004	0.9267
		0.06	0.086	0.086	-0.0004	0.9163	-0.0004	0.9196
		0.09	0.129	0.129	-0.0004	0.9027	-0.0005	0.9115
	MAR ($\eta = 0.9$)	0.03	0.034	0.034	-0.0004	0.9281	-0.0004	0.9269
		0.06	0.069	0.069	-0.0004	0.9203	-0.0004	0.9217
		0.09	0.103	0.103	-0.0004	0.9139	-0.0004	0.9154

表 10　検出力 (層の参加者数の比=1:2, 対照群の応答割合=(0.88, 0.91))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	検出力	バイアス	検出力
228	MCAR	0.03	0.060	0.060	-0.0004	0.9185	-0.0004	0.9262
		0.06	0.120	0.120	-0.0004	0.9039	-0.0004	0.9150
		0.09	0.180	0.180	-0.0004	0.8850	-0.0004	0.9082
	MAR ($\eta = 0.3$)	0.03	0.052	0.051	-0.0004	0.9225	-0.0004	0.9270
		0.06	0.103	0.103	-0.0004	0.9091	-0.0004	0.9172
		0.09	0.155	0.154	-0.0004	0.8954	-0.0004	0.9107
	MAR ($\eta = 0.6$)	0.03	0.043	0.043	-0.0004	0.9262	-0.0004	0.9280
		0.06	0.086	0.086	-0.0004	0.9158	-0.0004	0.9204
		0.09	0.129	0.129	-0.0004	0.9058	-0.0004	0.9126
	MAR ($\eta = 0.9$)	0.03	0.034	0.034	-0.0004	0.9284	-0.0004	0.9290
		0.06	0.069	0.069	-0.0004	0.9209	-0.0004	0.9217
		0.09	0.103	0.103	-0.0004	0.9142	-0.0004	0.9152

表 11　検出力 (層の参加者数の比=1:1, 対照群の応答割合=(0.78, 0.82))

N	欠測メカニズム	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
			被験薬群	対照薬群	バイアス	検出力	バイアス	検出力
378	MCAR	0.03	0.060	0.060	-0.0001	0.9188	0.0000	0.9276
		0.06	0.120	0.120	-0.0001	0.8997	-0.0001	0.9176
		0.09	0.180	0.180	0.0000	0.8773	0.0000	0.9061
	MAR ($\eta = 0.3$)	0.03	0.052	0.052	0.0000	0.9226	0.0000	0.9267
		0.06	0.105	0.105	-0.0001	0.9071	-0.0001	0.9182
		0.09	0.157	0.157	0.0000	0.8908	0.0000	0.9090
	MAR ($\eta = 0.6$)	0.03	0.045	0.045	0.0000	0.9258	0.0000	0.9276
		0.06	0.090	0.089	0.0000	0.9129	0.0000	0.9191
		0.09	0.134	0.134	0.0000	0.8993	-0.0001	0.9103
	MAR ($\eta = 0.9$)	0.03	0.037	0.037	0.0000	0.9289	0.0000	0.9283
		0.06	0.074	0.074	0.0000	0.9209	-0.0001	0.9208
		0.09	0.112	0.111	0.0000	0.9107	-0.0001	0.9127

表 12　検出力 (層の参加者数の比=1:1, 対照群の応答割合=(0.68, 0.92))

N	欠測	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
	メカニズム		被験薬群	対照薬群	バイアス	検出力	バイアス	検出力
378	MCAR	0.03	0.060	0.060	0.0002	0.9263	0.0003	0.9344
		0.06	0.120	0.120	0.0003	0.9106	0.0002	0.9235
		0.09	0.180	0.180	0.0003	0.8887	0.0003	0.9140
	MAR ($\eta = 0.3$)	0.03	0.052	0.052	0.0003	0.9287	0.0003	0.9351
		0.06	0.105	0.105	0.0003	0.9171	0.0003	0.9266
		0.09	0.157	0.157	0.0003	0.8983	0.0003	0.9165
	MAR ($\eta = 0.6$)	0.03	0.045	0.045	0.0003	0.9333	0.0003	0.9365
		0.06	0.090	0.089	0.0003	0.9231	0.0002	0.9282
		0.09	0.134	0.134	0.0003	0.9089	0.0002	0.9179
	MAR ($\eta = 0.9$)	0.03	0.037	0.037	0.0003	0.9369	0.0003	0.9367
		0.06	0.074	0.074	0.0003	0.9283	0.0002	0.9291
		0.09	0.111	0.111	0.0003	0.9206	0.0002	0.9222

表 13　検出力 (層の参加者数の比=1:2, 対照群の応答割合=(0.78, 0.81))

N	欠測	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
	メカニズム		被験薬群	対照薬群	バイアス	検出力	バイアス	検出力
378	MCAR	0.03	0.060	0.060	-0.0001	0.9193	0.0000	0.9263
		0.06	0.120	0.120	-0.0001	0.9002	-0.0001	0.9174
		0.09	0.180	0.180	0.0000	0.8779	0.0000	0.9059
	MAR ($\eta = 0.3$)	0.03	0.052	0.052	0.0000	0.9227	0.0000	0.9266
		0.06	0.105	0.105	-0.0001	0.9068	0.0000	0.9190
		0.09	0.157	0.157	0.0000	0.8901	0.0000	0.9089
	MAR ($\eta = 0.6$)	0.03	0.045	0.045	0.0000	0.9256	0.0001	0.9283
		0.06	0.090	0.089	0.0000	0.9130	0.0000	0.9214
		0.09	0.134	0.134	0.0000	0.8994	0.0000	0.9119
	MAR ($\eta = 0.9$)	0.03	0.037	0.037	0.0000	0.9290	0.0000	0.9298
		0.06	0.074	0.074	0.0000	0.9195	0.0000	0.9217
		0.09	0.111	0.111	0.0000	0.9073	-0.0001	0.9127

表 14　検出力 (層の参加者数の比=1:2, 対照群の応答割合=(0.68, 0.86))

N	欠測	$p_{(0,0)}$	欠測割合		完全ケースの解析		MI-LMB	
	メカニズム		被験薬群	対照薬群	バイアス	検出力	バイアス	検出力
378	MCAR	0.03	0.060	0.060	0.0004	0.9233	0.0005	0.9301
		0.06	0.120	0.120	0.0004	0.9052	0.0004	0.9222
		0.09	0.180	0.180	0.0005	0.8829	0.0004	0.9106
	MAR ($\eta = 0.3$)	0.03	0.052	0.052	0.0004	0.9268	0.0005	0.9319
		0.06	0.105	0.105	0.0004	0.9121	0.0004	0.9219
		0.09	0.157	0.157	0.0004	0.8952	0.0003	0.9138
	MAR ($\eta = 0.6$)	0.03	0.045	0.045	0.0004	0.9295	0.0005	0.9318
		0.06	0.090	0.089	0.0004	0.9185	0.0004	0.9228
		0.09	0.134	0.134	0.0004	0.9051	0.0004	0.9150
	MAR ($\eta = 0.9$)	0.03	0.037	0.037	0.0005	0.9303	0.0005	0.9326
		0.06	0.074	0.074	0.0004	0.9245	0.0004	0.9250
		0.09	0.112	0.111	0.0004	0.9146	0.0004	0.9171

5. まとめと今後の課題

本検討では，欠測がある場合に，Mantel-Haenszel 法による割合の群間差と層別 Newcombe 法による 95%信頼区間を計算する方法として，MI-LMB の適用について紹介した．シミュレーションの結果より，完全ケー

スの解析の第 I 種の過誤確率は名目の水準を超える場合があったが，MI-LMB では第 I 種の過誤確率を名目の水準以下を保持しており，検出力は完全ケースの解析と同等以上であったことから，MI-LMB を感度分析として利用する意義があると考えられた．また，試験計画の時点で，完全ケースの解析の第 I 種の過誤確率が名目の水準を超えることが懸念される状況として，補助変数と関連する欠測の発生が予想され，その割合が大きくなる可能性があると想定される場合は，MI-LMB を主解析として設定することが考えられた．

一方で，層別 Newcombe 法による 95%信頼区間について，MI-LMB を適用したときの第 I 種の過誤確率は名目の有意水準より小さく保守的な傾向がある．より高い検出力を期待でき，第 I 種の過誤確率を名目の水準に近い値となるような，欠測がある場合の層別 Newcombe 法に基づく信頼区間の構成方法の検討は今後の課題である．また，欠測がある場合の Mantel-Haenszel 法による割合の群間差に対する信頼区間として，層別 Newcombe 法以外の方法との性能比較については今後の課題である．

参考文献

1. 医薬品医療機器総合機構. 新型コロナウイルス（SARS-CoV-2）ワクチンの評価に関する考え方 (補遺４) ― 親ワクチンを改変した変異株ワクチン及び新たな有効成分を用いた追加接種用ワクチンの免疫原性に基づく評価について. 2022. https://www.pmda.go.jp/files/000247491.pdf (accessed 2025-08-10).
2. Suzuki R, et. al. Booster vaccination using bivalent DS-5670a/b is safe and immunogenic against SARS-CoV-2 variants in children aged 5–11 years: a phase 2/3, randomized, active-controlled study. *Front. Immunol.* 2024;15:1445459.
3. 医薬品医療機器総合機構. 臨床試験のための統計的原則 補遺 臨床試験における estimand と感度分析. 2024. https://www.pmda.go.jp/files/000269154.pdf (accessed 2025-08-10).
4. van Buuren S. Flexible Imputation of Missing Data, 2nd edition. Boca Raton, FL: CRC Press, 2018.
5. Li X, Mehrotra DV, Barnard J. Analysis of incomplete longitudinal binary data using multiple imputation. *Statist. Med.* 2006;25:2107-2124.
6. Yan X, Su GX. Stratified Wilson and Newcombe Confidence Intervals for Multiple Binomial Proportions. *Stat. Biopharm. Res.* 2010;2:3,329-335.
7. 佐々木英磨, 舟尾暢男. Mantel-Haenszel 法により調整したリスク差の信頼区間に関する一考察. 2017, https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2017-b-09.pdf (accessed 2025-08-10).
8. Lu K, Guo H. Multiple imputation analysis of Miettinen-Nurminen interval for difference in proportions. *Pharm Stat.* 2022;21:196-208.

ジョブ理論を組み込んだAIプロフィール生成と DataMaker活用の新展開 ～ SAS Hackathon 2024 の実証実験に基づく～

○村田 粋堂　下山 晃平　八木橋 和也
田中 蒼斗　渡慶次 一哲　小野 潔
(コムチュア株式会社 デジタルイノベーション本部)

The Job Theory-Driven AI Customer Profiling and Structured Data Generation with SAS DataMaker
― Evidence from the 2024 SAS Hackathon ―

要旨

本研究は、1 カ月の開発期間に新人メンバーが中心となって実施した SAS Hackathon 2024 における実証実験を通じ、SAS DataMaker による構造化データの生成とジョブ理論（Jobs to Be Done Theory）を基盤とした AI 顧客プロファイル生成を統合したハイブリッド CRM モデル構築の可能性を示した。SAS DataMaker は、生成された定性的情報と整合する構造データを高い再現性で生成し、実データを直接扱わずに分析を可能にした。また、Google Gemini と SAS Viya Workbench を API 連携させ、顧客が達成したい目的（ジョブ）を反映したプロファイルと、それに適合するインセンティブ（プレゼント）を自動生成した。実証では米国クレジットカード利用ダミーデータ 1000 件を用い、従来施策と比較した結果、短期間かつ少人数で多様かつ一貫性のある CRM 施策データを構築可能であることが確認された。

キーワード

ジョブ理論　生成 AI　AI プロフィール生成　SAS DataMaker
構造化データ生成　インセンティブ設計　ハイブリッド CRM　プライバシー保護

1. はじめに

1.1. 背景と課題意識

近年、マーケティング領域では、顧客ごとに最適化された施策の重要性が高まっている。従来の年齢・収入・購買履歴といった静的属性に基づく一律施策は、顧客の潜在的動機や行動理由を十分に捉えられない。本研究ではこの限界を解決するため、ジョブ理論を核に生成 AI により顧客プロファイルとインセンティブ案を自動生成し、SAS DataMaker により合成データを生成して構造化データを補完するハイブリッド型 CRM 戦略を提案する。

本稿では、その理論的背景、モデル設計、SAS Hackathon 2024 における実証結果を報告する。

1.2. 研究の目的とアプローチ

本研究の目的は、従来の属性ベースの分析では把握しきれなかった顧客の潜在的動機を明らかにし、より個別性の高い CRM 施策を可能にすることである。そのために、ジョブ理論を基盤とし、生成 AI（大規模言語モデル：LLM）と SAS DataMaker を用いたハイブリッド型 CRM モデルを設計・実証し

た。

具体的には、顧客データから Google Gemini を用いてプロファイルとインセンティブを自動生成した。また、SAS DataMaker を利用することで、実データを直接用いることなく構造化データを生成し、プライバシー保護に配慮しながら再現性の高い検証を行った。さらに、金融分野では個人情報の取り扱いに厳しい制約があり、実データを直接利用できない場面が多い。本研究では、SAS DataMaker を活用することでこの制約を克服し、合成データを併用することにより実データを直接扱わずに分析フローの再現性を確認した。

2. 理論背景

2.1. ジョブ理論と CRM への応用

ジョブ理論（Jobs to Be Done Theory）は、米ハーバード・ビジネス・スクールのクレイトン・クリステンセン（Clayton M. Christensen, 1952-2020）教授によって提唱された。彼は顧客が製品やサービスを選ぶ本当の理由は「属性」ではなく、「達成したい進歩（Job）」であると主張した。この理論の本質は、顧客が商品やサービスを「何かの目的を達成するために“雇う（hire）”」と考える。たとえば「家族と楽しく過ごしたい」「将来の不安を軽減したい」といった目的が購買の背景にある。この視点を取り入れることで、単なる年齢・性別などの属性では捉えきれない、顧客の内面的・動機的な行動理由を明らかにできる。

近年は生成 AI の発展により、行動データや属性から潜在的なジョブを推定することが可能となり、再び注目を集めている。CRM においてジョブ理論を導入することで、顧客の深層的な動機に沿った施策設計が可能となる。

2.2. 生成 AI とプロファイル生成

生成 AI（Generative AI）とは、大規模なテキストデータを学習した言語モデル（LLM）を活用し、自然言語や構造化情報を生成する技術である。本研究では、Google Gemini を用い、条件付きのプロンプト（例：与信データ、属性情報）から、顧客の性格、価値観、生活スタイルなどを含むプロファイル文を自動生成した。

生成 AI の強みは、与えられた情報を基に多様なパターンのアウトプットを柔軟に作成できる点にある。たとえば、同じ年齢・年収帯の顧客であっても、「貯蓄にあまり余裕がないため支出を控える傾向がある」「既婚で子供がいるため家族でプライベートを楽しみたい傾向がある」といった異なる心理傾向を記述できる。

2.3. SAS DataMaker の役割と価値

金融業におけるデータの利活用には、量と質の面での課題がある。例えば、各業務が異なるシステムに分散されて管理されるサイロ化により、部門間を横断したデータの利活用が困難となるケースが多い。また、コールセンターの応対記録やメールなどの非構造化データは加工が難しく、更にトランザクションデータのような膨大なデータはリアルタイムな処理が追い付かずデータ品質の低下に繋がる。

この問題を解決するために、近年は Gemini のような生成 AI を用いたプロファイル文から構造化データを作成する試みもあるが、統計的特徴の再現はできても、項目間の関係性や論理的一貫性は弱い。

一方で SAS DataMaker は、項目間の依存関係を保持した再現性の高い構造化データを生成できる。機械学習手法である制約付き VAE（Variational Autoencoder）や Conditional GAN（Generative Adversarial Network）などの生成モデルは、データ構造を学習し、項目間の依存関係を含んだ合成データを出力す

ることができる。これらの手法は安定性や各パラメータの調整（学習率、潜在変数の次元、損失関数設計など）に高度な専門知識が求められる点が課題である。そのため、本研究では再現性と開発効率を重視し、構造制約の明示化と GUI ベースでの設定が可能な SAS DataMaker を採用した。

なお、LLM・SAS DataMaker・VAE/GAN の比較結果は 9 章 Appendix の表 9.1.1 に示す。

3. モデルの設計

本章では、ジョブ理論・生成 AI を用いた CRM モデルの設計と、SAS DataMaker による合成データの作成について述べる。本モデルは、①顧客プロファイル生成(生成 AI による動機・心理の言語化)、②インセンティブ設計（ジョブ理論に基づく提案)、③AI と人の協働による改善、の 3 要素から構成される。

3.1. 顧客プロファイル生成のアプローチ

顧客プロファイルは、年齢・職業・年収などの定量的な属性だけではなく、「価値観」「悩み」「生活スタイル」といった定性的な情報を含むことが望ましい。そこで本研究では、まず図 3.1.1 に示すような債務負担比率・延滞歴・貯蓄率などの列を含む与信データをもとに、Google Gemini を用いて自然言語によるプロファイル文を生成した。一部の列の説明については、図 3.1.2 に示す通りである。

Customer_ID	Debt_Burden_Ratio	DPD_30_6M	DPD_60_6M	Savings	Job_Experience_Level	Years_Employed	•••	Default_Flag
0	< = 0.10	0	0	2.55335	15	15		0
1	< = 0.10	0	1	5.5156	1	19		1
2	< = 0.10	0	0	0.875	2	8		0
⋮								

図 3.1.1 使用した与信データの一部

カラム	説明	変数型	値の例
Customer_ID	顧客番号	数値	1, 2, 3…
Debt_Burden_Ratio	債務負担比率 (月間の債務返済額/月間の総収入)	カテゴリ	< = 0.10, < = 0.50, > 0.50
DPD_30_6M	過去6カ月に30日の延滞をしたかどうか	binary	0, 1
DPD_60_6M	過去6カ月に60日の延滞をしたかどうか	binary	0, 1
Savings	貯蓄率	数値	-0.71, 2.28…
Job_Experience_Level	役職の高さ	数値	0, 1, 2, …
Years_Employed	雇用年数	数値	0, 1, 2, …
⋮	⋮	⋮	⋮
Default_Flag	デフォルトしたかどうか	binary	0, 1

図 3.1.2 使用した与信データの一部の列の説明と値の例

たとえば、プロフィール生成のために次のようなプロンプトを用いる：

① プロンプトの入力例：
添付データから、想像で顧客プロファイルを生成してください。またデフォルト顧客はデフォルトを反映したプロファイルにしてください。

② プロフィールの出力例：

顧客 A は 39 歳の女性で、7 年間民間企業に勤務しています。学歴は 2 段階の教育を受けており、扶養家族が 1 人いて既婚です。デフォルト確率は 0.10 以下と低く、信用度は良好です。10 年間現住所に住んでおり、所有物件で居住しています。資産は 0 ドル、貯蓄率は 2.448%と、貯蓄にあまり余裕がない様子が見えます。

この LLM 出力により、顧客データから読み取れる情報を文で表現できる。

3.2. インセンティブ設計とジョブの対応

生成された顧客プロファイルに対して、LLM を用いてプレゼント候補リストを作成し、各人物に適したインセンティブとその理由を同時に生成する。金額帯はリスク水準に応じて段階的に設定し、従来の優良顧客向け施策に加え、デフォルト予備群にも財務健全化を促す無料プログラム（例：資金管理アプリ）を候補に含めた。

① プロンプトの入力例：

「プロファイルを元に正常顧客へのプレゼントを生成してください。
デフォルトおよびデフォルト予備軍の顧客向けに提供できるプレゼントやサービスのアイデアを以下に 10 個挙げました。・・・・・・・・・顧客のプロファイルからプレゼントとそれを選択した理由を生成し、プロファイルデータとプレゼントとその理由を出力してください。」

② 出力インセンティブ・提案理由の出力例：

「選定：映画チケット 2 枚」
「理由：既婚で扶養家族が 1 人いることから、家族で楽しめるプレゼントとして映画チケットを提案します。」

このように、プロファイルに基づく提案を自動生成することで、マーケティング担当者の作業を補完し、優良顧客だけでなくデフォルト予備群も含めたパーソナライズされた CRM 施策の実現につなげることが可能となる。この出力により、属性データだけでは読み取れない顧客の“心理的ジョブ”を浮き彫りにできる。

このようなプロファイル・インセンティブを 1000 件、理由については 30 件を自動生成したうえで、内容の整合性や多様性を確認し、インセンティブのカテゴリごとの出現傾向も分析した。

3.3. AI と人間の協働、および構造化データの統合

本研究では、まず LLM により 200 件のインセンティブ候補を生成し、担当者 6 人が修正・加筆した後に再学習させた。これにより、人の知識を反映した多様性の高い提案が可能となった。また、Google Gemini や SAS DataMaker を用いて顧客データから新たな構造データを生成し、生成されたデータの精度を検証した。

3.4. モデル全体の流れ

図 3.4.1 に示すフローにより、生成 AI の柔軟性を用いた CRM 基盤モデルの構築・評価を行った。まず初めにサンプルとしてクレジットカード会社の与信データを 2000 件作成し、学習用データと評価用

データに 1000 件ずつ分割した。次に、これらのデータから Google Gemini を用いて顧客それぞれの特徴を反映したプロファイルを生成した。更に、生成したプロファイルを基に適切と考えられるインセンティブ(特典)を生成した。プロファイル・インセンティブを生成する前のデータからデフォルト顧客を予測するベースモデルを構築し、評価用データを用いた評価を行った。また、インセンティブ列を説明変数に含めてデフォルト顧客を予測するインセンティブ使用モデルについても同様に評価用データを用いた評価を行った。

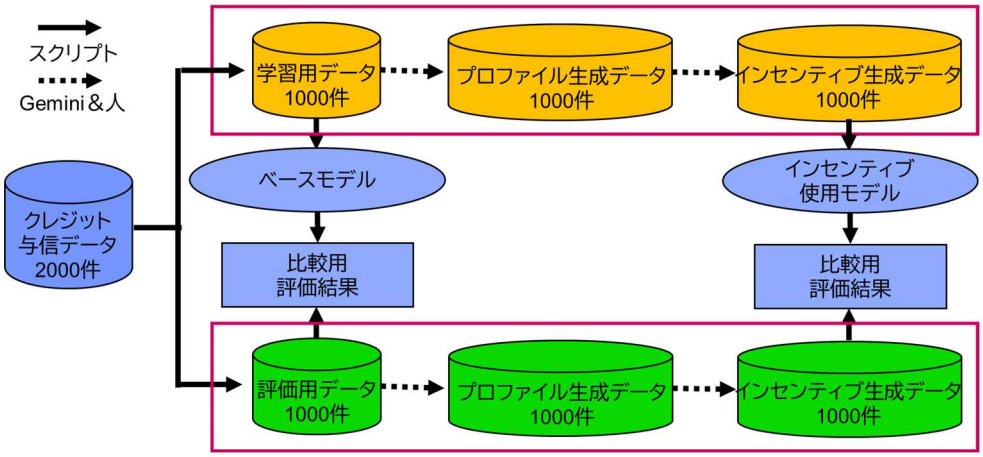


図 3.4.1 インセンティブの生成とモデル評価のフロー

また、図 3.4.2 に示すフローのように、1000 件の学習用データから SAS DataMaker・Google Gemini を用いて 1000 件の合成データをそれぞれ作成した。また、これらのデータを合わせたユニオンデータ 3000 件を作成し、それぞれ図 3.4.1 と同様にインセンティブ使用モデルの構築・評価を行った。

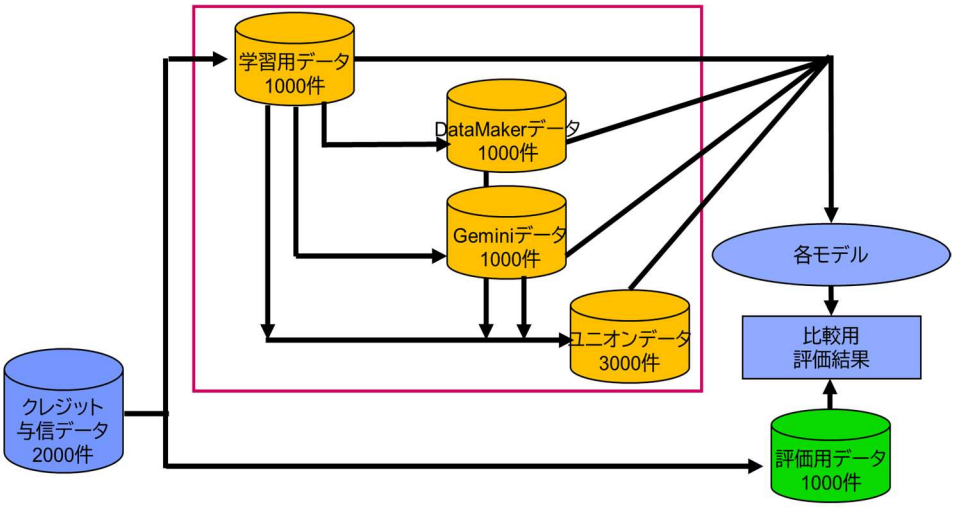


図 3.4.2 合成データの生成とモデル評価のフロー

4. 実証実験：SAS Hackathon 2024

4.1. 実験の目的とデータ概要

本実証実験の目的は、ジョブ理論に基づくプロファイル生成とインセンティブ設計を自動化し、それ

らを構造化データと統合した CRM 施策が従来手法に比べてどの程度有効であるかを定量的に評価することである。

使用した元データは、米国におけるクレジットカード利用者のダミーデータ（2000 件）であり、このうち 1000 件を学習用データとして、残りの 1000 件を評価用データとして用いた。データには、年齢、職業、収入、支出、借入状況、デフォルト有無などが含まれている。なお、元データの 30%は支出超過またはデフォルト傾向があると見なされる顧客であった。

4.2. 従来施策との比較

本モデルと従来施策を比較した結果を表 4.2.1 に示す。従来は手動で約 600 件の施策作成に 3〜5 人日を要し、担当者による表現のばらつきも大きかった。一方、本モデルではプロンプト調整を含めても 2〜3 時間、処理時間のみなら十数分で 1000 件以上のプロファイルとインセンティブを生成できた。文体は全体として一貫しており、各プロファイルにはジョブ的背景が明示された。さらに、プレゼント候補の多様性も LLM により拡張され、構造データとの整合性も自動的に担保された。

比較項目	従来施策（手動）	本モデル（LLM＋DataMaker）
作成速度	約 600 件で 3〜5 人日	約 2〜3 時間（プロンプト調整含む） 処理時間のみなら十数分
表現の一貫性	担当者によりバラつきあり	全体として文体が安定
顧客ジョブへの対応	明示的には未定義	各プロファイルにジョブ的背景が明記
プレゼントの多様性	実務経験に依存	LLM により多様なカテゴリが登場 （LLM でプレゼント生成に制約を入れることも可能）
構造データとの整合性	手動設計に依存	自動生成されたルールベースで整合

表 4.2.1　従来施策（手動）と本モデル（自動生成）の比較

4.3. 分析結果

図 4.3.1 に、3 章の図 3.4.1 で示したベースモデルとインセンティブ使用モデルの評価結果を示す。本研究では、生成 AI や SAS DataMaker によるデータ生成の有効性を定量的に検証するため、与信モデルを評価対象とした。与信は金融分野における典型的かつ重要なユースケースであり、ベースモデルとの精度比較を通じて、生成データが実際に予測性能の向上に寄与するかを確認した。ここで、モデルの目的変数はデフォルトフラグ(Default_Flag)とし、アルゴリズムは高速で精度が良い決定木系モデルである LightGBM を使用した。学習データは 750 件、検証データは 250 件とし、カテゴリ変数はターゲットエンコーディングによる処理を行った。インセンティブ使用モデルにおいては、学習用データの上位 3 つのインセンティブを判別分析モデルに組み込み、他は「その他」として説明変数に使用した。インセンティブを含めたカテゴリ変数はターゲットエンコーディングして使用した。図 4.3.1 の結果から、ベースモデルよりもインセンティブを説明変数として使用したモデルの方が全体の精度が向上していることが分かった。



図 4.3.1 ベースモデルとインセンティブ使用モデルの評価結果

また、図 4.3.2 にインセンティブ使用モデルにおける Shapley を示す。インセンティブの項目

(Present)が上位の貢献度にあることから、統計的にもおかしくないインセンティブが生成できていることが分かった。

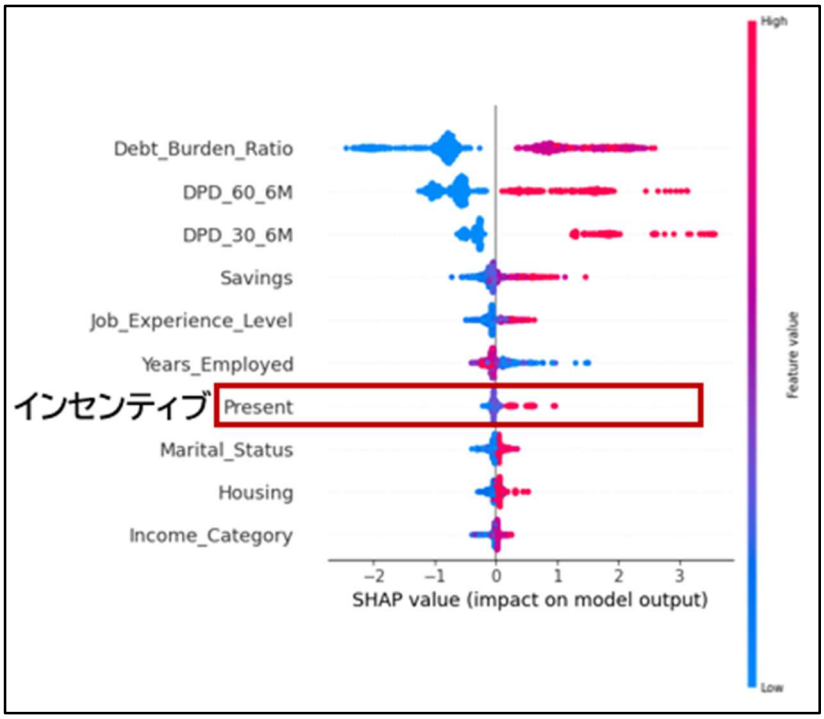


図 4.3.2 インセンティブ使用モデルにおける Shapley

図 4.3.3 に、3 章の図 3.4.2 で示したベースモデル、Gemini モデル、DataMaker モデル、ユニオンモデルの評価結果を示す。目的変数、アルゴリズムは先ほどと同様である。ベースモデル、Gemini モデル、DataMaker モデルにおいては、学習データは 750 件、検証データは 250 件とし、ユニオンモデルにおいては、学習データは 2250 件、検証データは 750 件とした。カテゴリ変数はターゲットエンコーディングによる処理を行った。図 4.3.3 の結果から、Gemini モデルはベースモデルよりも精度が悪化した一方で、DataMaker モデルとユニオンモデルでは精度の向上が確認できた。

	正解率	適合率	再現率	F1
ベースモデル CSV	0.9480	0.8590	0.8171	0.8375
Gemini モデル Gemini CSV	0.8480	0.0705	0.6111	0.1264
DataMaker モデル SAS DataMaker CSV	0.9580	0.8654	0.8654	0.8654
ユニオンモデル SAS DataMaker Gemini CSV CSV CSV	0.9540	0.8013	0.8929	0.8446

図 4.3.3 ベースモデルとデータ生成モデルの評価結果

4.4. 分析結果のまとめ

本実証実験を通じて、生成 AI と SAS DataMaker を用いた CRM モデルは、以下の点で有効であることが示された。

- ジョブ理論に基づくプロファイルとプレゼント提案の自動化が可能
- 生成された情報に整合する構造データを同時生成でき、分析と施策の往復が容易
- 再現性・効率性に優れ、少人数のチームでも短期間で大量の CRM データが構築可能

5. 実装とシステム構成

5.1. 概要（プラットフォームと役割）

本章では、本研究で用いたプラットフォームである SAS Viya Workbench と SAS DataMaker の役割を述べる。まず SAS DataMaker を用いて元データの統計的特性を保持した合成データを生成・検証した。具体的には、合成データを用いて与信モデルを構築し、F1 スコアや与信精度を指標として評価した。これにより、実データを直接利用することなく、プライバシーを保護しながら再現性の高いモデル学習と検証を行うことが可能となった。

次に、SAS Viya Workbench から Gemini API を利用して、顧客データに基づくプロファイル・インセンティブ・選定理由を生成した。DataMaker で作成した合成データと組み合わせることで、生成 AI と構造化データを統合した検証フローを構築し、短期間で効率的に CRM 施策を検討できることを確認した。詳細なプログラム例や画面キャプチャは 9 章の Appendix に示す。

5.2. 実装フローとシステム構成

本研究の実装では、まず SAS DataMaker により合成データを作成し、その統計的特性と項目間の関係性を確認した。次に、SAS Viya Workbench から Gemini API を連携し、合成データを含む顧客データに基づきプロファイル・インセンティブ・理由を生成した。この順序により、生成 AI と構造化データ生成を統合した検証フローを短期間で構築することが可能となった。詳細な手順は 9 章 Appendix 9.3 に示す。

① Gemini によるプロファイルとインセンティブ生成
- SAS Viya Workbench から Gemini API を呼び出し、顧客データに基づきプロファイル・インセンティブ・理由を生成。
- 生成結果を DataMaker で作成した合成データと統合し、CRM 施策設計およびモデル検証に活用。

② SAS DataMaker による合成データ生成と検証
- GUI で条件や行数を指定し、合成データを生成。
- 生成データを用いて与信モデルを構築し、F1 スコアや与信精度により妥当性を確認した。

5.3. SAS DataMaker による合成データの生成

SAS DataMaker は、合成データを生成するためのツールであり、ノーコードのインターフェースで操作できる。既存データを基に、現実世界の特徴を反映しつつ機密情報を保護したデータを作成できる点が特徴である。

本研究では、SAS DataMaker を用いて元データの統計的特性を保持した合成データを作成し、その

妥当性を与信モデルの F1 スコアおよび与信精度によって検証した。このアプローチにより、実データを直接利用できない環境でも再現性を確保し、分析とモデル評価を可能とした。詳細な手順は 9 章の Appendix 9.2 に示す。

6. 考察と結論

本研究では、ジョブ理論を軸に生成 AI と SAS DataMaker を用いた CRM モデルを構築し、SAS Hackathon 2024 にて実証実験を行った。その結果、以下の知見と課題が得られた。

① 本モデルの有効性と実務的意義

LLM によるプロファイル自動生成は、従来の属性分析では捉えにくい「動機」「価値観」を言語化できた。また、そのプロファイルに基づくインセンティブ提案は、顧客の内面的ニーズに即し、個別性の高い CRM 施策を可能にした。さらに SAS DataMaker により数値データを短時間で生成し、信用スコアリングや与信管理などへの応用可能性も確認された。

② ジョブ理論と生成 AI の限界

ジョブ理論は収支構造や家族構成の違いを反映できる有効な枠組みである一方で、その定義が抽象的であり、解釈の仕方に依存しやすいという限界を持つ。また、生成 AI の出力は人間よりばらつきが少ないとはいえ、完全に一貫しているわけではなく、ジョブ分類の定量化や自動化にはさらなる技術的補完が必要である。さらに、LLM は柔軟にプロファイルを生成できる一方で、項目間の整合性や統計的再現性には課題があり、SAS DataMaker による補完は有効であったものの、プロファイルと構造データを双方向に連携させるには追加の工夫が求められる。制約付き VAE や Conditional GAN といった生成手法も存在するが、専門知識を要するため一般的な業務での利用には依然としてハードルが高い。

③ 実装とスケーラビリティの課題

本モデルは 1000 件規模で検証したが、実務で数千万件以上を扱うにはパフォーマンス・コスト・運用体制の整備が必要である。また、生成 AI の出力には倫理的・差別的バイアスを含むリスクもあり、監査体制や説明可能性の確保が今後の導入における重要課題となる。

④ 今後の展望

今後の発展方向としては、まず①ジョブ分類をテンプレート化し、自動的に分析へ活用できる仕組みを整備することが挙げられる。さらに②生成 AI（LLM）と構造データを双方向に学習させるモデルを構築し、③SAS Intelligent Decisioning と連携させることでリアルタイムに施策を適用する体制を整えることが求められる。その上で④金融・教育・小売に加え、医薬分野など多様な領域への応用分野がある（詳細は Appendix.4 参照）。⑤生成 AI が持つ倫理的リスクやバイアスを制御する仕組みを強化することが必要である。加えて⑥従来は専門家依存で最優良顧客に限られていたジョブ理論施策を、生成 AI 活用により幅広い顧客層へ展開可能にすることが期待される。

本研究は、生成 AI と SAS DataMaker を組み合わせた新しい CRM 基盤の可能性を示したものである。今後の課題克服を通じて、より実務的で汎用性の高い顧客中心型 CRM へと発展することが期待される。

7. 謝辞

本研究のプログラム実装は、SAS ハッカソン（期間：１カ月）の短期間に、SAS プログラミング経験半年の村田および 2〜3 年の経験を持つ新人メンバーが中心となって実施された。限られた期間での成果創出にあたり、チームメンバーや関係者の協力に深く感謝する。

8. 参考文献

- Clayton M. Christensen, Taddy Hall, Karen Dillon, David S. Duncan. Competing Against Luck: The Story of Innovation and Customer Choice. Harper Business, 2016.
- Clayton M. Christensen. The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail. Harvard Business Review Press, 1997.
- クレイトン・クリステンセン他. 『ジョブ理論 イノベーションを予測可能にする消費のメカニズム』 翔泳社, 2017.
- Google DeepMind. Gemini 1.5 Technical Report, 2024.
- SAS Institute Inc. SAS DataMaker: “SAS Acquires Hazy Synthetic Data Software to Boost Generative AI Portfolio.” SAS Newsroom, https://www.sas.com/en_us/news/press-releases/2024/november/hazy-syntheticdata.html, 2024.
- SAS Institute Inc. SAS Viya Documentation. https://documentation.sas.com/
- 日本ディープラーニング協会. 『日本における生成 AI 活用ガイドライン』, 2023.
- 生成 AI とジョブ理論で作る顧客中心型 CRM, 人工知能学会全国大会, 小野, 2025.
- 医療ビックデータを用いた AI による合成医療データの生成および品質評価手法.関, 2025.
- Synthetic data generation methods in healthcare: A review, Pezoulas et al., 2024
- Synthetic data generation: a privacy-preserving approach to rare disease research. PMC 11958975, 2025.
- A review on generative AI models for synthetic medical text, time-series, and tabular data. npj Digital Medicine,nature, 2024.

9. Appendix

9.1. データ生成手法の比較

比較項目	LLM（生成 AI）	SAS DataMaker	VAE/GAN (機械学習型生成モデル)
生成対象	テキスト・疑似的な表形式	構造化された表形式データ（数値・カテゴリ含む）	構造化データ（潜在空間を通じた学習出力）
項目間の整合性	弱い（独立項目になりがち）	強い（事前定義条件や生成ロジックにより関係性を保持）	中程度～強い（学習により関係を再現可能）
統計特性（平均・分散等）の再現	単項目レベルで可能	高精度で再現可能	学習済みモデルにより再現可能だが、調整が難しい
柔軟性・表現力	高い（文脈に応じた自然文生成）	定義条件の範囲で柔軟（業務要件に沿った生成が可能）	非常に高い（複雑な分布や関係性も学習可能）
パラメータ設定の難易度	高（プロンプト設計が中心）	低（GUI）	高（学習率、損失関数、潜在変数数など調整）
実装のしやすさ	高い（API 等で簡易に利用可能）	高い（GUI 操作＋SAS 環境と統合）	低～中（専門知識とハイパーパラメータ調整が必要）
再現性・信頼性	低～中（生成内容に揺らぎあり）	高い（同条件で同一結果が得られる）	中（学習済みモデルに依存し、再学習が必要な場合あり）
適用例	顧客プロファイル文、施策案生成など	模擬顧客データ、仮想 CRM データ生成	クレジットスコアデータ生成、異常検知用合成データ

表 9.1.1　LLM/SAS DataMaker/VAE・GAN の構造化データ生成における比較

（補足）

本研究で利用した SAS DataMaker は 2024 年 9 月時点の GA（General Availability：一般提供）版ではなく、プライベートプレビュー版であり、合成データ生成には SMOTE（Synthetic Minority Over-sampling Technique）が用いられていた。その後、SAS 社は 2024 年 11 月に英国の Hazy 社を買収し、同社の高度な生成アルゴリズムを統合した GA 版を公開した。したがって、本研究の実証結果は開発段階の機能に基づくものであり、現行 GA 版とは性能面で差異がある可能性がある。

9.2. SAS DataMaker 操作画面

SAS DataMaker は下図のような画面で構成されており、元データをアップロードした後に「データが始まる行の番号」と「生成する行数」を指定することで合成データを生成できる。生成後は、元データと合成データそれぞれの相関行列が表示され、カラム間の関係性を比較することで再現度を確認することができる。

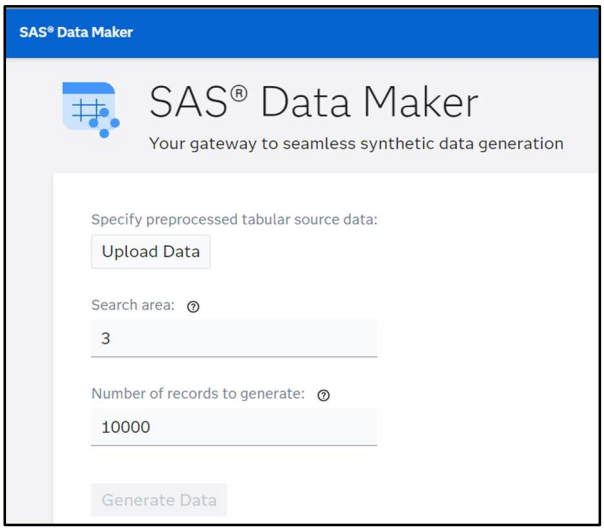


図 9.2.1 SAS DataMaker の画面

9.3. SAS Viya Workbench の使い方と Gemini API を呼び出す方法

SAS Viya Workbench は、SAS Viya プラットフォームの一部である、クラウドベースのオンデマンド環境である。AI や機械学習モデルの開発、データ分析、モデリングを効率的に行うことができ、Jupyter Notebook や Visual Studio Code と統合されており Python や SAS のコードを使用することができる。

SAS Viya Workbench を利用する際には、下図のようにワークベンチを作成する。ワークベンチごとに環境を切り替えることが可能であり、メモリを設定することで計算能力に応じた環境を構築できる。

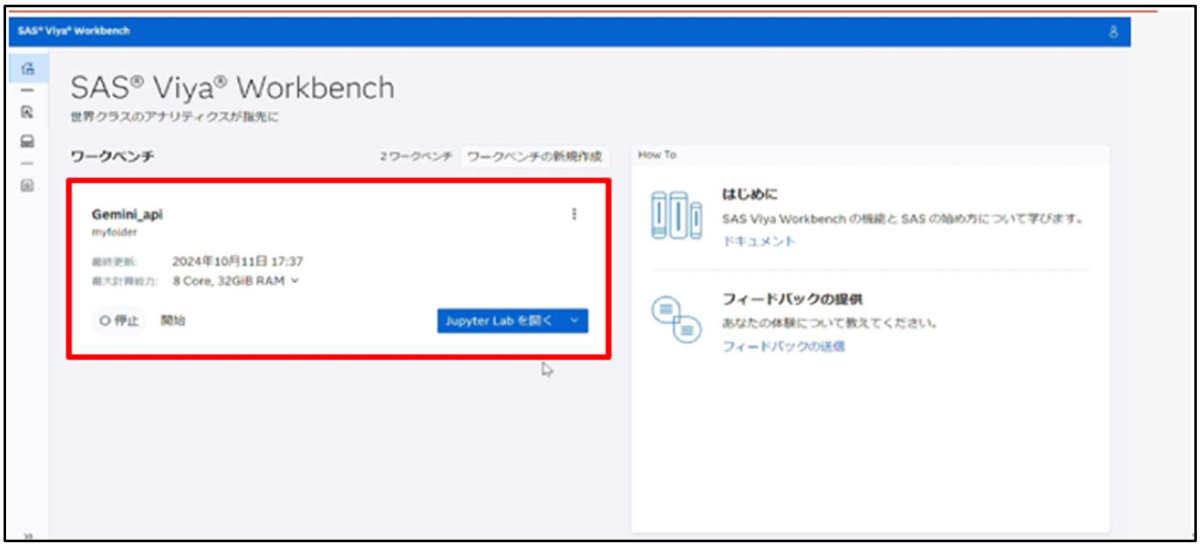


図 9.3.1 SAS Viya Workbench の起動画面

「Jupyter Lab を開く」を選択すると、下図のような画面が表示され、Jupyter Lab と同様に開発・分析をすることが可能である。

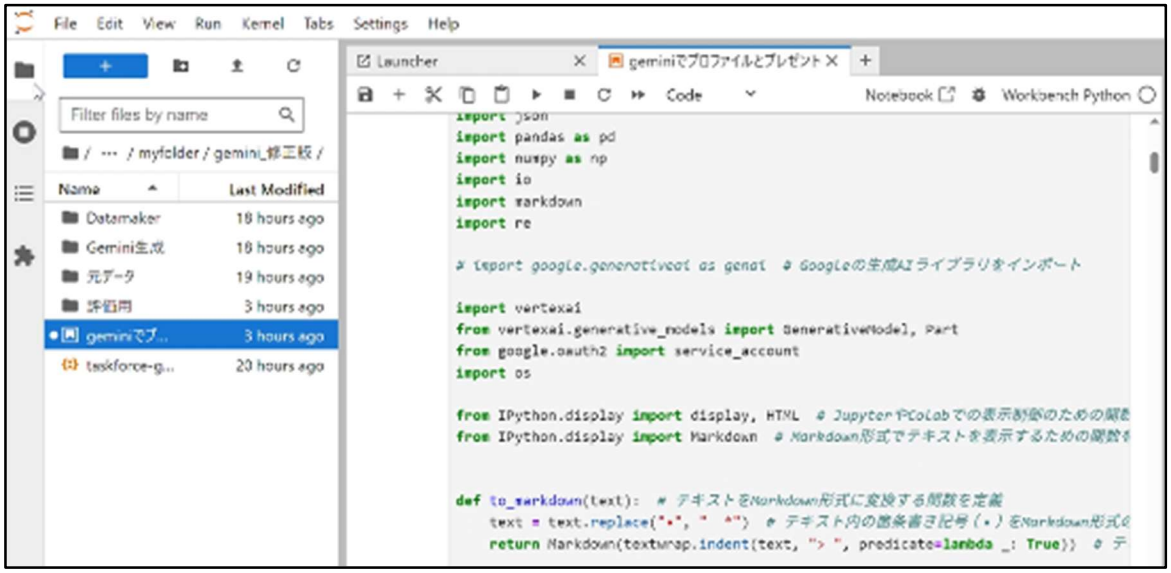


図 9.3.2 Jupyter Lab の起動画面

このワークベンチで Gemini を用いるためには、Google AI Studio から Gemini API キーを作成する。下図のように「API Keys」から「API キーを作成」を選択することで、コード内で設定するキーを取得することができる。

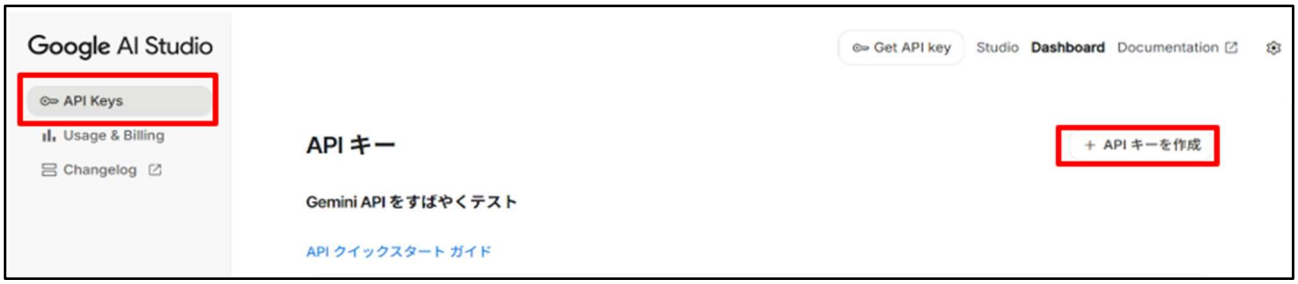


図 9.3.3 Google AI Studio から API キーを作成する

作成した API キーは、下図のように Gemini API を Python から利用するための Google 公式ライブラリである google.generativeai をインポートし、configure 関数 の api_key で指定する。(下図では google.generativeai を genai 、API キーを"xxxxxxxxxxxxxxxx"としている。)

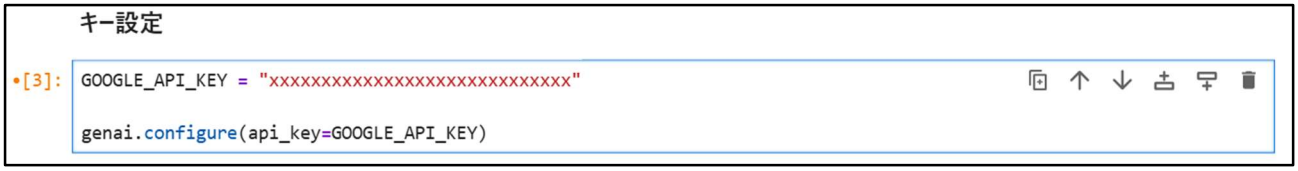


図 9.3.4 Google AI Studio で取得した API キーを設定する

これにより SAS Viya Workbench から Google Gemini を呼び出し、顧客データそれぞれに対してプロファイル・インセンティブ・選定理由の生成、また学習用データを基に合成データの生成を行った。プロファイル・インセンティブ・選定理由の生成については、プロンプトエンジニアリングと Chain-of-Thought の発想を基に具体的な例の提示と段階的な思考の指示を与えることにより妥当性の高い結果の生成を目指した。SAS Viya Workbench から Google Gemini を呼び出すプログラムのサンプルは、付録の「01_ConnectToGemini.ipynb」を、プロファイル・インセンティブ・選定理由を生成するプログラムのサンプルは付録の「02_GenerateIncentive.ipynb」を参照。

また、学習用データとインセンティブ生成データを用いてデフォルトフラグの予測を行うモデルを構築し、評価用データを用いて評価を行った。評価を行うプログラムのサンプルについては、付録の「03_BuildAndValidateModel.ipynb」参照。

9.4. 医薬分野への展開

医薬分野は個人差が大きく、プライバシー制約も強い領域であるため、生成 AI と合成データの枠組みが特に有効である。本研究で構築した「ジョブ理論 × 生成 AI × 合成データ」の枠組みは、医薬分野においても応用可能である。以下に代表的な活用シナリオを示す。

- 患者プロファイル生成：生活習慣・心理的要因を含むプロファイルを自動生成し、医療従事者の判断を補助する。
- 服薬アドヒアランス支援：リマインド施策やインセンティブ提案により、治療の継続性を向上。
- 治療プランとの統合：患者の「達成したい進歩（Job）」に基づき、治療計画を最適化。
- 臨床試験・創薬サポート：合成データにより希少疾患や副作用リスクの検証を補完。
- 予防医療・公衆衛生：生活習慣病対策やワクチン普及施策の個別化に応用可能。

ベイズ統計における有効サンプルサイズ(ESS)のSASによる実装

○玉乃井崚太[1]

（シミック（株）データサイエンス事業本部）

Implementation of Prior Effective Sample Size (ESS) in Bayesian Statistics Using SAS

Ryota Tamanoi

Data Science Div., CMIC Co.,Ltd.

要旨

近年, 医療統計分野においては, ベイズ統計の応用が活発に進展している. ベイズ統計では観察データに加え外部情報を「事前分布」として組み込むことができる他, 複雑な試験デザインにも柔軟に対応できるなど, 多くの利点がある. 一方, 事前分布の情報量が過度に大きい場合, 解析結果が事前分布に大きく依存し, 観察データの影響が十分に反映されないという課題も指摘されている. したがって適切な情報量の事前分布を選択することが重要である. このとき有効サンプルサイズ (Effective Sample Size: ESS) を事前分布の情報量の指標として用いることができる. ESS は事前分布の情報量を観察データの人数に換算して定量化するものであり, 2010 年の医療機器におけるベイズ統計の FDA ガイダンスでは「プロトコルに含めることを推奨する情報」の一つとして挙げられている. 近年複数の ESS が提案されており, それぞれ定義や適用範囲など特徴が異なる. 本稿ではまずベイズ統計の基礎の説明を行い, その後 3 つの ESS の解説, SAS による実装を行う.

キーワード: ベイズ統計, 事前分布, 有効サンプルサイズ, ESS, 無情報事前分布

1. ベイズ統計と ESS の導入

本節ではベイズ統計になじみのない読者を対象に基礎的な内容から紹介し, 事前分布の ESS を導入する.

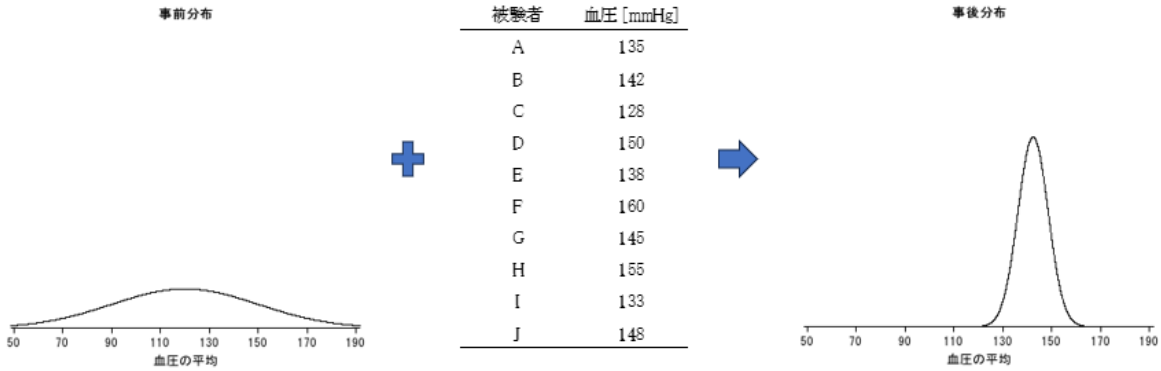
1.1. ベイズ統計と事前分布

ベイズ統計では一般的な統計と異なり, 事前分布とよばれる確率分布を観察データとは別に用意する必要がある. 事前分布と観察データの結果から事後分布を導出し, この事後分布から信頼区間や p 値に対応する値を算出することができる.

図 1 は事前分布と観察データから事後分布を求める流れを示している. この例では観察データを得る以前には, 血圧の真の平均は 120 前後, 80 から 160 程度であろうと考えていることがわかる. ここに観察データを加えた結果, 真の平均は 140 前後, 130 から 150 程度の範囲にあると予測できる.

事前情報がない場合には無情報事前分布 (non-informative prior. 情報の少ない事前分布;uninformative prior, 曖昧な事前分布;vague prior とよぶこともある) を用いることで事前分布の影響を制限できる. どのような事

図 1 事前分布 (左図) と観察データ (中央表) から事後分布 (右図) が算出される. 事前分布は平均 120, 標準偏差 30 であり, 事後分布は平均 142.4, 標準偏差 6.2 である.



前分布が無情報といえるかはモデルによって異なり課題も多い [1]が一般的には Jefferys prior [2]を含む reference prior [3]がよく用いられる.

ベイズ統計のメリットとして, 中間解析の結果やリアルワールドデータなどの事前情報がある場合, これを事前分布にすることで意思決定に用いるデータを増やすことができる他, 少ない症例数で検出力を確保できる場合がある [4]. また事前情報がない場合にも, 事後分布を用いた治療効果の定量的評価や中間解析による柔軟な意思決定が可能になる点で有用である [5]. 一方ベイズ統計のデメリットとして, 理論的な専門性や難しさの他に, 事前分布の設定の問題が指摘されている [4, 5, 6]. 情報のある事前分布を用いる場合, その選択によって解析結果が大きく変わりうる. 例えば事前分布の情報量が観察データに比べて過剰に大きい場合, 解析結果が事前分布に大きく依存し, 観察データの影響が十分に反映されない. そのため適切な情報量を持つ事前分布を選択することが重要となる. 平川ら (2021) [5]では情報のある事前分布を用いる場合, 感度分析として無情報事前分布を用いた結果との比較が提案されている.

1.2. ESS について

事前分布の情報量を「観察データ何人分であるか」で表した指標が事前分布の有効サンプルサイズ (prior effective sample size; ESS) である. すなわち ESS=1 の事前分布は一人分の情報を持ち, ESS=100 の事前分布は 100 人分の情報を持っていることを表す. FDA2010 [4]では Suggested Information to Include in Your Protocol の一つとして Effective Sample Size が挙げられている.

どの程度の ESS が適切であるかは試験によって異なる. 事前情報と試験の類似性が高い場合は ESS の大きい事前分布によって試験効率を向上することができるが, そうでない場合は ESS の小さい事前分布を用いることで事前分布と観察データの衝突 (prior-data-conflict) を防ぐことができる. FDA (2010) [4]では事前分布の ESS が試験の症例数を上回る場合事前分布は情報過多である可能性があり, 修正する必要があると述べられている.

注意として ESS という単語の意味はいくつかの文脈で用いられる. 本稿では事前分布の有効サンプルサイズを指しているが, FDA (2010) [4]では観察データを加えた後の事後分布の有効サンプルサイズを ESS とよんでいる. また PROC MCMC でも ESS が output に表示されるがこれは事前分布の ESS とは関係がない. 混乱が生じる場合は prior ESS などと示すべきである. 本稿では特に指定がない場合 ESS は事前分布の ESS を指す.

2. 様々な ESS の定義と性質

ESS はベイズ統計で古くから用いられており, Novick and Hall (1965) [7]では二項モデルに対するベータ分布の ESS (論文内では prior information とよばれている) について言及されている. 一方で一般の事前分布について ESS を求めることは容易ではなく, 近年も複数の ESS が提案されている. 本稿では提案されている ESS の定義をいくつか挙げ, 特にESS_{MTM}, ESS_{ELIR}, $ESS_{E\text{-}F}$の説明と SAS での実装を行う.

2.1. 共役事前分布における ESS

ベイズ統計において古典的に用いられている事前分布に共役事前分布とよばれる事前分布がある. これはモデルに指数型分布族を指定している場合に与えられる事前分布(族)であり, 事後分布が容易に計算できる特徴がある. 表 1 は代表的な共役事前分布の一覧である.

表 1: 代表的な共役事前分布 代表的なモデルと共役事前分布の組とその ESS の一覧 ここで N, invGa, Ber, Beta, Exp, Ga, Pois はそれぞれ正規分布, 逆ガンマ分布, ベルヌーイ分布, ベータ分布, ガンマ分布, ポアソン分布を指す.

尤度分布	パラメータ	共役事前分布とハイパーパラメータ	古典的に与えられている ESS の例
正規分布（分散既知）	$N(Y \mid \mu, \sigma^2)$（σ^2既知）	$N(\mu \mid m_0, s_0^2)$	σ^2/s_0^2
正規分布（平均既知）	$N(Y \mid \mu, \sigma^2)$（μ既知）	$invGa(\sigma^2 \mid \alpha, \beta)$	2α
正規分布	$N(Y \mid \mu, \sigma^2)$	$N(\mu \mid m_0, \sigma^2/\kappa)$ $\times invGa(\sigma^2 \mid \alpha, \beta)$	$\kappa = 2\alpha$のときκ. そうでないとき ESS は古典的に与えられていない
ベルヌーイ分布（二項分布）	$Ber(Y \mid p)$	$Beta(p \mid \alpha, \beta)$	$\alpha + \beta$
指数分布	$Exp(Y \mid \lambda)$	$Ga(\lambda \mid \alpha, \beta)$	α
ポアソン分布	$Pois(Y \mid \lambda)$	$Ga(\lambda \mid \alpha, \beta)$	β
線形回帰	$N(Y \mid \sum_{k=1}^{p} \beta_k X_k, \sigma^2)$	g prior	n/g ただしnは症例数

多くの共役事前分布に対しては ESS が一般的に用いられており, その値が表 1 の右側である. 例えば尤度分布がベルヌーイ分布, 事前分布にベータ分布 $Beta(\alpha, \beta)$ を用いるとき, 観察データとして$Y = 1$をn_1例, $Y = 0$をn_0例得たとすると, 事後分布は再びベータ分布$Beta(\alpha + n_1, \beta + n_0)$となる. (事後分布が, 事前分布のハイパーパラメータを変えるだけで得られることが共役事前分布の最大の特徴である.) ここからα, βはそれぞれ$Y = 1,0$の事前情報といえる. したがって ESS を合計$\alpha + \beta$と考えることができる.

注意として, ESS は文献によって 1,2 程度異なる場合がある. 例えばベータ分布の ESS に関して無情報事前分布, すなわち$ESS = 0$の基準として Haldane prior ($Beta(0,0) \propto p^{-1}(1-p)^{-1}$) を用いれば$ESS = \alpha + \beta$となるが, Jefferys' prior $Beta(0.5,0.5)$ や一様分布 $Beta(1,1)$ を用いれば ESS はそれぞれ $\alpha + \beta - 1, \alpha + \beta - 2$となる.

2.2. 分散比を用いた ESS

前節の ESS は共役事前分布に対してしか与えられていない. 一般の事前分布に対する ESS の定義は明らかではなく, 状況に応じて選択することが必要になる. 最初に, Malec (2001) [8]や FDA (2010) [4]で与えられている, 事後分布と「無情報事後分布」の分散の比を利用して ESS を算出する方法を紹介する.

ESS を求めたい事前分布を$p(\theta)$, 無情報事前分布を$\pi_0(\theta)$, m例の観察データを$Y_1, \cdots, Y_m$とおく. このとき, 無情報事後分布$\pi_m(\theta) = \pi_0(\theta \mid Y_1, \cdots, Y_m)$と$p(\theta)$の事後分布$p(\theta \mid Y_1, \cdots, Y_m)$を比較する. $\pi_m(\theta)$は無情報事前分布の事後分布であるから ESS はちょうどmであると考えられる. $p(\theta \mid Y_1, \cdots, Y_m)$の情報量がこの何倍であるかによって事後分布の ESS を算出し, 事前分布の ESS を求める.

$\pi_m(\theta)$の分散をV_1, 事後分布$p(\theta \mid Y_1, \cdots, Y_m)$の分散を$V_2$とおき, ESS を次で与える :

$$ESS = m \cdot (V_1/V_2) - m$$

正規モデルにおいて, この定義はうまく機能する. 分散の比が無情報事後分布と事後分布の情報量の比になると考えれば, $\pi_m(\theta)$の ESS はmであるから, $m \cdot (V_1/V_2)$が事後分布$p(\theta \mid Y_1, \cdots, Y_m)$の ESS である. したがって事前分布の ESS は$m \cdot (V_1/V_2) - m$と考えることができる. 一方でこの定義はいくつかの問題がある.

- 観察データ$Y_1, \cdots, Y_m$によって ESS が変わるため計画段階で ESS を算出することが難しい.
- 分散の比は変数変換によって変わるため, 本質的に同じ分布であっても異なる ESS が与えられる. 例えば対数変換によって ESS が変わる.
- モデルによっては分散を情報量の指標として用いることはできない. モデルごとに適切な指標を考える必要がある. 特に多変数パラメータの場合これは簡単でない.

2.3. モデルベースの ESS

この節では観察データによらず, モデルによって与えられる ESS を考える.

2.3.1. ESS_{MTM}の定義

ESS_{MTM}は Morita et al. (2008) [9]で与えられた ESS である. この ESS ではまず, 事前分布$p(\boldsymbol{\theta})$の情報量を次の式で定義する. パラメータ$\boldsymbol{\theta}$は多変数でもよい:

$$i(p(\bar{\boldsymbol{\theta}})) = tr\left\{-\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log p(\boldsymbol{\theta})\right\}_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}$$

ここで, 記法は以下のとおり.

- trは行列の対角成分の和を意味する.
- $\bar{\boldsymbol{\theta}}$は事前分布の期待値であり, $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$は偏導関数$\partial \boldsymbol{\theta}^2 \log p(\boldsymbol{\theta})$の$\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$における値であることを意味する.

この情報量において, m例のデータを与えた無情報事後分布$\pi_m(\boldsymbol{\theta})$と$p(\boldsymbol{\theta})$の情報量の差が最も小さくなる$m$を$ESS_{MTM}$とおく:

$$ESS_{MTM} = \underset{m}{\operatorname{argmin}} \left| i(p(\bar{\boldsymbol{\theta}})) - E_{\mathbf{Y}^m}[i(\pi_0(\bar{\boldsymbol{\theta}} \mid Y_1, \cdots, Y_m))] \right|$$

ここで$E_{\mathbf{Y}^m}$は事前分布の予測分布$\int f(Y \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$による期待値. これを式変形すると次を得る [10].

$$ESS_{MTM} = \frac{i(p(\bar{\boldsymbol{\theta}})) - i(\pi_0(\bar{\boldsymbol{\theta}}))}{i_{MODEL}(\bar{\boldsymbol{\theta}})}$$

ここで$i_{MODEL}(\bar{\boldsymbol{\theta}})$はモデル$f(Y \mid \boldsymbol{\theta})$の$\bar{\boldsymbol{\theta}}$における情報量である:

$$i_{MODEL}(\bar{\boldsymbol{\theta}}) = -E_{Y^1}\left[tr\left\{\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log f(Y \mid \boldsymbol{\theta})\right\}_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}\right].$$

すなわちESS_{MTM}は「事前分布の情報量」を「モデルの一人分の情報量」で割った値であるといえる．この定義は多くのモデルに当てはめることが可能である．一方，問題点として$\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$の周りの局所的な値のみによって ESS を定義していることが挙げられる．事前分布が正規分布であれば$i(p(\boldsymbol{\theta}))$は任意の$\boldsymbol{\theta}$で一定であり問題ないが，例えば t 分布のような裾の広い分布では$i(p(\bar{\boldsymbol{\theta}}))$は周辺の$i(p(\boldsymbol{\theta}))$より大きい値であり，ESS を過大評価することになる．

2.3.2. ESS_{ELIR}の定義

Neuenschwander et al. (2020) [10]は$\boldsymbol{\theta}$が一変数の場合にESS_{MTM}を改善し，*expected local-information-ratio ESS* (以下ESS_{ELIR}) を与えた．これは次で定義される:

$$ESS_{ELIR} = E_\theta\left[\frac{i(p(\theta))}{i_F(\theta)}\right]$$

ここでE_θは事前分布$p(\theta)$による期待値であり，$i_F(\theta)$はフィッシャー情報量:

$$i_F(\theta) = -E_{Y|\theta}\left[\frac{d^2}{d\theta^2}\log f(Y \mid \theta)\right]$$

である．$i_{MODEL}(\theta)$との違いは予測分布$\int f(Y \mid \theta)p(\theta)d\theta$による期待値を用いるか，モデル$f(Y \mid \theta)$による期待値を用いるかの差である．$ESS_{ELIR}$は各$\theta$での「事前分布の情報量と一人分の情報量の比」(local-information ratio) の期待値であり，大域的な値になっている．また次の予測的無矛盾性 (predictive consistency) を持つことが証明できる [10].

予測的無矛盾性 (predictive consistency)

予測分布に従う観察データ$\mathbf{Y^m} = Y_1, Y_2, \cdots, Y_m$を与えるとき，事後分布の$ESS_{ELIR}$の期待値は事前分布の$ESS_{ELIR} + m$となる．

$$E_{\mathbf{Y^m}}[p(\theta \mid \mathbf{Y^m})のESS_{ELIR}] = (p(\theta)のESS_{ELIR}) + m$$

すなわち「観察データをm例与えると ESS がm大きくなる」という性質であり，自然である．このようにθが 1 次元の場合ESS_{ELIR}は自然な値といえるが，一方で回帰分析のような多次元の場合に定義されていない．

2.3.3. Clarke (1996) における ESS の定義

予測的無矛盾性を持ち一部の多変数パラメータモデルで定義できる ESS として$ESS_{E\text{-}F}$がある．この定義の前に Clarke (1996) [11]で定義された ESS について述べる．簡単にいえばこれは「無情報事後分布$\pi_m(\boldsymbol{\theta} \mid \mathbf{Y^m})$のうち最も事前分布$p(\boldsymbol{\theta})$に*似ているもの*」を探索し，その$m$を ESS として定義する．ここで*似ている*とは Kullback-Leibler divergence [12](以下 KLd) を最小にする$\mathbf{Y^m}$を選択することを指す:

$$KL\langle p(\boldsymbol{\theta}) || \pi_m(\boldsymbol{\theta} \mid \mathbf{Y^m})\rangle = \int p(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta})}{\pi_m(\boldsymbol{\theta} \mid \mathbf{Y^m})} d\boldsymbol{\theta}$$

KLd は 0 以上の実数であり 0 に近いほど$p(\boldsymbol{\theta})$と$\pi_m(\boldsymbol{\theta} \mid \mathbf{Y^m})$が近いことを示す指標として統計学でよく用いられる．すなわち Clarke (1996) は effective sample を

$$\text{effective sample} = \underset{\mathbf{Y^m}}{\operatorname{argmin}}\, KL\langle p(\boldsymbol{\theta}) || \pi_m(\boldsymbol{\theta} \mid \mathbf{Y^m})\rangle$$

で与え, ESS を effective sample の例数mとして与える．

Clarke (1996)の ESS は簡明であり任意のモデルに対して用いることができる．一方で ESS が大きいとき計算

量が非常に大きくなる．実際mが大きくなると$\mathbf{Y^m}$の取り方も多くなり，KLd についての最適化問題は莫大な量の計算を必要とする．また$KL\langle p(\boldsymbol{\theta})||\pi_m(\boldsymbol{\theta}\mid\mathbf{Y^m})\rangle$は「$\pi_m(\boldsymbol{\theta}\mid\mathbf{Y^m})$を基準とする$p(\boldsymbol{\theta})$のエントロピー」を意味しており，$\mathbf{Y^m}$によって基準が変わる．これが一因しこの ESS は予測的無矛盾性を持たない．

2.3.4. $ESS_{E\text{-}F}$の定義

ESS for exponential family (以下$ESS_{E\text{-}F}$) は Tamanoi (2024) [13]で定義された ESS であり，モデルが指数型分布族の場合に Clarke (1996) の問題点を解決したものである．指数型分布族には観察データのモデルが正規分布や二項分布，線形回帰，ロジスティック回帰の場合が含まれ，逆にワイブル分布，プロビット回帰，コックス回帰，混合モデルなどが含まれない．まず，次の*一人分の尤度*を考える:

$$L(\boldsymbol{\theta}\mid\mathbf{Y^m})=\left(\prod_{i=1}^{m}f\,(Y_i\mid\boldsymbol{\theta})\right)^{1/m}$$

これは$Y_1,\cdots,Y_m$の尤度を平均したものであり，例えば 100 例のデータから 80 例分の情報を得るときなど (power prior) によく用いられる．$f(Y\mid\theta)$が指数分布族の場合，この一人分の尤度は$Y_1,Y_2,\cdots,Y_m$の十分統計量のみで与えられる [13]．よって十分統計量$t_1,t_2,\cdots,t_K$を用いて一人分の尤度を$L(\boldsymbol{\theta}\mid t_1,t_2,\cdots,t_K)$と表すことができる．指数分布族の十分統計量の次元$K$は高々パラメータの個数であるから，ESS が大きい場合にも計算量が大きくなることはない．

次に一人分の尤度を用いて次の*事前分布$p(\theta)$からα例分の情報を引いた分布*を与える:

$$p_{-\alpha}(\boldsymbol{\theta}\mid t_1,t_2,\cdots,t_K)=\frac{1}{Z_\alpha}p(\boldsymbol{\theta})L(\boldsymbol{\theta}\mid t_1,t_2,\cdots,t_K)^{-\alpha}$$

ここでZ_αは正規化定数である．通常の事後分布と異なり$-\alpha$乗をとることで情報を引くことを実現している．しかし$Z_\alpha=\int p(\boldsymbol{\theta})L(\boldsymbol{\theta}\mid t_1,t_2,\cdots,t_K)^{-\alpha}d\boldsymbol{\theta}$は一般に収束しない．そこで$p(\theta)$の 95%の区間 (Bayesian evidence interval [14]) 上で$p_{-\alpha}(\boldsymbol{\theta}\mid t_1,t_2,\cdots,t_K)$を考える．

最後に無情報事前分布$\pi_0(\boldsymbol{\theta})$と$p_{-\alpha}(\boldsymbol{\theta}\mid t_1,t_2,\cdots,t_K)$の間の KLd を考え，これを最小にする$\alpha$を$ESS_{E\text{-}F}$とおく:

$$ESS_{E\text{-}F}=\underset{\alpha}{\operatorname{argmin}}\left(\min_{t_1,\cdots,t_K}KL\langle p_{-\alpha}(\boldsymbol{\theta}\mid t_1,\cdots,t_K)||\pi_0(\boldsymbol{\theta})\rangle\right)$$

この定義では固定された$\pi_0(\boldsymbol{\theta})$に対する$p_{-\alpha}(\boldsymbol{\theta}\mid t_1,\cdots,t_K)$のエントロピーを考えており，$ESS_{E\text{-}F}$は予測的無矛盾性を持つ [13]．また，定義の中で微分を用いていないためモンテカルロ積分による近似計算が可能であり，複雑な事前分布に対しても尤度と事前分布に従うシミュレーションサンプル$\boldsymbol{\theta_1},\cdots,\boldsymbol{\theta_N}$ ($N=10000$など) があれば近似可能である．一方で以下の問題点が挙げられる．

- モデルが指数型分布族に含まれなければ定義できない．
- Clarke (1996) に比べて改善しているものの，回帰分析などでKが大きいとき依然計算量が大きい．
- KLd の最小値は必ずしも 0 にならない．この値が十分小さくない場合$ESS_{E\text{-}F}$は正当化できない．

KLd の最小値に関して，Tamanoi (2025) で RKL という指標が導入されている．これは次で定義される．

$$RKL=\frac{KL_{min}}{KL\langle p(\boldsymbol{\theta})||\pi_0(\boldsymbol{\theta})\rangle}$$

ここでKL_{min}は KLd の最小値である．すなわち RKL は KL の最小値と KL の初期値の比である．$0\le RKL\le 1$であり，$RKL=0$のとき理想的である．RKL の大小によって$ESS_{E\text{-}F}$の正当性を評価できるが，許容される RKL の上限については議論されていない．

3. 実装例

この節ではESS_{MTM}, ESS_{ELIR}, $ESS_{E\text{-}F}$の実装例を紹介する. 添付の macro ファイルと sas ファイルによって再現が可能である. SAS version 9.4 を用いている.

3.1. 一変数の場合 (二項モデルにおけるベータ分布の ESS)

最も単純な例として, $Y = 0$ or 1の二項モデルにおけるベータ分布$Beta(a, b)$の ESS を sas を用いて計算する. 前述の通りこのベータ分布の ESS は古典的に$a + b$であることが知られている.

3.1.1. ESS_{MTM}

添付の ESS_MTM.mac に SAS マクロ関数「macro_ESS_MTM」が含まれており, これを用いてESS_{MTM}を計算することが可能である. 計算するために以下の値が必要である:

- 事前分布の期待値 $\bar{\boldsymbol{\theta}}$
- 事前分布の対数二階微分. 事前分布の情報量$i(p(\bar{\boldsymbol{\theta}}))$の算出に用いる.
- モデルの対数二階微分. モデルの情報量 $i_{MODEL}(\boldsymbol{\theta})$の算出に用いる.
- 無情報事前分布の対数二階微分. 無情報事前分布の情報量 $i(\pi_0(\boldsymbol{\theta}))$の算出に用いる.
- 予測分布に従うサンプルが 10000 例程度含まれるデータセット

事前分布の期待値, 対数二階微分については直接計算する必要がある. 対数二階部分は一般に簡単ではないが, Wolfram Alpha のようなオープンシステムによっても計算可能である. ベータ分布$Beta(p \mid a, b) \propto p^{a-1}(1-p)^{b-1}$について, 期待値は$a/(a+b)$, 対数二階微分は

$$\frac{d^2}{dp^2}\log Beta(p \mid a, b) = -\frac{a-1}{p^2} - \frac{b-1}{(p-1)^2}$$

である.

モデルの情報量について, モデル$f(Y \mid \boldsymbol{\theta})$の対数二階微分の期待値をとる必要がある. 主要なモデルについては対数二階偏微分の一覧が ESS_MTM.sas に含まれている. 二項モデルの場合 (すなわちベルヌーイ分布 $Ber(Y \mid p) = p^Y \cdot (1-p)^{1-Y}$) では

$$\frac{\partial^2}{\partial p^2}\log Ber(Y \mid p) = -\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}$$

である. この対数二階微分の期待値をとる必要があるが, このためにモンテカルロ積分による近似を用いる. すなわち予測分布 $\int Ber(Y \mid p)Beta(p \mid a, b)dp$ に従うシミュレーションサンプル $Y_1, Y_2, \cdots, Y_n$を 10000 例程度発生させることで次の近似が可能になる.

$$i_{MODEL}(p) \simeq \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial p^2}\log f(Y_i \mid p)$$

このため macro_ESS_MTM では予測分布に従うサンプルが必要になる. 以下のコードによって計算可能である. 引数はそれぞれ次を意味する:

- DATA: 予測分布に従うサンプルが含まれるデータセット. 変数 Y にサンプルを与える.
- numParameter: パラメータの個数. 今回はpのみであるから 1.
- parameter_1: パラメータを表す記号. 以降の引数で用いる記号.
- priorExpectedValue_1: 事前分布の期待値. ここでは$a/(a+b)$.

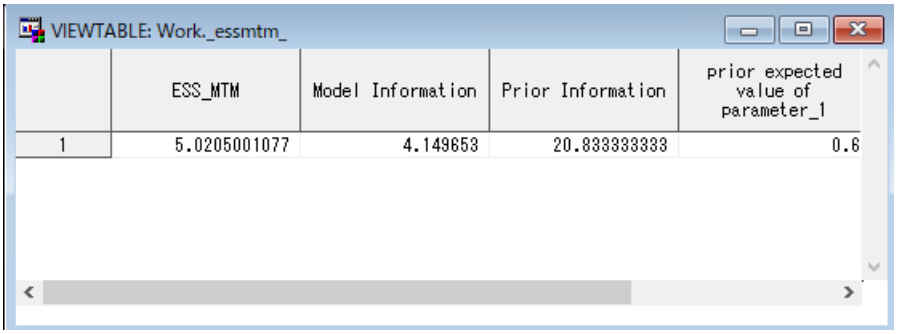
- logModel_second_deriv_1: モデルの対数二階微分．ここでは$\frac{\partial^2}{\partial p^2}\log Ber(Y \mid p)$.

- logPrior_second_deriv_1: 事前分布の対数二階微分．ここでは$\frac{d^2}{dp^2}\log Beta(p \mid a, b)$.

- logNoninfo_second_deriv_1: 無情報事前分布の対数二階微分．ここでは$\frac{\partial^2}{\partial p^2}\log(p^{-1} \cdot (1-p)^{-1})$.

```
%inc 'ESS_MTM.mac' ; /* マクロの読み込み */
/* ハイパーパラメータを指定する． */
%let alpha = 3;
%let beta  = 2;     /* ESSはalpha + betaになることが知られている */
/* 予測分布に従うサンプルを10000例発生させる */
DATA priorPredictiveSamples;
  DO n = 1 TO 10000;
    p = RAND("BETA", &alpha., &beta.);
    Y = RAND("BERNOULLI", p);
    OUTPUT;
  END;
  DROP n;
RUN;
/* マクロの実行 */
%macro_ESS_MTM(
  DATA = priorPredictiveSamples, /* 予測分布のサンプル */
  numParameter = 1, /* パラメータの個数 */
  parameter_1 = p, /* パラメータの記号.多変量パラメータではその分与える */
  priorExpectedValue_1 = &alpha / (&alpha + &beta), /* 事前分布の期待値 */
  logModel_second_deriv_1 = -Y/p**2 - (1-Y)/(1-p)**2, /* モデルの対数二階微分 */
  logPrior_second_deriv_1 = -(&alpha.-1)/p**2 - (&beta.-1)/(1-p)**2, /* 事前分布の対数二階微分 */
  logNoninfo_second_deriv_1 = 1/p**2 + 1/(1-p)**2 /* 無情報事前分布の対数二階微分 */
);
```

計算結果はデータセット_ESSMTM_で与えられる．



	ESS_MTM	Model Information	Prior Information	prior expected value of parameter_1
1	5.0205001077	4.149653	20.833333333	0.6

変数 ESS_MTM はESS_{MTM}の計算結果であり，Model Information, Prior Information はそれぞれ$i_{MODEL}(\bar{\boldsymbol{\theta}})$,

$i(p(\bar{\boldsymbol{\theta}}))$である. DATA のサンプル数を増やすことで精度を上げることができる.
logModel_second_deriv_1 や logNoninfo_second_deriv_1 について，主要なモデルでは一覧が与えられている．

3.1.2. ESS_{ELIR}

添付の ESS_ELIR.mac に SAS マクロ関数「macro_ESS_ELIR」が含まれており，これを用いてESS_{ELIR}を計算できる. 計算するために以下の値が必要である:

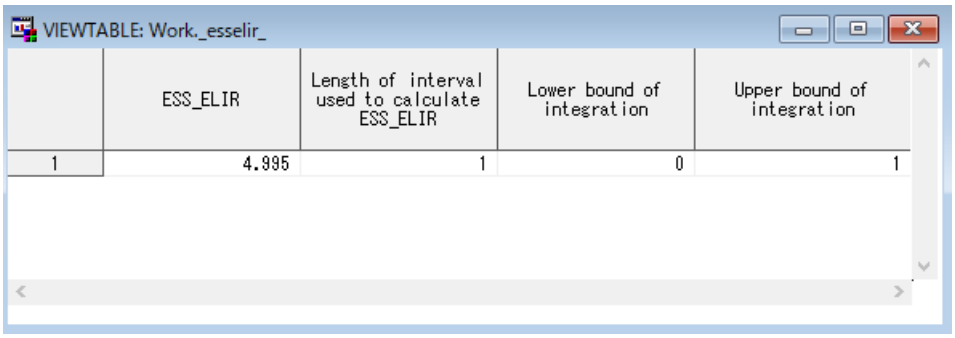
- 事前分布の情報量, すなわち事前分布の対数二階微分 $i(p(\theta))$
- モデルのフィッシャー情報量 $i_F(\theta)$
- 情報量の比の期待値を計算するために以下が必要
    - 事前分布の確率密度関数
    - パラメーターの積分範囲

$i(p(\theta))$についてはESS_{MTM}と同様である. 主要なモデルのフィッシャー情報量については ESS_ELIR.sas に含まれている. 与えられた情報量の比に対して期待値を計算する. ESS_{ELIR}は一変数パラメータしか扱わないためこれは一次元の積分でありモンテカルロ積分による近似は必要ない. 代わりに積分範囲と事前分布の確率密度関数が必要である. 以下のコードによって計算可能である. 引数はそれぞれ次を意味する:

- parameter: パラメータを表す記号. 以降の引数で用いる記号.
- FisherInformation: モデルのフィッシャー情報量.
- logPrior_second_deriv: 事前分布の対数二階微分.
- from: 積分範囲の下限.
- to: 積分範囲の上限.
- breaks: 積分の近似に用いる分割の個数.
- priorDensity: 事前分布の確率密度関数.

```
/* マクロ読み込み*/
%inc 'ESS_ELIR.mac' ;
/* ハイパーパラメータの指定 */
%let alpha = 3;
%let beta  = 2;
/* マクロを実行 */
%macro_ESS_ELIR(
  parameter = p, /* パラメータの記号 */
  FisherInformation = 1/p/(1-p), /* フィッシャー情報量 */
  logPrior_second_deriv = -(&alpha.-1)/p**2 - (&beta.-1)/(1-p)**2, /* 事前分布の対数二階微分 */
  from  = 0, /* 積分範囲の下限 */
  to    = 1, /* 積分範囲の上限 */
  breaks= 100, /* 積分の近似に用いる分割の個数 */
  priorDensity= pdf("BETA", p, &alpha, &beta) /* 事前分布の確率密度関数 */
);
```

データセット_ESSELIR_にESS_{ELIR}の計算結果が表示される.



ESS_ELIR はESS_{ELIR}の計算結果であり，他は期待値の計算に用いた積分範囲を示している．
代表的なモデルについてはフィッシャー情報量の一覧が ESS_ELIR.sas にあたえられている．

3.1.3. $ESS_{E\text{-}F}$

$ESS_{E\text{-}F}$では偏微分は用いない．一方で$KL\langle p_{-\alpha}(\boldsymbol{\theta} \mid t_1,\cdots,t_K)||\pi_0(\boldsymbol{\theta})\rangle$を最小化する$\alpha$，$t_1,\cdots,t_K$を探索する必要があるためマクロの中身は非常に複雑である．主要なモデルについては簡易に計算可能なマクロをそれぞれ用意しており，二項モデルでは「macro_ESS_EF_Bernoulli」を用いて$ESS_{E\text{-}F}$を計算することが可能である．計算するため，以下の値が必要である．

- 事前分布に従うサンプル 10000 例程度
- Bayesian evidence interval の信頼水準. 95%が標準.
- 事前分布の対数確率密度関数
- $t_1,t_2,\cdots,t_k$が与えられたとき$KL\langle p_{-\alpha}(\boldsymbol{\theta} \mid t_1,\cdots,t_K)||\pi_0(\boldsymbol{\theta})\rangle$を最小にする$\alpha$を求めるために以下が必要(二分探索法を用いる)
    - αの初期値
    - 計算回数．すなわち二分割する回数
- $KL\langle p_{-\alpha}(\boldsymbol{\theta} \mid t_1,\cdots,t_K)||\pi_0(\boldsymbol{\theta})\rangle$を最小にする$t_1,t_2,\cdots,t_k$を Nelder Mead 法 [15]によって求める．そのために以下が必要:
    - simplex の初期値の設定方法．自動で作成するか，自分で作成するか選択が可能.
    - Nelder Mead 法のループ回数

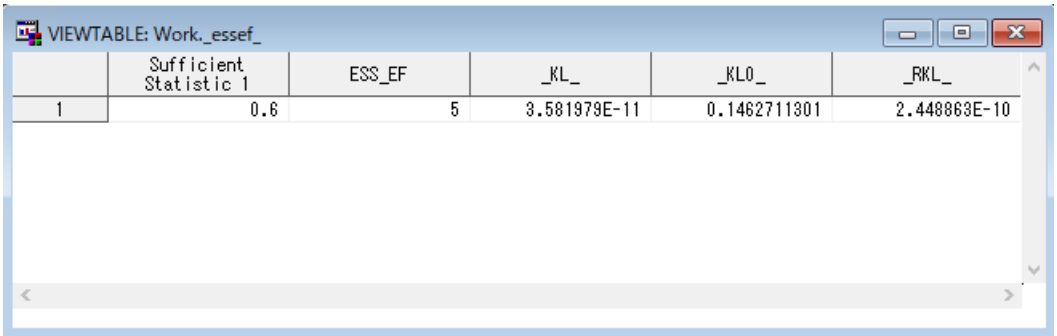
注意として，計算過程で作業用のデータセットを複数用いる．同じ名前のデータセットがある場合削除される．以下のコードによって計算可能である．引数はそれぞれ次を意味する:

- data_priorSamples: 事前分布のサンプルが含まれるデータセット．parameter で指定する変数名を用いる．
- logPriorDensity: 事前分布の確率密度関数を対数変換したもの.
- initial_alpha: KLd を最小にするαを求めるための初期値.
- loop_depth: 二分探索法の分割回数.
- simplex_init_method: Nelder Mead 法で用いる simplex の初期値を指定する方法. RANDOM の場合マクロ内で作成する．そうでない場合はデータセット_current_simplex_を用意する必要がある.
- simplex_seed: RANDOM の場合シード値を指定可能．指定しないことも可能.
- LoopTimes_NM: Nelder Mead 法のループ回数．パラメータが多いほどこの回数を増やす必要がある.
- sampleLimittingLevel: Bayesian evidence interval の信頼水準．デフォルトは 0.95.

```
/* マクロを読み込む */
```

```
%inc 'ESS_EF_given_t.mac';
%inc 'ESS_EF_NelderMead.mac';
%inc 'ESS_EF.mac';
/* ハイパーパラメータの設定 */
%let alpha = 3;
%let beta  = 2;
/* 事前分布のサンプルを10000例発生させる */
DATA priorSamples;
  DO n = 1 TO 10000;
    p = RAND("BETA", &alpha, &beta);
    OUTPUT;
  END;
RUN;
/* マクロを実行 */
%macro_ESS_EF_Bernoulli(
  data_priorSamples = priorSamples, /* 事前分布のサンプル */
  logPriorDensity = %str(logpdf("BETA", p, &alpha, &beta)), /* 事前分布の対数確率密度関数 */
  initial_alpha = 1, /* alphaの初期値 */
  loop_depth = 6, /* 各tにおける二分探索法のループ回数 */
  simplex_init_method = RANDOM,  /* RANDOM or MANUAL */
  simplex_seed = , /* RANDOMの場合シードを指定できる */
  LoopTimes_NM = 10, /* Nelder Mead 法のループ回数 */
  sampleLimittingLevel = 0.95 /* Bayesian evidence intervalの水準 */
);
```

データセット_ESSEF_に$ESS_{E\text{-}F}$の計算結果が表示される.



	Sufficient Statistic 1	ESS_EF	_KL_	_KL0_	_RKL_
1	0.6	5	3.581979E-11	0.1462711301	2.448863E-10

Sufficient Statistics 1 は$KL\langle p_{-\alpha}(\boldsymbol{\theta} \mid t_1,\cdots,t_K)||\pi_0(\boldsymbol{\theta})\rangle$を最小にする$t_1,\cdots,t_K$の値であり, ESS_EF は$ESS_{E\text{-}F}$, _KL_ は KLd の最小値, _KL0_は事前分布と無情報事前分布の間の KLd, _RKL_は$RKL = KL_{min}/KL_0$である. この例は RKL=0 であり, 理想的な例になっている.

3.2. 回帰モデルの場合 (線形回帰における多変量正規分布の ESS)

回帰モデルでは説明変数の分布が ESS に影響を与える. 例えば説明変数として年齢と投与群 (0 or 1) を用いる場合, 年齢の標準偏差は投与群の標準偏差より大きい. よって年齢の回帰係数 = 1 と投与群の回帰係数 =

1 では意味が全く異なる．したがって回帰係数の事前分布が同じであっても ESS は同じにはならない．説明変数の分布は試験前に予測することは一般に難しい．そのため説明変数に合わせて事前分布を調整するか，各説明変数を正規化，すなわち平均 0，分散 1 になるようにスケールを直してから計算を行う必要がある．
本節では線形回帰の事前分布の ESS を計算するマクロを紹介する．上述のように説明変数の分布が影響することに注意が必要である．説明変数x_1, x_2, x_3に対して応答変数Yが

$$f(Y \mid \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_p x_p + \epsilon$$
$$\epsilon \sim N(0, \sigma^2)$$

に従うとする．説明変数は互いに独立かつ標準正規分布に従うとし，事前分布は

$$\beta_i \sim N(0, 1^2)$$

とする．また簡単のためここではσ^2を既知とするが，未知とした場合も計算可能である．

3.2.1 ESS_{MTM}

基本的に 4.1.1 節と大きく変わらないが次の 2 点に注意が必要である．

- 予測分布のサンプルを作成する際，説明変数のサンプルを同時に作成する必要がある．
- パラメータが 4 変数であるから偏微分や期待値もその分計算する必要がある．引数 logModel_second_deriv_1, logModel_second_deriv_2, logModel_second_deriv_3,…のように増やせばよい．

```
%let model_var = 1;

DATA priorPredictiveSamples;

  DO n = 1 TO 10000;

    x_1 = RAND("NORMAL", 0, 1); x_2 = RAND("NORMAL", 0, 1); /* 説明変数のサンプル */

    beta_0 = RAND("NORMAL", 0, 1);

    beta_1 = RAND("NORMAL", 0, 1);

    beta_2 = RAND("NORMAL", 0, 1);

    Y = beta_0 + x_1 * beta_1 + x_2 * beta_2 + RAND("NORMAL", 0, sqrt(&model_var.));

    OUTPUT;

  END;

  DROP n beta_0 beta_1 beta_2;

RUN;

%macro_ESS_MTM(

  DATA                    = priorPredictiveSamples, /* 予測分布のサンプルYと説明変数のサンプルx_nを含む */

  numParameter             = 3, /* 切片を含む */

  parameter_1              = beta_0,

  parameter_2              = beta_1,

  parameter_3              = beta_2,

  priorExpectedValue_1      = 0,

  priorExpectedValue_2      = 0,

  priorExpectedValue_3      = 0,

  logModel_second_deriv_1   = - 1 / &model_var.,

  logModel_second_deriv_2   = - x_1**2 / &model_var.,
```

```
  logModel_second_deriv_3   = - x_2**2 / &model_var.,
  logPrior_second_deriv_2   = - 1 / 1,
  logPrior_second_deriv_1   = - 1 / 1,
  logPrior_second_deriv_3   = - 1 / 1,
  logNoninfo_second_deriv_1 = 0, /* 無除法事前分布は一様分布 */
  logNoninfo_second_deriv_2 = 0, /* 無除法事前分布は一様分布 */
  logNoninfo_second_deriv_3 = 0  /* 無除法事前分布は一様分布 */
);
```

3.2.2 $ESS_{E\text{-}F}$

前述の通り$ESS_{E\text{-}F}$はモデルが指数分布族であるときにしか定義されないが，回帰分析の場合これは正準リンク (canonical link) を持つ一般化線形モデルに対応する．すなわち

$$f(Y \mid \boldsymbol{\beta}, \boldsymbol{X}) = \exp\left\{\frac{Y \cdot \boldsymbol{X}^{\top}\boldsymbol{\beta} - a(\boldsymbol{X}^{\top}\boldsymbol{\beta})}{\phi} + c(Y, \phi)\right\}$$

で表されるモデルである．例えば線形回帰，ロジスティック回帰，ポアソン回帰は含まれるが，Cox 回帰は一般化線形モデルではないため含まれず，プロビット回帰は正準リンクではないため含まれない．
回帰分析では説明変数のサンプルが含まれるデータセットを用意する必要がある．

```
/* 説明変数のサンプル */
DATA explanatory_vars;
  DO n = 1 TO 1000;
    x0 = 1; x1 = RAND("NORMAL", 0, 1); x2 = RAND("NORMAL", 0, 1); /* x0は切片 */
    OUTPUT;
  END;
RUN;
/* 事前分布のサンプル */
DATA PriorSamples;
  DO n = 1 TO 10000;
    beta0 = RAND("NORMAL", 0, 1); beta1 = RAND("NORMAL", 0, 1); beta2 = RAND("NORMAL", 0, 1);
    OUTPUT;
  END;
RUN;
/* マクロを実行 */
%macro_ESS_EF_linearRegression(
  variance_error = 1, /* 既知の分散の値 */
  data_priorSamples = PriorSamples, /* 事前分布のサンプル */
  data_explanatory_vars = explanatory_vars, /* 説明分布のサンプル */
  logPriorDensity = logpdf("NORMAL", beta0, 0, 1) + logpdf("NORMAL", beta1, 0, 1) + logpdf("NORMAL",
beta2, 0, 1),
  initial_alpha = 0.1, /* alphaの初期値 */
```

```
  loop_depth = 6,  /* 各tにおける二分探索法のループ回数 */
  simplex_init_method = RANDOM,  /* RANDOM or MANUAL */
  simplex_seed = , /* RANDOMの場合シードを指定できる */
  NumParameter = 3, /* パラメータの個数 */
  parameter_1 = beta0, /* パラメーターの記号．例えば切片 */
  parameter_2 = beta1, /* パラメーターの記号．例えば説明変数1の傾き */
  parameter_3 = beta2, /* パラメーターの記号．例えば説明変数2の傾き */
  explanatory_var_1 = x0, /* 説明変数の記号．切片 x0 = 1 */
  explanatory_var_2 = x1, /* 説明変数の記号．*/
  explanatory_var_3 = x2, /* 説明変数の記号．*/
  LoopTimes_NM = 50,
  sampleLimittingLevel = 0.95 /* Bayesian evidence intervalの水準 */
);
```

4. まとめ

本稿でベイズ統計の簡単な説明と ESS の紹介，いくつかの ESS の定義, SAS による実装を行った. ESS は事前分布の解析結果への影響の大きさを評価するうえで重要な指標であるが，その定義は簡単ではなく，状況に合わせて適切な ESS を選択する必要がある．以下の表は本稿で紹介した ESS を簡単に比較したものである．

	メリット	デメリット
分散比による ESS	・定義が簡明 ・観察データによって値を変えることができる	・観察データによって値が変わらない ・分散が指標として適しているとは限らない ・多変数パラメータに対して定義されていない
ESS_{MTM}	・観察データによって値が変わらない ・多変数パラメータにも定義できる	・事前分布の期待値周辺のみによって値が決まり，事前分布が正規分布でないとき，ESS を過大評価，過小評価する場合がある
ESS_{ELIR}	・観察データによって値が変わらない ・任意の事前分布で評価ができる ・予測的無矛盾性を持つ	・多変数パラメータでは定義されない.
ESS in Clarke(1996)	・観察データによって値が変わらない ・定義が簡明 ・多変数パラメータにも定義できる	・ESS が大きいとき計算量が非常に大きい ・KLd が小さくならないとき，ESS が正しい値である保証がない.
$ESS_{E\text{-}F}$	・観察データによって値が変わらない ・多変数パラメータにも定義できる ・予測的無矛盾性を持つ	・モデルが指数型分布族でないとき定義されない ・KLd が小さくならないとき ESS が正しい値である保証がない.

ESS はこれらがすべてではない. Morita et al. (2012) [16]は conditionally independent hierarchical models (CIHMs) とよばれる階層ベイズにおける ESS を定義している．また，Wiesenfarth and Calderazzo (2020) [17]は事前分布

と観察データの傾向が大きく異なる場合 (prior-data-conflict) に同じ ESS の事前分布でも解析結果への影響が異なることに注目し, effective current sample size (ECSS) を定義した.

SAS マクロについて, ESS_{MTM}やESS_{ELIR}ではモデルや事前分布の二階偏微分が必要になる. 二階微分の近似は簡単ではなく, 本マクロには含まれていないためユーザーが計算する必要がある. $ESS_{E\text{-}F}$では計算に積分しか用いないため近似計算が可能である. 一方で KLd に関する最適化問題を解く必要があるためコードは非常に複雑である.

臨床開発が複雑化する今日, ベイズ統計の果たす役割は大きくなることが予想されている. 本稿がベイズ統計や ESS について理解する一助になれば幸いである.

参考文献

[1] 柳本武美, “望ましい無情報事前分布の基準,” 数理解析研究所講究録, 京都大学, 2022.

[2] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society A,* vol. 186, no. 1007, p. 453–461, 1946.

[3] J. M. Bernardo, "Reference Posterior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society. Series B,* vol. 41, no. 2, pp. 113-147, 1979.

[4] FDA, "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," The United States Food and Drug Administration, Rockville, Maryland, 2010.

[5] 平川晃弘, 佐藤宏征, 井桁正尭, 藤川桂, 堀口剛, 大門貴志 , 手良向聡, “希少疾患領域の治験におけるベイズ流アプローチの利用可能性と留意事項,” *薬理と治療,* 第 49 巻, 第 1, pp. 72-79, 2021.

[6] FDA, "Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products," The United States Food and Drug, Rockville, Maryland., 2020.

[7] M. R. Novick and W. J. Hall, "A Bayesian Indifference Procedure," *Journal of the American Statistical Association,* vol. 60, no. 312, pp. 1104-1117, 1965.

[8] D. Malec, "A closer look at combining data among a small number of binomial experiments," *Statistics in Medicine,* vol. 20, no. 12, p. 1811–1824, 2001.

[9] S. Morita, P. F. Thall and P. Müller, "Determining the effective sample size of a parametric prior," *Biometrics,* vol. 64, no. 2, p. 595–602, 2008.

[10] B. Neuenschwander, S. Weber, H. Schmidli and A. O’Hagan, "Predictively consistent prior effective sample sizes," *Biometrics,* vol. 76, no. 2, p. 578–587, 2020.

[11] B. Clarke, "Implications of reference priors for prior information and for sample size," *Journal of the American Statistical Association,* vol. 91, no. 433, p. 173–184, 1996.

[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics,* vol. 22, no. 1, p. 79–86, 1951.

[13] R. Tamanoi, "Prior effective sample size for exponential family distributions with multiple parameters," *Statistical*

Analisis and Data Mining, vol. e11685, 2024.

[14] R. Kelter, "The evidence interval and the Bayesian evidence value: On a unified theory for Bayesian hypothesis testing and interval estimation," *British Journal of Mathematical and Statistical Psychology,* vol. 75, no. 3, p. 550–592, 2022.

[15] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal,* vol. 7, no. 4, p. 308–313, 1965.

[16] S. Morita, P. F. Thall and P. Müller, "Prior effective sample size in conditionally independent hierarchical models," *Bayesian Analysis,* vol. 7, no. 3, p. 591–614, 2012.

[17] M. Wiesenfarth and S. Calderazzo, "Quantification of prior impact in terms of effective current sample size," *Biometrics,* vol. 76, no. 1, p. 326–336, 2020.

SASによる条件付き検出力の算出と性能比較

○栗田凪，魚住龍史

東京科学大学工学院経営工学系

Calculation of conditional power and performance comparison using SAS

Nagi Kurita and Ryuji Uozumi
Department of Industrial Engineering and Economics, Institute of Science Tokyo

要旨

臨床試験において，有効性や安全性に関する試験治療群間の比較を目的として試験完了前に行われる解析を中間解析という．中間解析における早期試験中止には，治療法の効果を早い段階で認める有効中止と，治療法に見切りをつけ早い段階であきらめる無効中止が存在する．中間解析における無効中止指標のひとつとして，条件付き検出力 (Conditional Power，CP) が提案されている．条件付き検出力には，最終解析における棄却確率のみを考える算出方法と，当該中間解析以降全ての中間解析における中止確率を考慮する算出方法が存在するが，これらの性能は詳しく評価されていない．本稿では，無効中止のみを想定した中間解析に着目し，SAS による条件付き検出力の算出方法を示すとともに，算出結果を真値と比較することで性能を評価した．シミュレーションの結果，2 群間の平均の差における標準偏差を過大に推定した場合を除き、当該中間解析以降すべての中間解析における中止確率を考慮する算出方法が優れていることが示唆された．

キーワード：中間解析，条件付き検出力，SEQDESIGN プロシジャ，SEQTEST プロシジャ，検定統計量，情報量，情報分数

1. はじめに

中間解析は，臨床試験において，有効性や安全性に関する試験治療群間の比較を目的として，試験の途中段階で行われる解析である．中間解析の役割のひとつは，早期中止の判断を下すことである．早期中止には，治療法の効果を早い段階で認める有効中止と，治療法に見切りをつけ早い段階であきらめる無効中止が存在する．無効中止を目的とした中間解析では，無効中止指標として条件付き検出力(Conditional Power，以下 CP)が用いられることがある．条件付き検出力は，ある中間解析時点で得られたデータで条件付けた下，帰無仮説が偽であるときに正しく帰無仮説を棄却する確率として定義される (Lan et al., 1982).

現状，多くの臨床試験で，最終解析において帰無仮説を正しく棄却する確率に相当する条件付き検出力の算出方法が用いられている．しかし，中間解析を 2 回以上実施する場合，最終解析のみを考慮して算出される条件付き検出力では，最終解析より前の時点における帰無仮説の棄却および試験継続確率を考慮しないため，真値から乖離するおそれがある．そこで，最終解析だけでなく，当該中間解析後に実施される中間解析

の情報も考慮した，条件付き検出力の算出方法(Zhu et al., 2011) が提案されている．これら 2 種類の条件付き検出力の算出方法は，いずれも SEQTEST プロシジャの CONDPOWER オプションを用いて指定することができる．

本稿では，無効中止のみを中止条件とする中間解析デザインに焦点を当て、最終解析のみを考慮する算出方法と試験継続確率を考慮する算出方法の 2 種類による条件付き検出力を SAS で求める方法について説明する．また，シミュレーションによって条件付き検出力の真値を算出し，2 種類の条件付き検出力と真値を比較する．

2. 条件付き検出力の算出方法

2.1.記法

試験全体で最大$K(\geq 3)$回の解析を行うとする．このとき，第k中間解析の検定統計量と情報量をそれぞれ$Z_k, I_k (k = 1,2 \dots, K)$として，検定統計量の列$\{Z_1, Z_2, \dots, Z_K\}$は多変量正規分布に従っているものとする．また，最終解析における棄却限界値をa_K，第k中間解析における無効中止境界をb_k，効果量をθと表す.

2.2. 第k中間解析の条件付き密度関数

任意の$k < k'$に対して，次の式が成り立つ(Jennison and Turnbull, 2000).

$$Z_{k'}|Z_k \sim N\left(\frac{\theta(I_{k'} - I_k) + z_k\sqrt{I_k}}{\sqrt{I_{k'}}}, \frac{I_{k'} - I_k}{I_{k'}}\right) \tag{1}$$

したがって，第k中間解析の検定統計量Z_kで条件付けた下での第k'中間解析の検定統計量$Z_{k'}$の密度関数は，標準正規分布の確率密度関数φを用いて次のように表せる．

$$f_{k',k}(z_{k'}|z_k) = \frac{\sqrt{I_{k'}}}{\sqrt{I_{k'} - I_k}}\varphi\left\{\frac{z_{k'}\sqrt{I_{k'}} - z_k\sqrt{I_k} - \theta(I_{k'} - I_k)}{\sqrt{I_{k'} - I_k}}\right\} \tag{2}$$

2.3. 条件付き検出力$\mathrm{CP}_{\mathrm{final}}$の算出方法と近似式

第k中間解析以降の中間解析時点を考慮せず，最終解析のみに着目した算出方法として，

$$\mathrm{CP}_{\mathrm{final}}(\theta) = \Pr\{Z_K > a_K | Z_k = z_k\} \tag{3}$$

が提案されている (Lan et al., 1982)．すなわち，(3) の条件付き検出力は，無効中止のみを考える中間解析に対して，第k中間解析時点以降の中間解析において無効中止される確率を考慮せず，最終解析において帰無仮説が正しく棄却される確率として算出される．以降，(3) によって算出された条件付き検出力を$\mathrm{CP}_{\mathrm{final}}$と呼ぶ.

実際に$\mathrm{CP}_{\mathrm{final}}$を算出する場合は，近似式が用いられる．ここで，標準正規分布の累積分布関数Φを用いると，一般に$X \sim N(\mu, \sigma^2)$と定数cに対して，$\Pr\{X > c\} = \Phi\{-(c - \mu)/\sigma\}$が成り立つ．ゆえに，$\mathrm{CP}_{\mathrm{final}}$の近似式は，(1)式において$k' = K, k = k$として，次のように与えられる．

$$\begin{aligned}\mathrm{CP}_{\mathrm{final}}(\theta) &= \Pr\{Z_K > a_K | Z_k = z_k\} \\ &= \Phi\left[-\left\{a_K - \frac{\theta(I_K - I_k) + z_k\sqrt{I_k}}{\sqrt{I_K}}\right\} \Big/ \sqrt{\frac{I_K - I_k}{I_K}}\right]\end{aligned}$$

$$= \Phi\left\{\frac{z_k\sqrt{I_k} + \theta(I_K - I_k) - a_K\sqrt{I_K}}{\sqrt{I_K - I_k}}\right\} \tag{4}$$

近似式(4)を用いて直接CP_{final}を求める SAS コードを次に示す.

```
data CPfinal;
 CP_final = CDF('NORMAL', (z_k1*sqrt(i_k1)+theta*(i_k2-i_k1)-a_k*sqrt(i_k2))/sqrt(i_k2-i_k1));
run;
```

例えば、ステージ数 3 の 1 回目の中間解析に関する条件付き検出力を算出するには，上記の SAS コードにおいて，$k_1 = 1, k_2 = 3$とすればよい.

2.4. 条件付き検出力CP_{all}の算出方法と近似式

一般に，CP_{final}が条件付き検出力として用いられている．しかし近年，無効中止のみを中止条件とする場合について，Zhu et al. (2011) によって新たに次のような条件付き検出力が提案された.

$$CP_{all}(\theta) = \Pr\{b_{k+1} \le Z_{k+1}, b_{k+2} \le Z_{k+2}, \dots, b_{K-1} \le Z_{K-1}, Z_K > a_K | Z_k = z_k\} \tag{5}$$

(5) の条件付き検出力は，第k中間解析時点以降の中間解析を考慮した算出方法である．すなわち，無効中止のみを考える中間解析に対しては，第k中間解析時点以降の中間解析において無効中止となる確率を考慮した，最終解析において帰無仮説が正しく棄却される確率として算出される．以降，(5) によって算出された条件付き検出力をCP_{all}と呼ぶ．CP_{final}との違いとしては，CP_{all}は最終解析より前の時点における帰無仮説の棄却および試験継続の確率を考慮する一方，CP_{final}は最終解析時点のみの帰無仮説の棄却割合を考えている．したがって，中間解析 2 回以上のステージ数 3 以上の中間解析デザインを計画する場合，どちらの条件付き検出力を用いるかで，結果が変わり得る.

無効中止のみを中止条件とする場合，(2)式を用いて，CP_{all}の近似式は次のように与えられる.

$$\begin{aligned} CP_{all}(\theta) &= \Pr\{b_{k+1} \le Z_{k+1}, \dots, b_{K-1} \le Z_{K-1}, Z_K > a_K | Z_k = z_k\} \\ &= \int_{a_K}^{\infty}\int_{b_{K-1}}^{\infty}\cdots\int_{b_{k+1}}^{\infty} f_{k+1,k}(z_{k+1}|z_k)\cdots f_{K,K-1}(z_K|z_{K-1})dz_{k+1}\cdots dz_K \end{aligned} \tag{6}$$

近似式(6)を用いて 1 回目の中間解析におけるCP_{all}を求める SAS コードの一部を次に示す．(6) は積分計算をともなうため，IML を用いた SAS マクロとして示している.

```
%macro k_CPall(k=, z1=, i_values=, theta=, a_k=, b_values=);
    proc iml;
        /* グローバル変数の宣言 */
        k_val = &k;z_vec = j(k_val, 1, .); z_vec[1] = &z1; i_vec = {&i_values};theta = &theta;
        a_k = &a_k;b_vec={&b_values};
        /* 各レベルの被積分関数を定義（k-1個の関数）*/
        %do level = 1 %to %eval(&k-1);
            start f&level(x) global(z_vec, i_vec, theta, a_k, b_vec, k_val);
                z_vec[&level + 1] = x; /* 現在のz値を設定（z_2からz_kまで）*/
                  %if &level < %eval(&k-1) %then %do;
                    %if &level = %eval(&k-2) %then %do;
                        lower = a_k; /*積分範囲の下限に最終解析の棄却限界値を設定*/
                    %end;%else %do;
                    lower = b_vec[&level + 1];%end;upper = .P;
                    /* 次のレベルの積分を実行 */
                    call quad(result, "f%eval(&level+1)", lower||upper);
```

```
                    %end;%else %do;
                      /* 条件付き密度関数の積を計算 */
                      do j = 1 to k_val-1;
                          term = (sqrt(i_vec[j+1])/sqrt(i_vec[j+1]-i_vec[j])) *
                          PDF("NORMAL", (z_vec[j+1]*sqrt(i_vec[j+1]) - z_vec[j]*sqrt(i_vec[j]) -
                          theta*(i_vec[j+1]-i_vec[j])) / sqrt(i_vec[j+1]-i_vec[j]));
                  end;%end;finish;%end;
          /* 最外側の積分を実行 */
          call quad(CP, "f1", lower_outer||upper_outer);print CP;quit;
%mend k_CPall;
```

2.5.SEQTEST プロシジャによる条件付き検出力の算出

2.3 節及び 2.4 節で示したCP_{final}，CP_{all}の算出方法は SAS の SEQTEST プロシジャの CONDPOWER オプションによっても求めることができる．CONDPOWER オプションでは，サブオプションとして TYPE=FINALSTAGE または TYPE=ALLSTAGES を指定することにより，それぞれCP_{final}とCP_{all}に対応した値を算出可能である．

例として，無効中止のみを中止条件とするステージ数 3 の中間解析デザインにおける，SEQTEST プロシジャを用いた条件付き検出力の算出方法として次の SAS コードを示す．

```
/*中間解析デザイン*/
proc seqdesign altref=0.1 ;
   SidedOBrienFleming: design nstages=3 alpha=0.025 beta=0.1
                       stop = accept /*中止条件*/
                       method=obf /*棄却域obf型*/ alt=upper;
   ods output Boundary=Bnd_LDL; /*無効中止境界の取得*/
run;
/*CP_finalの算出*/
proc seqtest Boundary = Bnd_LDL_ Parms(Testvar = Trt) = Parms_IA1_ infoadj=prop
    condpower(cref=0 to 2 by 0.1 TYPE=FINALSTAGE) plots=condpower;
    ods output Test=Test_LDL1_FINAL;
run;
/*CP_allの算出*/
proc seqtest Boundary = Bnd_LDL_ Parms(Testvar = Trt)=Parms_IA1_ infoadj=prop
    condpower(cref=0 to 2 by 0.1 TYPE=ALLSTAGES) plots=condpower;
    ods output Test=Test_LDL1_ALL;
run;
```

SEQTEST プロシジャの前に SEQDESIGN プロシジャによって無効中止境界を求めている．これは，CP_{all}を(6)式によって求めるために，無効中止境界値が必要なためである．また，SEQTEST プロシジャでは，altlef の値と cref の値の積が効果量θに相当する．したがって上記のコード例では，効果量を 0 から 0.2 まで 0.01 ずつ変化させた条件付き検出力の値が得られる．SEQDESIGN プロシジャと SEQTEST プロシジャについての詳細は，付録において述べる．

3. 数値実験

3.1. 状況設定

無効中止のみを中止条件にして，片側 2 標本Z検定を，最終解析を含むステージ数 3 で行う状況を考える．1 回目の中間解析に注目し 1 回目の中間解析データを固定データ（IAfix）と呼ぶ．以降，1 回目の中間解析，すなわち固定データに対する条件付き検出力について議論する．各被験者データは互いに独立とし，

固定データのサンプルサイズを各群n_1，2 回目の中間解析までに得られるデータのサンプルサイズを各群n_2，全解析データのサンプルサイズを各群N，θを効果量とする．このとき 1 回目の中間解析から 2 回目の中間解析にかけて得られるデータのサンプルサイズと 2 回目の中間解析から最終解析である 3 回目の解析にかけて得られるデータのサンプルサイズはそれぞれ各群，n_2-n_1，$N-n_2$と表される．1 回目の中間解析において得られる実現値$X_i \sim N(\mu,1), Y_i \sim N(0,1)$をそれぞれ治療群の$i$番目の被験者データ$(i=1,2,\ldots,n_1)$，対象群の$i$番目の被験者データとする．また 1 回目の中間解析から最終解析にかけて得られる実現値$X_j \sim N(\theta,1), Y_j \sim N(0,1)$をそれぞれ治療群の$j$番目の被験者データ，対象群の$j$番目の被験者データ$(j=n_1+1, n_1+2,\ldots,N)$とする． n_1の値は設定し，Nの値は固定データの標本標準偏差，有意水準α，検出力$1-\beta$，対立仮説のもとでの効果量θ_{alt}を与えることで決定される．n_2の値は，$n_2=(n_1+N)/2$として算出した．すなわち 2 回目の中間解析は，1 回目の中間解析と最終解析の中央の時点で行うものとした．α, βはそれぞれ 0.025，0.1 とし，θ_{alt}は 0.1，0.15，0.2 と変化させた．

3.2.条件付き検出力の評価と真値の算出方法

発生させた固定データに対する条件付き検出力の真値CP*を次の式によって求めた．

$$\mathrm{CP}^*(\theta)=\frac{\sum_{s=1}^{s=S} I(Z_2^s>b_2, Z_3^s>a_3|Z_1)}{S} \tag{7}$$

ただし，$Z_k^s(k=2,3)$は第sシミュレーションの第k中間（最終）解析における2標本Z検定の検定統計量を表し，次の式によって算出した．

$$Z_2^s=\sqrt{\frac{n_2}{2}}\left(\frac{\sum_{i=1}^{i=n_1}X_i+\sum_{j=n_1+1}^{j=n_2}X_j}{n_2}-\frac{\sum_{i=1}^{i=n_1}Y_i+\sum_{j=n_1+1}^{j=n_2}Y_j}{n_2}\right)$$

$$Z_3^s=\sqrt{\frac{N}{2}}\left(\frac{\sum_{i=1}^{i=n_1}X_i+\sum_{j=n_1+1}^{j=N}X_j}{N}-\frac{\sum_{i=1}^{i=n_1}Y_i+\sum_{j=n_1+1}^{j=N}Y_j}{N}\right)$$

また，b_2は 2 回目の中間解析における無効中止境界値，a_3は最終解析における棄却限界値，Sはシミュレーション回数，Iは指示関数を表す．シミュレーション回数Sは 10 万回とした．ここで，X_i及びY_iはシミュレーションごとに同一である一方，X_j及びY_jはシミュレーションごとに異なることに注意されたい．(7) 式の分子は，2 回目の中間解析における検定統計量が無効中止境界値を上回り継続判断が下され，最終解析において帰無仮説が棄却されて有意になる回数を表している．すなわち(7) 式は，同一の固定データが得られた状況を条件としたシミュレーションのうち，最終解析において帰無仮説が棄却されて有意になる場合の割合と解釈できる．

真値を求める SAS コードを次に示す．

```
%macro simulate_IA1;
    %do diff3 = 0 %to 20; /*各効果量に対してシミュレーション*/
/*固定データの生成*/
data IAfix;/*１回目の固定中間解析データ*/
call streaminit(00000);
n1 = 500;/*1回目の中間解析データサンプルサイズ*/ diff1 = 0;/*固定データの平均パラメータ*/
mu10 = 0; mu11 = mu10 + diff1;sd = 1;
    do i = 1 to n1;Ldl = rand('NORMAL',mu11,sd);Trt = 1;output;end;
    do j = 1 to n1;Ldl = rand('NORMAL',mu10,sd);Trt = 0;output;end;
drop mu10 mu11 diff1 sd i j;run;
proc reg data=IAfix; model Ldl=Trt;ods output ParameterEstimates=Parms_IA1;run;quit;
data Parms_IA1; set Parms_IA1;if Variable='Trt';_Scale_='MLE';_Stage_= 1;
   keep _Scale_ _Stage_ Variable Estimate StdErr;run;
```

```
data IA; set Parms_IA1;
/*シミュレーション*/
Nsim = 100000;/*シミュレーション回数*/n1 = 500;/*１回目のデータ数*/
n2 = 731;/*２回目までの総データ数*/n3 = 962;/*３回目までの総データ数*/
zalpha = 1.89868;/*分位点*/beta2 = 1.26741;/*２回目の中間解析における無効中止境界*/
effect = 1;/*効果量の刻み幅*/
do j = 1 to Nsim;
    sum1 = 0;sum2 = 0;diff2 = effect * &diff3 / 100;
        mu20 = 0; mu21 = mu20 + diff2;sd = 1;sign = 0;
        do i = n1 to n2;
                sum1 = sum1 + rand("NORMAL",mu21,sd);/*Trt=1の中間解析データの合計*/
                sum2 = sum2 + rand("NORMAL",mu20,sd);/*Trt=0の中間解析データの合計*/
        end;
        delta = (sum1 - sum2)/(n2-n1);/*全データの平均の差*/
        mean = (n1*Estimate + (n2-n1)*delta)/n2;/*差の加重平均*/
        z = mean*sqrt(n2)/sqrt(2);
        if z > beta2 then do;
            do k = n2 to n3;
                    sum1 = sum1 + rand("NORMAL",mu21,sd);/*Trt=1の中間解析データの合計*/
                    sum2 = sum2 + rand("NORMAL",mu20,sd);/*Trt=0の中間解析データの合計*/
            end;
            delta = (sum1 - sum2)/(n3-n1);/*全データの平均の差*/
            mean = (n1*Estimate + (n3-n1)*delta)/n3;/*差の加重平均*/
            z = mean*sqrt(n3)/sqrt(2);
        if z > zalpha then sign=1;
        end;output;
end;
keep mean z sign;run;
proc means data=IA ;/*signの平均が条件付き検出力の真値*/var sign;output;run;
%end;%mend simulate_IA1;
%simulate_IA1;; /*真値算出のためのSASマクロ実行*/
```

3.3.実験方法

実験方法について具体的に説明する．次の図 1 は，実験手順を表したフローチャートである．

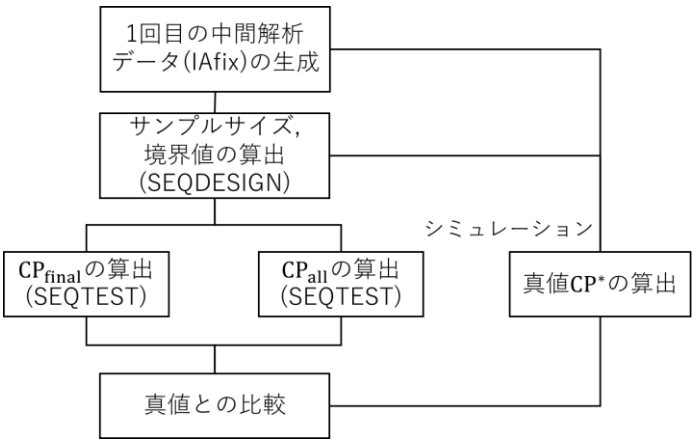


図 1 実験手順

はじめに，1 回目の中間解析データである固定データを生成する．次に，各中間解析時点におけるサンプルサイズや境界値を SEQDESIGN プロシジャによって算出する．特に，2 回目の中間解析における無効中止境界は，CP_{all}およびCP^*の算出に必要である．その後，前章で示した SAS コードによりCP_{final}とCP_{all}を算出する．また，固定データに対して，シミュレーションを行い，真値CP^*を算出する．そして，求めた真値CP^*とCP_{final}，CP_{all}を比較する

3.4. 結果

条件付き検出力と，効果量の関係に関する結果の一部を示す．次の図 2，3 は，固定データの母平均$\mu = 0$として生成した，平均が 0.006，標準偏差が 0.081，サンプルサイズが各群 300 の固定データに対して，数値実験を行った結果である．この固定データを IAfix_A と呼ぶ．図の縦軸は条件付き検出力，横軸は効果量を表す．

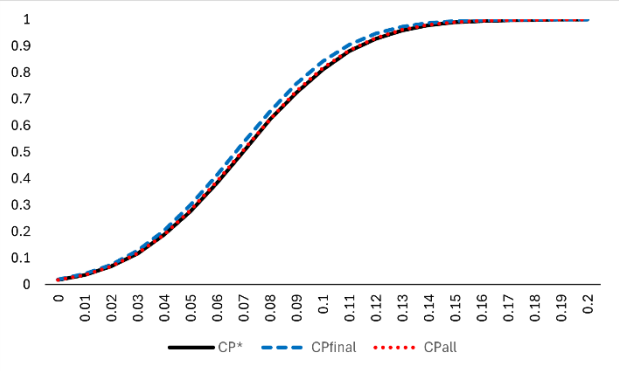


図 2 $\theta_{\mathrm{alt}} = 0.1$

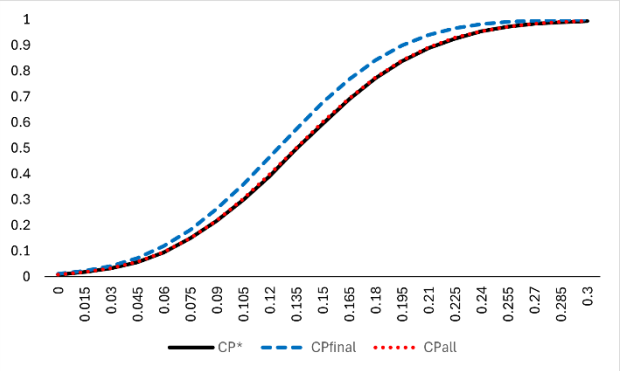


図 3 $\theta_{\mathrm{alt}} = 0.15$

次の図 4，5 は，固定データの母平均$\mu = 0$として生成した，平均が 0.176，標準偏差が 0.079，サンプルサイズが各群 300 の固定データに対する結果である．この固定データを IAfix_B と呼ぶ．

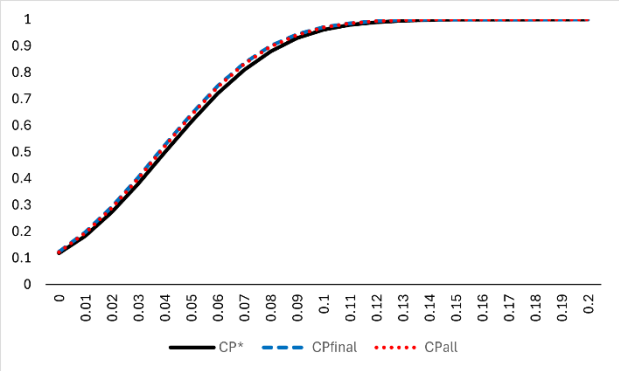


図 4 $\theta_{\mathrm{alt}} = 0.1$

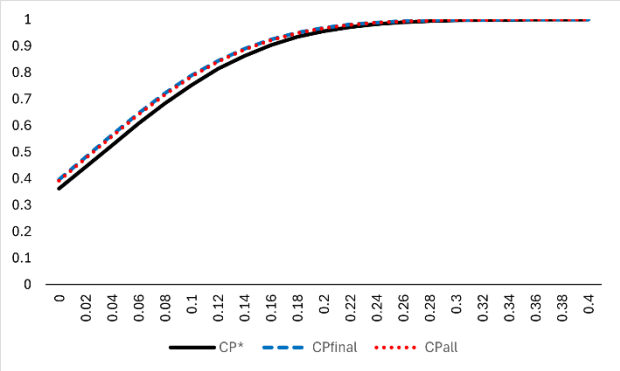


図 5 $\theta_{\mathrm{alt}} = 0.2$

次の図 6，7 は，固定データの母平均$\mu = -0.3$として生成した，平均が−0.115，標準偏差が 0.133，サンプルサイズが各群 100 の固定データに対する結果である．この固定データを IAfix_C と呼ぶ．

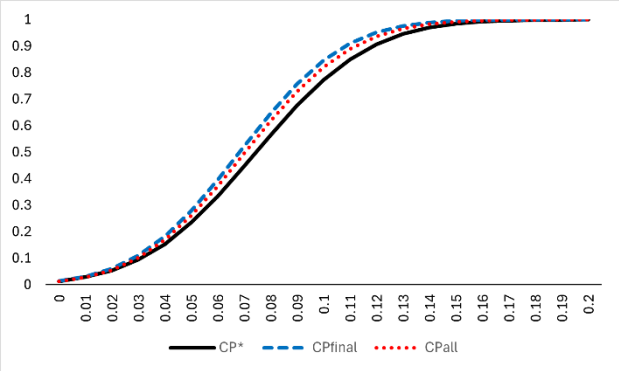


図 6 $\theta_{\mathrm{alt}} = 0.1$

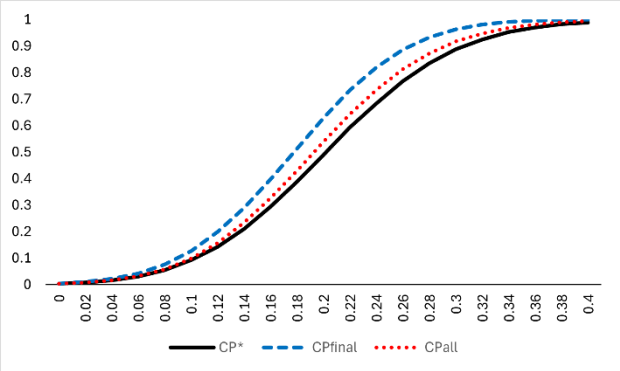


図 7 $\theta_{\mathrm{alt}} = 0.2$

次の図 8，9 は，固定データの母平均$\mu = 0$として生成した，平均が−0.085，標準偏差が 0.150，サンプルサイズが各群 100 の固定データに対する結果である．この固定データを IAfix_D と呼ぶ．

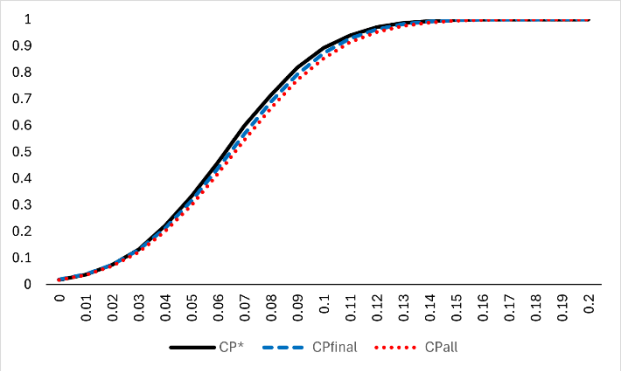


図 8 $\theta_{\mathrm{alt}} = 0.1$

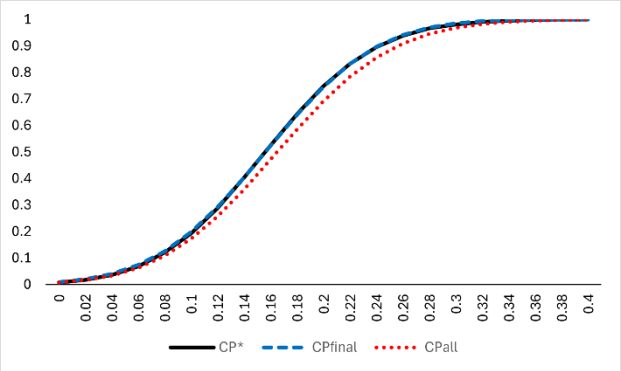


図 9 $\theta_{\mathrm{alt}} = 0.2$

次の図 10，11 は，固定データの母平均$\mu = 0.1$ として生成した，平均が 0.069，標準偏差が 0.063，サンプルサイズが各群 500 の固定データに対する結果である．この固定データを IAfix_E と呼ぶ．

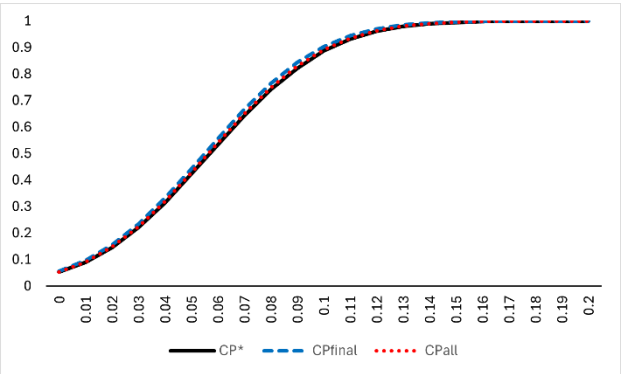


図 10 $\theta_{\mathrm{alt}} = 0.1$

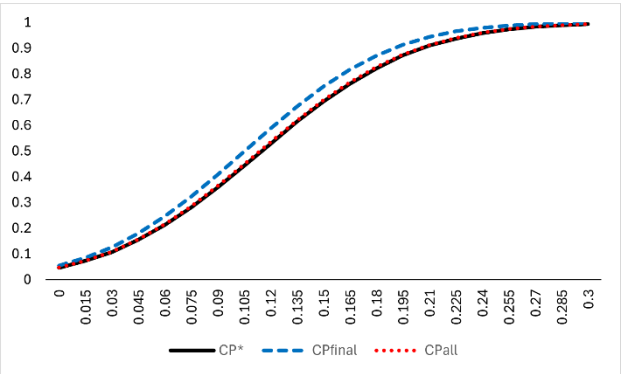


図 11 $\theta_{\mathrm{alt}} = 0.15$

次の図 12，13 は，固定データの母平均$\mu = 0$として生成した，平均が 0.073，標準偏差が 0.063，サンプルサイズが各群 500 の固定データに対する結果である．この固定データを IAfix_F と呼ぶ．

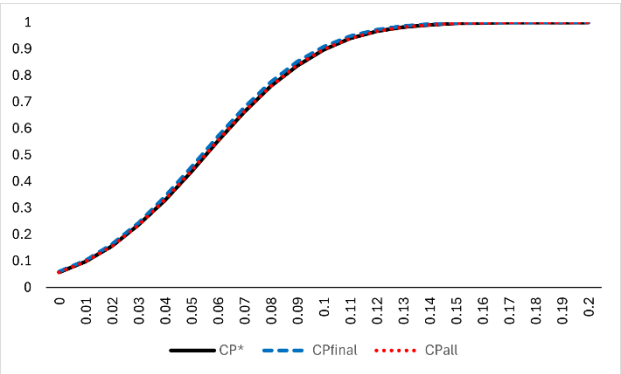


図 12 $\theta_{\mathrm{alt}} = 0.1$

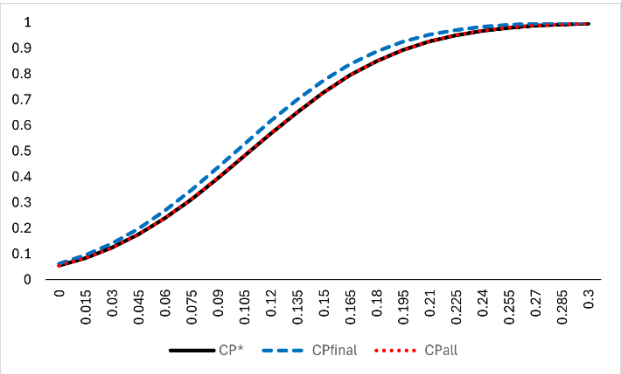


図 13 $\theta_{\mathrm{alt}} = 0.15$

3.5. 考察

数値実験の結果，固定データとして IAfix_D を用いた，図 8，図 9 を除いたすべての状況において$\mathrm{CP}_{\mathrm{final}}$よりも$\mathrm{CP}_{\mathrm{all}}$の方が真値に近い値を算出することが確認された．逆に，図 8，図 9 の状況では，$\mathrm{CP}_{\mathrm{all}}$よりも$\mathrm{CP}_{\mathrm{final}}$の方が真値に近い値を算出することが確認された．

固定データの違いによって，より真値に近い条件付き検出力の算出を達成する方法が変化するという結果に関して，標準偏差の誤推定が影響していると考える．数値実験では，治療群と対照群に関して，分散1の正規分布を仮定したため，Z統計量の標準偏差の真値は，固定データのサンプルサイズn_1を用いて，$\sigma =$

$\sqrt{2/n_1}$と求められる．一方で，実際に中間解析を実施する際は，標準偏差を中間解析データから推定する必要がある．標準偏差の真値σと固定データから推定した標準偏差$\hat{\sigma}$の差を次の表 1 に示す．

表 1 標準偏差の真値$\boldsymbol{\sigma}$と固定データから推定した標準偏差$\hat{\boldsymbol{\sigma}}$の差

固定データ	IAfix_A	IAfix_B	IAfix_C	IAfix_D	IAfix_E	IAfix_F
真値$\boldsymbol{\sigma}$	0.082	0.082	0.141	0.141	0.063	0.063
推定値$\hat{\boldsymbol{\sigma}}$	0.081	0.079	0.133	0.150	0.063	0.063
真値と推定値の差$\boldsymbol{\sigma}-\hat{\boldsymbol{\sigma}}$	0.001	0.003	0.008	-0.009	0.000	0.000

表 1 と数値実験の結果から，真値と推定値の差$\sigma-\hat{\sigma}$が比較的大きい固定データである IAfix_C，IAfix_D を用いた状況では，$\mathrm{CP}_{\mathrm{all}}$の真値からの乖離が大きくなる傾向が確認された．逆に，真値と推定値の差$\sigma-\hat{\sigma}$が小さい固定データである IAfix_B， IAfix_E，IAfix_F を用いた状況では，$\mathrm{CP}_{\mathrm{all}}$の曲線が$\mathrm{CP}^*$の曲線に重なっており，$\mathrm{CP}_{\mathrm{all}}$の真値からの乖離が小さくなる傾向が確認された．実際，IAfix_C のように標準偏差を過小推定した場合，当該固定データにおけるZ統計量は過大評価され，それに伴い条件付き検出力も真値と比較してインフレーションを起こすと考えられる．この場合，$\mathrm{CP}_{\mathrm{final}}$，$\mathrm{CP}_{\mathrm{all}}$いずれの算出方法でも条件付き検出力の真値を上回る値を算出することになるが，$\mathrm{CP}_{\mathrm{all}}$は中止確率を考慮する分，$\mathrm{CP}_{\mathrm{final}}$よりも真値に近い値を算出することができる．一方で，IAfix_D のように標準偏差を過大推定した場合，当該固定データにおけるZ統計量は過小評価され，それに伴い条件付き検出力も真値と比較してデフレーションを起こすと考えられる．この場合，$\mathrm{CP}_{\mathrm{final}}$，$\mathrm{CP}_{\mathrm{all}}$いずれの算出方法でも条件付き検出力の真値を下回る値を算出することになるが，$\mathrm{CP}_{\mathrm{final}}$は中止確率を考慮しない分，$\mathrm{CP}_{\mathrm{all}}$よりも真値に近い値を算出することができる．

標準偏差の誤推定を修正することにより，どの程度条件付き検出力の算出精度が改善するか確認した結果を示す．IAfix_C，IAfix_D に対し，固定データから推定した標準偏差ではなく，サンプルサイズから求めた標準偏差の真値を用いてシミュレーションを行った．次の図 14 は，固定データ IAfix_C に対して，真の標準偏差 0.141 を用いた結果である．図 15 は，図 7 と同じ，推定された標準偏差 0.133 を用いた結果である．

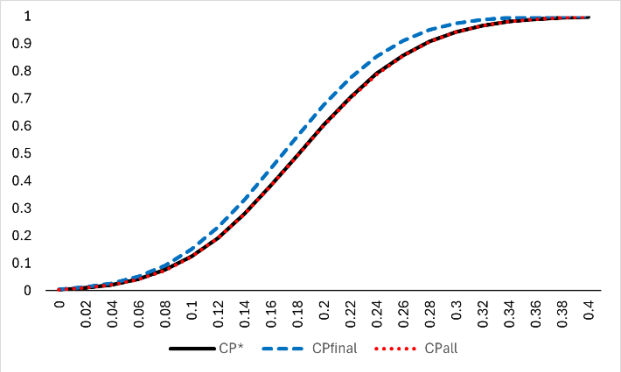


図 14 $\hat{\sigma}=\sigma=0.141, \theta_{\mathrm{alt}}=0.2$

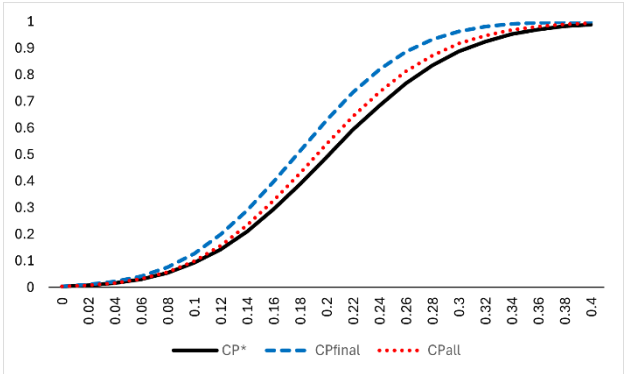


図 15 $\hat{\sigma}=0.133, \theta_{\mathrm{alt}}=0.2$

次の図 16 は，固定データ IAfix_D に対して，推定された標準偏差 0.150 ではなく，真の標準偏差 0.141 を用いた結果である．図 17 は，図 9 と同じ，推定された標準偏差 0.150 を用いた結果である．

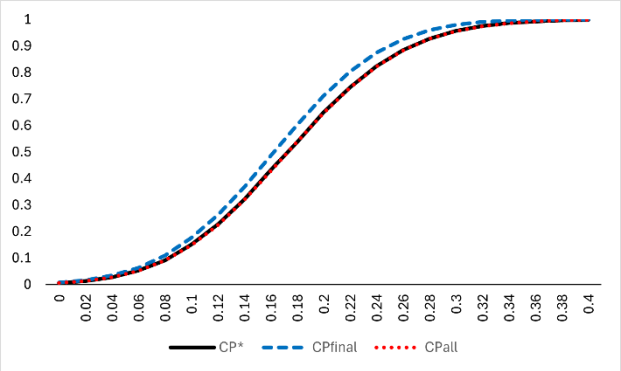


図 16 $\hat{\sigma} = \sigma = 0.141, \theta_{\text{alt}} = 0.2$

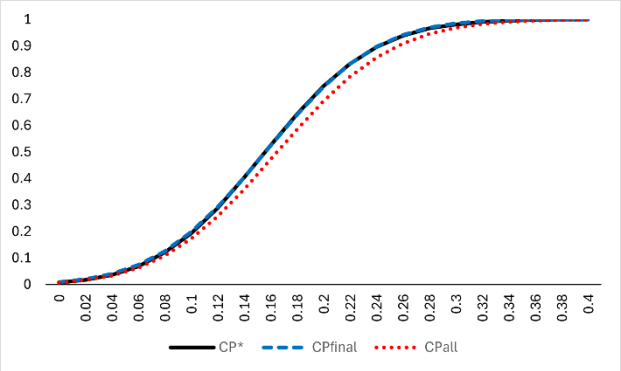


図 17 $\hat{\sigma} = 0.150, \theta_{\text{alt}} = 0.2$

図 14 と図 15，図 16 と図 17 いずれにおいても，標準偏差の値を固定データからの推定値から真値に変更することで，CP_{all}が真値に近い条件付き検出力を算出できることが確認された．

次に，標準偏差の誤推定と優れている算出方法の関係をシミュレーションによって確認した結果を示す．シミュレーションでは，固定データから標準偏差を推定することを 1 万回繰り返した．次の図 18 と図 19 は，それぞれ固定データのサンプルサイズ$n_1 = 100, n_1 = 500$の場合について，標準偏差の推定値の分布と優れている算出方法の関係を示した結果である．図中の点線は，標準偏差の真値を表す．

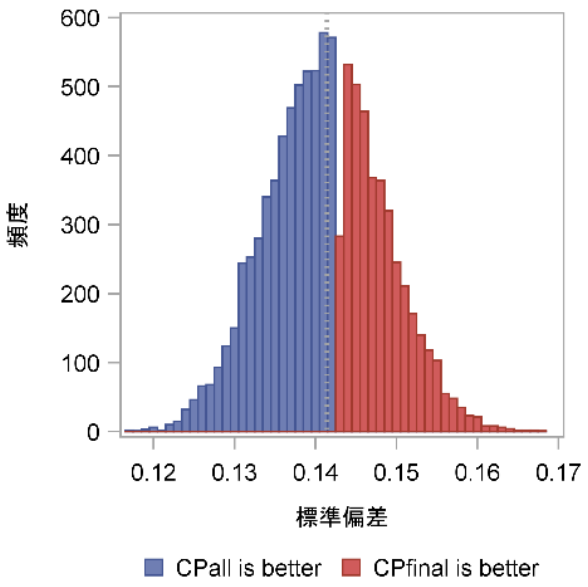


図 18 $n_1 = 100$ $\hat{\sigma}$の分布と CP の関係

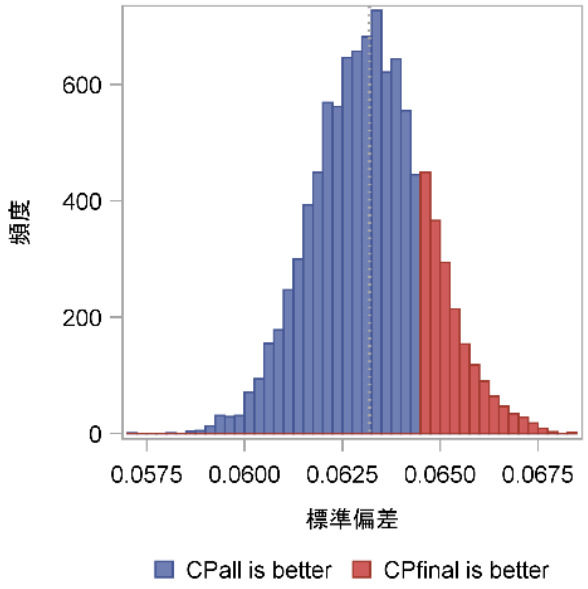


図 19 $n_1 = 500$ $\hat{\sigma}$の分布と CP の関係

図 18 の$n_1 = 100$の場合では，CP_{all}の方が真値に近い確率が 59.6%，CP_{final}の方が真値に近い確率が 40.4%であった．また，図 19 の$n_1 = 500$の場合では，CP_{all}の方が真値に近い確率が 81.1%，CP_{final}の方が真値に近い確率が 18.9%であった．

4. まとめ

本稿では，条件付き検出力に着目し，各中間解析時点における中止確率を考慮しない算出方法CP_{final}と考慮する算出方法CP_{all}の近似式及び SAS 上での算出方法を示した．数値実験の結果から，無効中止を目的とした中間解析では，常にCP_{all}よりもCP_{final}の方が値は大きいことが確認された．これは，CP_{final}では当該中間解析以降の中間解析による無効中止を考慮せず，無条件に最終解析での棄却確率を算出する一方，CP_{all}では

途中の中間解析による無効中止を考慮するため，最終解析まで試験が継続されるという条件の下，最終解析での棄却確率を算出するからである．加えて，$\mathrm{CP}_{\mathrm{final}}$と$\mathrm{CP}_{\mathrm{all}}$の条件付き検出力の算出精度には，標準偏差の誤推定が影響していることが示唆された．特に，標準偏差を過大推定した場合，当該固定データにおけるZ統計量は過小評価され，結果的に$\mathrm{CP}_{\mathrm{final}}$の方が優れた算出精度を示す状況が存在することが確認された．また，標準偏差の誤推定と真値に近い算出方法の関係には，当該中間解析データのサンプルサイズが影響しており，サンプルサイズが大きくなるほど，$\mathrm{CP}_{\mathrm{all}}$の方が優れた算出精度を示す確率が高くなる傾向が確認された．

本稿が SAS で条件付き検出力を用いた解析をする際の参考になると幸いである．

参考文献

Jennison, C., Turnbull, B. W. Group Sequential Methods with Applications to Clinical Trials, 2000. Chapman and Hall. (森川敏彦，山中竹春 (訳). (2012). 臨床試験における群逐次法：理論と応用．シーエーシー).

Kim, K. M., Tsiatis, A. A., Mehta, R. C. Computational Issues In Information-Based Group Sequential Clinical Trials. Journal of Japanese Society of Computational Statistics, 2003, 15, 153–167.

Lan, K. K. G., Simon, R., Halperin, M. Stochastically curtailed tests in long–term clinical trials. Commun Statist, 1982, 1, 207–219.

O'Brien, C. P., Fleming, R. T. A multiple testing procedure for clinical trials. Biometrics, 1979, 35, 549–556.

Pocock, S. J. Group sequential method in the design and analysis of clinical trials. Biometrika, 1977, 64, 191–199.

Yeatts, S. D., Martin, H. R., Coffey, S. C. et al. Challenges of decision making regarding futility in a randomized trial. Stroke, 2014, 45, 1408–1414.

Zhu, L., Ni, L., Yao, B. Group Sequential Methods and Software Applications. The American Statistician, 2011, 65, 127–135.

小山暢之．中間解析結果に依存した解析時点の選択と α 消費関数法．計量生物学, 2001, 22, 53–65.

髙沢翔，堂園剛司，冨金原悟，森川敏彦．中間解析結果に基づく被験者数再設定方法の特性の検討． SAS ユーザー総会論文集, 2011, 121–133.

丹後俊郎．臨床試験のデザイン．保険医療科学, 2011, Vol. 60, 27–32.

浜田知久馬. SAS による中間解析のデザインと解析. SAS ユーザー総会論文集, 2010, 111‐148.

付録

```
proc seqdesign altref=0.1 ;
   SidedOBrienFleming: design nstages=3 alpha=0.05 beta=0.05
                       stop = accept/*中止条件*/
                       method=obf /*棄却域obf型*/ alt=upper;
   ods output Boundary=Bnd_LDL;run;  /*無効中止境界の取得*/
run;
```

上記の SEQDESIGN プロシジャでは，対立仮説のもとでの効果量θ_{alt}=0.1，ステージ数 3，αエラー5%，βエラー5%，中止条件を無効中止のみ，β消費関数を O'Brien Fleming 型，上側対立仮説を考える中間解析デザインの情報を生成するものとなっている．中止条件を決定する stop=では無効中止のみを考慮する accept 以外に，有効中止のみを考慮する reject と無効中止と有効中止いずれも考慮する both をオプションとして指定することができる．また，β消費関数を表す method=では，次の図 20 のような O'Brien Fleming 型を用いる obf 以外に，図 21 のような Pocock 型を用いる poc をオプションとして指定することができる．

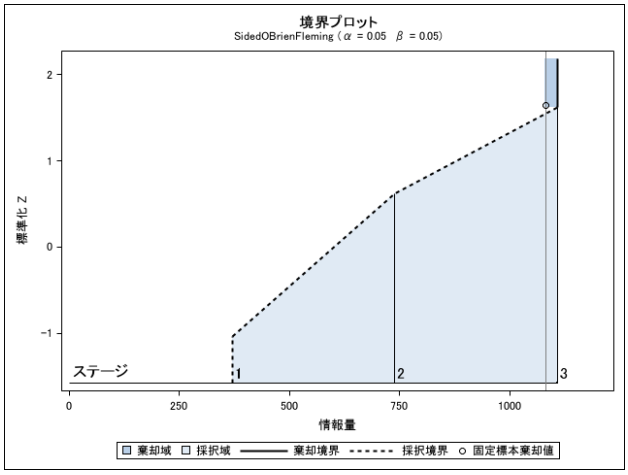


図 20 obf 型の境界プロット例

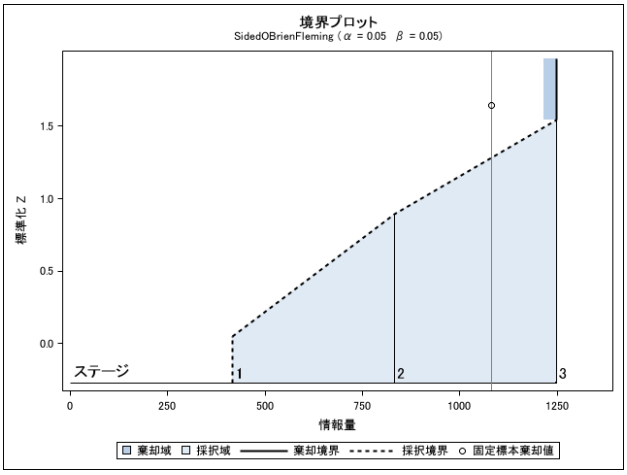


図 21 poc 型の境界プロット例

本稿では無効中止に注目したため，β消費関数を指定したが，同様にしてα消費関数についても指定することができる．加えて，alt=では上側対立仮説を考える upper 以外に，下側対立仮説を考える lower，両側対立仮説を考える two sided をオプションとして指定することができる．加えて，ここでは指定していないオプションとして、boundaryscale=が挙げられる．boundaryscale=では出力される境界値の種類を指定することができる．具体的には，図 22 のような平均値に基づく棄却域を算出する mle，図 23 のようなスコア統計量に基づく棄却域を算出する score，図 24 のような z 統計量に基づく棄却域を算出する stdz，図 25 のような p 値に基づく棄却域を算出する pvalue の中から選択することができる．境界域の種類として pvalue を指定すると，他の境界域の種類とは採択域と棄却域の関係が上下逆転するため，図の解釈には注意が必要である．

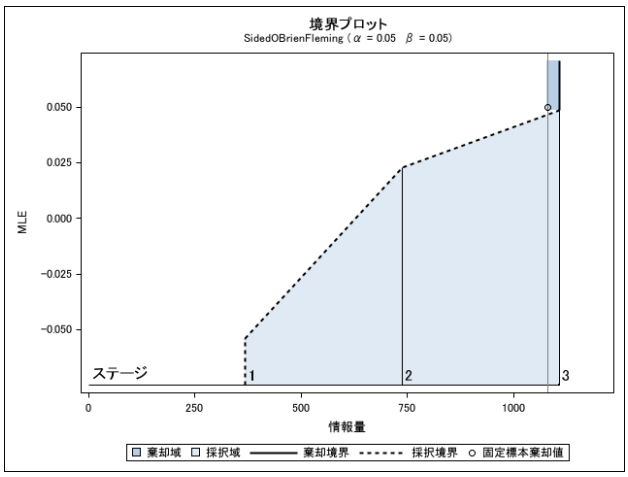


図 22 平均値に基づく境界プロット例(mle)

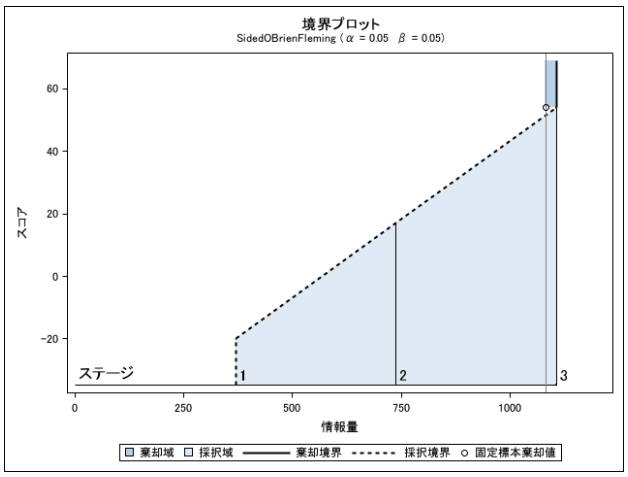


図 23 スコア統計量に基づく境界プロット例(score)

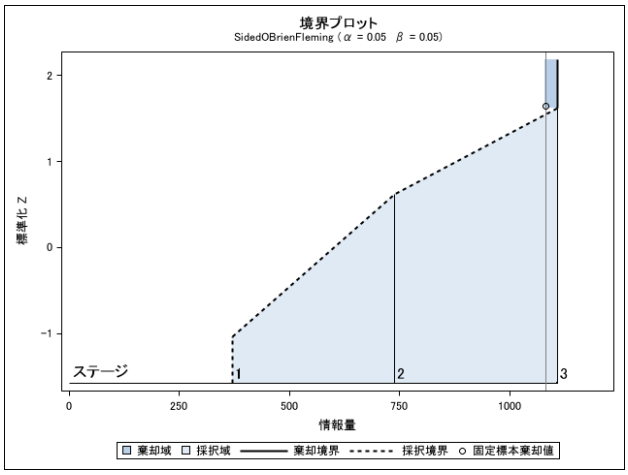


図 24 z 統計量に基づく境界プロット例(stdz)

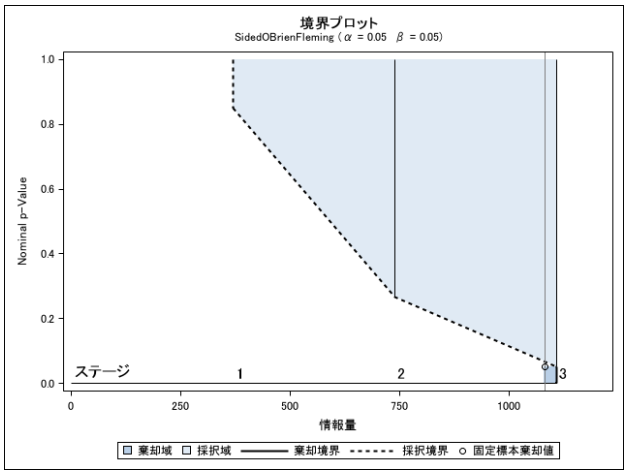


図 25 p 値に基づく境界プロット例(pvalue)

```
/*CP_finalの算出*/
proc seqtest Boundary = Bnd_LDL
Parms(Testvar = Trt) = Parms_IA1
infoadj=prop
    condpower(cref=0 to 1 by 0.1 TYPE=FINALSTAGE) plots=condpower;
ods output Test=Test_LDL1_FINAL;run;
/*CP_allの算出*/
proc seqtest Boundary = Bnd_LDL
 Parms(Testvar = Trt)=Parms_IA1
 infoadj=prop
    condpower(cref=0 to 1 by 0.1 TYPE=ALLSTAGES) plots=condpower;
ods output Test=Test_LDL1_ALL;run;
```

上記の SEQTEST プロシジャでは，中間解析データにおける中間解析情報を算出する．入力として SEQDESIGN プロシジャで算出した境界値 Bnd_ldl とシミュレーションコード内で生成した被験者データを加工した Parm_IA1 を用いている．Parm_IA1 はシミュレーションコード内の次の部分で加工している．

```
proc reg data=IAfix; model z=Trt; ods output ParameterEstimates=Parms_IA1; run;
data Parms_IA1; set Parms_IA1; if Variable='Trt'; _Scale_='MLE'; _Stage_= 1;
   keep _Scale_ _Stage_ Variable Estimate StdErr;
run;
```

このコードを実行すると，次のようなデータセットを得ることができる．

	変数	パラメータ推定値	標準誤差	_Scale_	_Stage_
1	Trt	-0.04932	0.13124	MLE	1

すなわちこのデータセットが中間解析で解析したいデータセットとなる．本文の 3.5 節で議論した標準偏差の推定値は，上記のデータセットにおける標準誤差に相当する．

SEQTEST プロシジャを実行すると出力として，中間解析における情報量，境界と検定統計量の関係，中間解析による中止・継続判断，効果量ごとの条件付き検出力等を得ることができる．本文で述べたように，条件付き検出力は condpower オプションによって得ることができる．デフォルトでは TYPE=ALLSTAGES が指定されており，CP_{all}が算出される．したがって SAS においては，サブオプションによって指定しなければ，これまでよく用いられてきた条件付き検出力CP_{final}ではなく，CP_{all}が算出されることに注意したい．加えて上記の SEQTEST プロシジャでは，plot=condpower を指定することにとり，横軸に効果量，縦軸に条件付き検出力の値をとった条件付き検出力プロットを得ることができる．

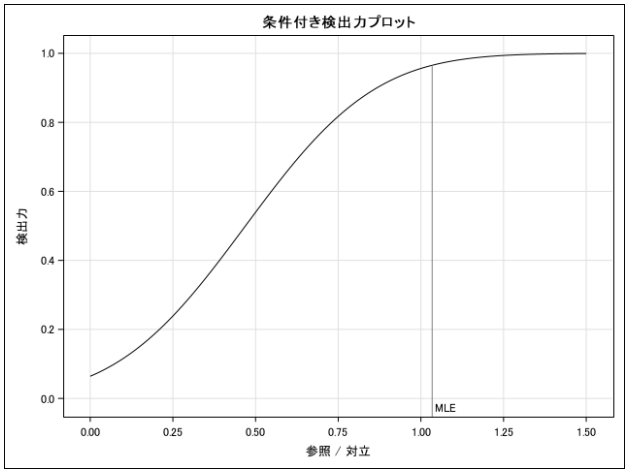


図 26　SAS による条件付き検出力プロットの出力例

トレンドの変化点を推定可能な区分線形回帰の SAS による実装と応用事例のご紹介

○ビービ　彩加[1]、菊池　宏和[1]

([1]シミック（株）データサイエンス事業本部）

Techniques for Estimating Changes in Trends: Piecewise Linear Regression and Its Implementation in SAS

Ayaka Beebe, Hirokazu Kikuchi
Data Science Div., CMIC Co., Ltd.

要旨

途中で傾向が変わる線形回帰を求めることができる、区分線形回帰を紹介する。また、例を用いて SAS の PROC NLIN の使用方法を理解する。
キーワード：区分線形回帰モデル、複数の区分、変化点、PROC NLIN

1　はじめに

実データでは、データの中に区分ごとの傾きの違いが存在する場合、説明変数の範囲ごとに異なる関係性が現れることもある。このような状況では、従来の単一の線形回帰モデルを用いるだけでは変数間の関係の変化を十分に捉えることができず、推定結果に偏りが生じるリスクがある。区分線形回帰モデルはデータを複数の区間に分け、それぞれの区間で異なる線形関係を容認することによって、このような状態でも適切に対応できる臨機応変なモデル化手法である。

区分線形回帰モデルの推定では、各区間の回帰係数だけでなく、区切りとなる変化点の位置も同時に推定することが可能である。SAS では、PROC NLIN を用いることで独自のモデル式を柔軟に定義できるため、回帰係数と変化点を同時に推定する区分線形回帰モデルを非線形回帰として実装することが可能である。

本論文では、区分線形回帰モデルの基本的な理論と SAS での PROC NLIN による実装方法について、概要を説明する。SAS HELP のデータを用いた例を通じて、本手法の有用性や注意点について解説し、具体的な応用方法も示す。

2　区分線形回帰モデル

2.1　区分線形回帰

区分線形回帰（piecewise linear regression）は、説明変数の値域を複数区間に分割し、各区間で独立に線形回帰式を適用することにより、変化点（ブレークポイント）を境に関係性が変化するデータをモデルで表現する統計的手法である。特に、変化点の位置が未知の場合は、モデルが非線形となるため、パラメータ推定には非線形最小二乗法（nonlinear least squares method）が用いられることが多い。本手法を用いることで、従来の単一の線形回帰では対応できない傾向の変化を読み取ることが可能となり、区間ごとに異なる傾向を柔軟にモデリングできる点が特徴である。

2.2　非線形最小二乗法

区分線形回帰において変化点が未知の場合、回帰モデルは線形部分と変化点自体の両方をパラメータとして含むため、全体として非線形な推定問題となる。このとき適用される非線形最小二乗法は、観測値とモデルによる推定値の差（残差）の二乗和を最小化するパラメータ値を求める手法である。具体的には、初期推定値を設定した上で、パラメータ（変化点の位置および各区間の回帰係数）を段階的に調整し、残差平方和が最も小さくなるように最適化を行うアルゴリズムが用いられる。

区分線形回帰を使用し、変化点(t)が未知の場合、モデルは次のように表される。

$$y_i = \begin{cases} \beta_0 + \beta_1 x_i, & x_i \le t \\ \beta_0 + \beta_1 t + \beta_2 (x_i - t), & x_i > t \end{cases} \qquad (1)$$

このとき、$(\beta_0, \beta_1, \beta_2)$は各区間の回帰係数、$t$は変化点の位置であり、説明変数を$x_i$、応答変数$y_i$と設定する。変化点$(t)$がモデル式の中に含まれ、区分ごとに線形性が異なるため、モデルはパラメータに対して非線形となり、通常の線形最小二乗法では推定ができない。

非線形最小二乗法は、次の目的関数を最小化するパラメータ$(\beta_0, \beta_1, \beta_2, t)$を探索する。

$$S(\beta_0, \beta_1, \beta_2, t) = \sum_{i=1}^{n} [y_i - f(x_i; \beta_0, \beta_1, \beta_2, t)]^2 \qquad (2)$$

ここで、$f(x_i; \beta_0, \beta_1, \beta_2, t)$は式(1)の区分線形回帰モデルを表している。この方法により、観測値(y_i)とモデルによる推定値との差（残差）の二乗和を最小にするパラメータセットが求められ、変化点と各区間の回帰係数を同時に推定することが可能となる。数式における iは、データの各観測点（各サンプル）を表すインデックス（番号）である。たとえば、データが N件あれば、i は 1 からN までの値を取り、上記の数式では、$i = 1$ から N までの各データについて、「観測値 y_i」と「モデルから予測した値 $f(x_i; \beta_0, \beta_1, \beta_2, t)$」の差を二乗し、それらを全て足し合わせることで推定できる。

3　SAS の実装について

3.1　解析データセット

本分析で使用するデータセットは SAS Help ライブラリに記載されている LIDAR データセット[*1] である。本データセットは、平滑化手法や回帰モデルの例として、多くの書籍や論文で利用されている。LIDAR（Light Detection and Ranging）は、レーザーの反射を用いて大気中の化学物質を検出する技術であり、科学者によって広く用いられている。データセットの各変数の詳細を Table 1 に、データセットの値の中身を Table 2 に示す。

Table 1. LIDAR データセットの内容

変数名	内容
RANGE	光が反射して戻ってくるまでの距離
LOGRATIO	レーザーから受け取る光の量の比を対数にしたもの

Table 2. LIDAR データセット

RANGE	LOGRATIO
390	-0.05035573
391	-0.06009706
…	…
718	-0.557754
720	-0.8026684

3.2　RANGE と LOGRATIO の基本情報

本論文では RANGE を説明変数、LOGRATIO を応答変数として設定する。まず、LIDAR データセットの説明変数を均等に 3 つに分けたときの変化点（33 パーセンタイルと 66 パーセンタイル）を算出したものを Table 3 にまとめた。

Table 3. RANGE の 33 パーセンタイルと 66 パーセンタイル

変数	33rd Percentile	66th Percentile
RANGE	498	607

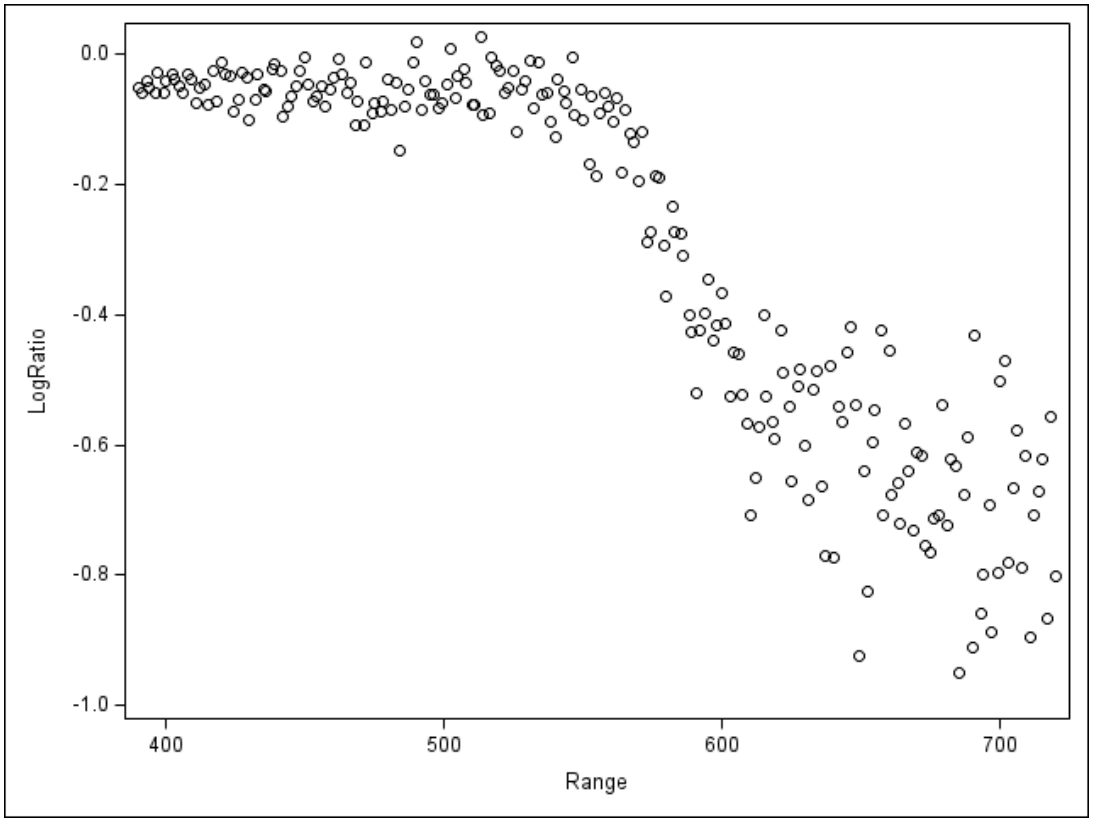
33 パーセンタイルと 66 パーセンタイルで RANGE のデータを分けた区分ごとに算出した LIDAR データセットの主要な統計量を Table 4 に記載した。

Table 4. LIDAR データセットの要約統計量

区分	変数	N	Mean	SD	Min	Q1	Median	Q3	Max
全体	RANGE	221	554.8	95.9	390.0	472.0	555.0	637.0	720.0
	LOGRATIO	221	-0.29	0.28	-0.95	-0.54	-0.11	-0.05	0.03
$RANGE \leq 498$	LOGRATIO	73	-0.05	0.03	-0.15	-0.07	-0.05	-0.03	0.02
$498 < RANGE \leq 607$	LOGRATIO	73	-0.17	0.16	-0.53	-0.28	-0.09	-0.06	0.03
$RANGE > 607$	LOGRATIO	75	-0.64	0.13	-0.95	-0.72	-0.63	-0.54	-0.40

Figure 1 は、RANGE（x 軸）と LOGRATIO（y 軸）の関係を図に示している。

Figure 1. RANGE と LOGRATIO の散布図



3.3 区分線形回帰モデルを用いる背景

RANGE と LOGRATIO の関係性を考察する。Figure 1 から、両者の関係は明らかに非線形であり、通常の線形回帰モデルでは十分に捉えられないことが示唆される。そのため、より柔軟なモデリング手法が必要となる。また、図中には 2 つの変化点が認められ、RANGE と LOGRATIO の関係は 3 つの区間で異なる傾向を示している。本論文では、データをこれらの区間ごとに分割し、各区間に線形モデルを適用する区分線形回帰に着目する。

3.4 区分線形回帰モデルの数式

2 つの変化点を持った区分線形回帰モデルは以下の数式で表すことができる。パラメータの内容は Table 5 にまとめた。第一区間のモデル式はbを傾き、aを切片とした一般的な一次関数の式$y = a + bx$を設定し、第二区間では最初の変化点(t_1)の影響を考慮したモデル式を構築した。第三区間に対しても、第二区間のモデル式に 2 つ目の変化点(t_2)の影響を配慮したモデル式を設定している。

$$y_i = \begin{cases} a_1 + b_1 * x_i & x_i \le t_1 \\ a_1 + b_1 * t_1 + b_2 * (x_i - t_1) & t_1 < x_i \le t_2 \\ a_1 + b_1 * t_1 + b_2 * (t_2 - t_1) + b_3 * (x_i - t_2) & x_i > t_2 \end{cases} \quad (3)$$

Table 5. 区分線形回帰モデルのパラメータ

パラメータ	内容
a_1	第一区間の切片
b_1	第一区間の傾き

b_2	第二区間の傾き
b_3	第三区間の傾き
t_1	第一変化点
t_2	第二変化点

定義した区分線形回帰モデルを SAS で実装するには PROC NLIN が適応可能である。

3.5 PROC NLIN とは

SAS の PROC NLIN は、非線形最小二乗法によりパラメータを推定するためのプロシジャであり、モデル式の自由な設定やパラメータの初期値の指定が可能である。区分線形回帰モデルにおいても、変化点の位置や回帰係数など複数のパラメータを一括して推定ができる。

本論文では、PROC NLIN を用いて区分線形回帰モデルの構築およびパラメータ推定を行う。

最適なパラメータを決定するプロセスは反復的に行われる。まず初期パラメータ値を設定する必要がある。初期パラメータ値は、データの要約統計量および散布図を参考に設定する。PROC NLIN は、パラメータの値を順次調整しながら、モデルの当てはまりを向上させる。このパラメータの調整の 1 回分を反復（イテレーション）と呼ぶ。次の反復でも同様に、パラメータを修正することでさらに当てはまりの改善を試みる。各反復で当てはまりの改善が得られなくなった時点で、モデルは収束したと判断される。

3.6 PROC NLIN を使った区分線形回帰モデルの実装

初期パラメータは Table 4 の要約統計量や Figure 1 の散布図を元に判断し、決定する。Table 6 に今回設定した初期パラメータの値をまとめた。

Table 6. 区分線形回帰モデルの初期パラメータ

パラメータ	内容	値
a_1	第一区間の切片	-0.1
b_1	第一区間の傾き	0.0001
b_2	第二区間の傾き	-0.001
b_3	第三区間の傾き	-0.0005
t_1	第一変化点	500
t_2	第二変化点	600

以下のコードを用いて SAS で PROC NLIN を実装する。また、SAS での実装は、SAS Version 9.4 を使用した。

```
proc nlin data=lidar converge=1e-3;
   /* 初期パラメータ値の設定 */
   parms a1=-0.1        /* 第一区間の切片 */
         b1=0.0001      /* 第一区間の傾き */
         b2=-0.001      /* 第二区間の傾き */
         b3=-0.0005     /* 第三区間の傾き */
         t1=500         /* 第一変化点 */
         t2=600;        /* 第二変化点 */

   /* 区分線形回帰モデルの定義 */
   if (range <= t1) then pred = a1 + b1*range;
   else if (range <= t2) then pred = a1 + b1*t1 + b2*(range-t1);
   else pred = a1 + b1*t1 + b2*(t2-t1) + b3*(range-t2);

   /* モデル式を設定 */
   model logratio = pred;

   /* 変化点に制約を設定 */
   bounds t1 > 390, t2 > t1, t2 < 720;

   /* 結果の出力 */
   output out=results p=predicted r=residual;
run;
```

まず、parms ステートメントにより Table 6 に示した初期パラメータ値を指定し、model ステートメントでは logratio と pred の関係でモデル式を構築した。pred の数式については、区間ごとに if else ステートメントを用いて応答変数と予測式を定義した。さらに、bounds ステートメントによって変化点のパラメータ範囲を設定した。

実行の過程で、パラメータの調整は計 11 回の反復を経て収束判定基準に達し、反復ごとの推定値を Table 7 に掲載した。

Table 7. Iteration Phase

Iteration	**a_1**	**b_1**	**b_2**	**b_3**	**t_1**	**t_2**	**Sum of Squares**
0	-0.1000	0.000100	-0.00100	-0.00050	500.0	600.0	18.6247
1	-0.0685	0.000030	-0.00173	-0.00086	526.0	685.0	11.7717
2	-0.0428	-0.00002	-0.00400	-0.00062	541.5	603.0	8.6361
3	-0.0270	-0.00006	-0.00712	-0.00180	555.2	624.7	1.5101
4	0.0134	-0.00015	-0.00794	-0.00169	554.5	615.5	1.3612

5	-0.00415	-0.00011	-0.00873	-0.00189	555.7	608.4	1.3385
6	0.0135	-0.00015	-0.00893	-0.00175	557.3	611.1	1.3325
7	0.0151	-0.00015	-0.00929	-0.00182	558.1	608.6	1.3302
8	0.0152	-0.00015	-0.00917	-0.00179	557.9	609.5	1.3295
9	0.0163	-0.00015	-0.00916	-0.00177	558.0	609.9	1.3294
10	0.0166	-0.00015	-0.00916	-0.00176	558.0	610.0	1.3294
11	0.0163	-0.00015	-0.00915	-0.00176	558.0	610.0	1.3294

反復回数が増加するにつれて、誤差項の平方和は減少し、モデルの適合度が高まる。モデルの適合性は Table 8 の分散分析表にて表すことができる。

Table 8. Analysis of Variance

Source	Degrees of Freedom	Sum of Squares	F value	p-value
Model	5	16.2248	524.80	<0.0001
Error	215	1.3294		
Corrected Total	220	17.5542		

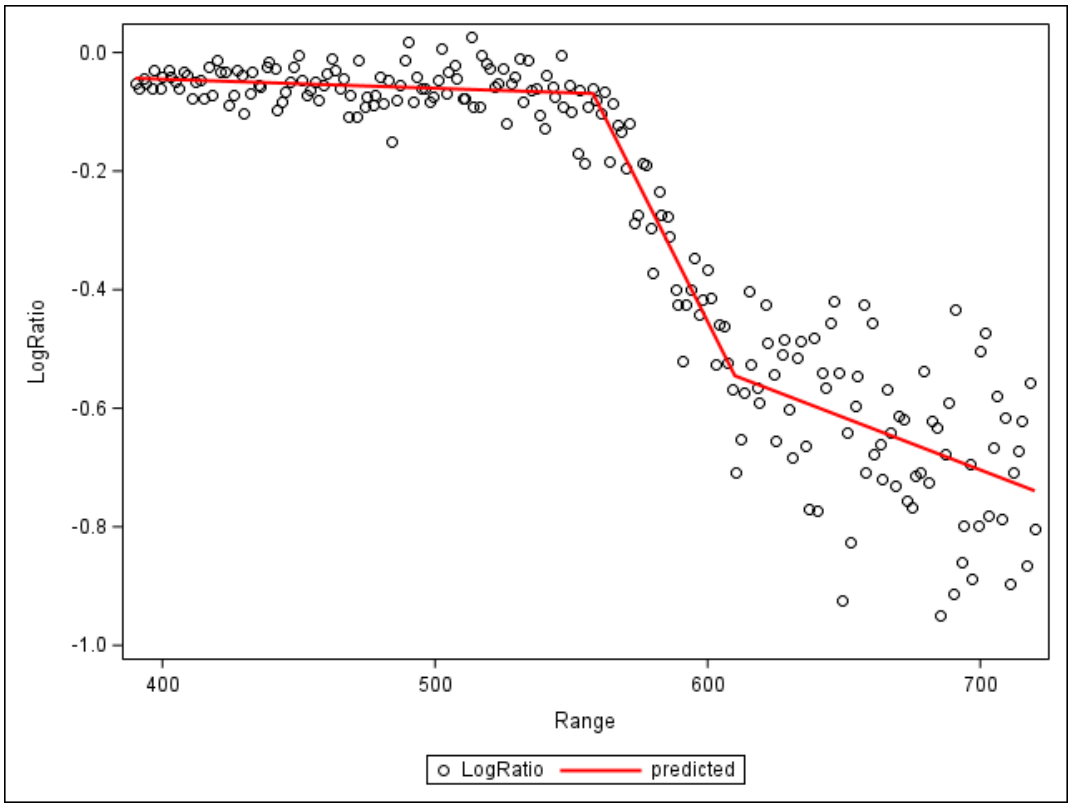
最終的に算出されたモデルのパラメータ推定値を Table 9 に示した。

Table 9. Parameter Estimates

Parameter Name	Parameter	Estimate	Standard Error	95% Confidence Limits	
第一区間の切片	a_1	0.0163	0.0729	-0.1273	0.1599
第一区間の傾き	b_1	-0.00015	0.000153	-0.00046	0.000149
第二区間の傾き	b_2	-0.00915	0.000877	-0.0109	-0.00742
第三区間の傾き	b_3	-0.00176	0.000285	-0.00233	-0.00120
第一変化点	t_1	558.0	3.3211	551.4	564.5
第二変化点	t_2	610.0	4.3981	601.3	618.6

PROC NLIN で推定した区分線形回帰モデルを Figure 1 の散布図に加えたグラフは以下のとおりである。

Figure 2. モデルを加えた散布図



3.7　PROC NLIN を使った区分線形回帰モデルの解説

PROC NLIN による算出結果の回帰係数と変化点の値を式(3)に代入と以下になる。

$$LOGRATIO = \begin{cases} 0.0163 + (-0.00015) * RANGE & 390 < RANGE \leq 558.0 \\ 0.0163 + (-0.00015) * 558.0 + (-0.00915) * (RANGE - 558.0) & 558.0 < RANGE \leq 610.0 \\ 0.0163 + (-0.00015) * 558.0 + (-0.00015) * (610.0 - 558.0) + (-0.00176) * (RANGE - 610.0) & 610.0 < RANGE < 720 \end{cases} \quad (4)$$

最終的な結果は以下のモデルとなった。

$$LOGRATIO = \begin{cases} 0.0163 - 0.00015 * RANGE & 390 < RANGE \leq 558.0 \\ 5.0363 - 0.00915 * RANGE & 558.0 < RANGE \leq 610.0 \\ 0.9984 - 0.00176 * RANGE & 610.0 < RANGE < 720 \end{cases} \quad (5)$$

本分析では、LOGRATIO と RANGE の関係を区分線形モデルにより表現した。PROC NLIN を用いることで、非線形最小二乗法によるパラメータ推定が行われ、二つの変化点に対しても推定値および信頼区間が算出された。Table 9 の結果では、最初の変化点が 558.0 (95%CI: [551.4, 564.5])、2 つ目が 610.0 (95%CI: [601.3, 618.6])に存在することを示すものであり、それぞれの区間における回帰係数も同時に推定される。具体的には、RANGE が 390〜558.0 の区間（第一変化点まで）では LOGRATIO は RANGE の増加とともに緩やかに減少(傾き: -0.00015)し、558.0〜610.0 の区間（第一変化点から第二変化点まで）ではより急激に減少(傾き: -0.00915)する傾向が認められた。610.0〜720 の区間（第二変化点以降）では再び減少の傾きが小さくなっている(傾き: -0.00176)。以上より、RANGE の区分ごとに LOGRATIO の減少傾向が異なり、558.0 および 610.0 付近で関係性が変化することが明らかとなった。

PROC NLIN による Table 8 の分散分析表では、p 値が 0.0001 未満となり、区分線形回帰モデルが有意に LIDAR データに適合していることが示された。

4　PROC NLIN の限界

本論文で用いた SAS の PROC NLIN は、非線形回帰モデルの推定において有用なプロシジャである。一方で、その使用にはいくつかの技術的な制約が存在する。特に、初期パラメータ値の設定が推定結果や収束性に大きく影響する点、線形モデル用プロシジャと比べて回帰診断に利用できる情報が限られている点、そしてランダム効果を含む階層構造や集団間変動を明示的にモデル化することができない点が挙げられる。以下では、これらの主な限界について詳述する。

4.1　初期パラメータ設定

非線形回帰に使用される PROC NLIN では、各パラメータの初期値を指定する必要がある。推定は初期値から反復的に進められるため、初期値が不適切であるとモデルが収束しない場合や、最適なパラメータ値に到達できず、局所解で推定が終了する可能性がある。その結果、得られる推定結果の信頼性が損なわれる場合がある。したがって、初期値の選定は推定結果の妥当性や収束性に大きく影響するため、十分な検討が必要となる。

4.2　回帰診断情報が少ない

PROC NLIN には診断指標の出力が非常に限定的であるという重要な技術的制約が存在する。例えば、PROC REG や PROC GLM などの線形モデル用プロシジャでは、標準化残差・スチューデント化残差、レバレッジ値、Cook の距離、DFFITS、分散比などの詳細な回帰診断指標が自動的に算出されるが、PROC NLIN ではこれらの指標が出力されない。そのため、データ点ごとの影響度や外れ値・高レバレッジ観測値の定量的評価、本格的なモデル適合度検証が困難となる。

4.3　ランダム効果導入不可

PROC NLIN は固定効果のみを持つ非線形モデルの推定しか行うことができず、ランダム効果や階層構造を持つモデルには対応していない。そのため、被験者内の繰り返し測定やグループ化されたデータなど、データ間の相関や構造的なばらつきを考慮する必要がある場合には使用できない。ランダム効果を含む非線形混合効果モデルについては、PROC NLMIXED など、専用の解析手法を用いる必要がある。

5　最後に

本論文では、データ内に区分ごとの変化点が存在する場合のモデル化手法として、区分線形回帰（piecewise linear regression）を取り上げ、その実装方法および解析上の有用性について検討を行った。SAS の PROC NLIN を活用し、変化点の位置および各区間の回帰係数を非線形最小二乗法で同時に推定する手順を具体的な LIDAR データセット（2 つの変化点を持つモデル）を用いて示した。その結果、区分線形回帰モデルは区間ごとに異なる傾向を明確に捉えることができ、従来の単純な線形回帰では十分に表現できないデータの構造的な変化を把握するうえで有効であることが示された。

一方で、本研究で用いた PROC NLIN にはいくつかの限界も認められる。特に、推定の際に設定する初期パラメータ値が結果に大きく影響するため、適切な初期値の選択が解の収束や妥当な推定結果の取得に不可欠となる。また、回帰モデルの診断情報が限定的であり、外れ値やモデル適合度の詳細な評価が困難であること、さらにランダム効果や階層構造を持つモデルへの対応ができないなど、解析の柔軟性に制約がある。これらの課題を十分に認識した上で、データや解析目的に応じて適切な手法を選択することが、今後の研究において重要となる。

参考文献

*1 Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). Semiparametric Regression. Cambridge: Cambridge University Press.

2値分類変数に対するベイズ流例数設計のSASによる実装

○高橋 朋寛，古賀 優一，平野 真太郎

（エイツーヘルスケア株式会社 データサイエンス本部 データサイエンスソリューション部）

Bayesian approach sample size determination for binary variables using SAS

Tomohiro Takahashi, Yuichi Koga and Shintaro Hirano
Data Science Solutions Department, A2 Healthcare Corporation

要旨

ベイズ流アプローチは，従来の頻度主義的手法と比べ，臨床試験の途中で結果をモニタリングしながら意思決定を行うような，柔軟な対応が必要な場合に特に有望である．しかし，臨床試験では科学的妥当性・倫理性・資源的制約の観点から，必要な標本サイズを事前に設計することが不可欠である．ベイズ流例数設計に対しては様々なアプローチが提案されており，推論上の性能に基づく方法として，例えば検出力を基準とした場合，解析事前分布とデザイン事前分布という「2 つの事前分布」を設定し，所望の検出力を満たす標本サイズを見積もる方法がある．

本稿では，主要評価項目を 2 値分類変数とする臨床試験を例に，ベイズ流例数設計の SAS による実装方法を検討し，検出力を基準とした 2 つの事前分布に基づくアプローチを中心に，動作特性を評価する．既報の論文で示された方法及び Web で公開されている計算ツールに対応する SAS プログラムを示して計算手順を詳述することにより，ベイズ流の例数設計手法の特徴を再考する．さらに，例数設計基準の設定や２つの事前分布それぞれの選択が必要例数に及ぼす影響を詳述し，試験設計時の留意点を考察する．

キーワード：標本サイズ設計，ベイズ流検出力，デザイン事前分布

1. 背景

医薬品開発の分野で実施される臨床試験は，基本的には第 I 相から第 III 相に分類される．第 II 相試験では, 開発中の薬剤が第 III 相の大規模なランダム化比較試験を実施するに値する十分な可能性を有しているかを判断することを目的に，一般的には複数の投与量を用いて探索的に検討される．このとき，倫理的な観点から，治療が明らかに不適切な場合，試験が過度に長期化しないようにステージを分けて試験がデザインされることもある (例えば，Simon 1989).

ベイズ流アプローチは，従来の頻度主義的手法と比べ，臨床試験の途中で結果をモニタリングしながら意思決定を行うような，柔軟な対応が必要な場合に特に有望である．しかし，臨床試験では科学的妥当性・倫理性・資源的制約の観点から，必要な標本サイズを事前に設計することが不可欠である (手良向 2023). ベイズ流例数設計に対しては様々なアプローチが提案されておる．大きく分けて決定理論に基づく方法と推論上

の性能に基づく方法があり，それぞれ日本語の文献としては坂巻ら (2020) および手良向 (2023) による解説がある．推論上の性能に基づく方法は，伝統的な頻度論による例数設計との類似点が多いため，利害関係者の理解が得られやすいとされることがある (手良向 2023). 推論上の性能に基づく方法において，性能の指標としては事後分散の大きさ, 信用区間の区間長や, ある仮説が満たされるであろう予測確率などが挙げられ，この予測確率を基準とする方法は，頻度論による検出力関数のベイズ流版とみなせるため「検出力基準」と呼ばれる (手良向 2023).

検出力基準を用いる場合，解析事前分布とデザイン事前分布という「2 つの事前分布」を設定し，所望の検出力を満たす標本サイズを見積もる方法がある (Sambucini 2008, Teramukai et al. 2012).「デザイン」事前分布は，解析時に直接ベイズの定理に代入するものではなく，データを入手する前に期待する「事前」予測分布を得るために用いるものである. 手良向 (2023) は，1 標本 2 項モデルにおける標本サイズの数値例を示し，更に単群の 2 値評価項目試験デザインの標本サイズ計算ツール（R Shiny アプリケーション）は，京都府立医科大学大学院医学研究科生物統計学のウェブサイト (https://bios-kpum.shinyapps.io/BayesSSD/) に公開されている．また，矢田ら (2016) は，比較的複雑な生存時間モデルを想定した群間比較における応用を考え，必要例数を解析的に算出できない状況でのシミュレーションに基づいた方法を SAS による実装例とともに示した．このとき，対数ハザード比が漸近的に正規分布に従うと仮定された.

手良向 (2023) は，平均に関する 1 標本片側検定に対して 2 つの事前分布に基づく例数設計を行ったときのベイズ流条件付き検出力が，古典的な頻度論による検出力関数と一致する条件も詳述している．臨床試験の文脈では伝統的に頻度論に基づく手法が採用されてきたため，ベイズ流アプローチを採用する際に，頻度論的動作特性の評価とともに，同様の条件において頻度論を採用した場合の結果の確認が有用になることがある．連続データではなく離散データの場合，頻度論においても近似を用いた手法，正確 (exact) な方法がそれぞれ存在し，ベイズ流の方法においても事前分布の選択が論点になる場合がある． De & Lecoutre (2010) は，特にサンプルサイズが小さい場合において，一般的に「無情報」事前分布と分類される分布のなかでも複数の分布を取り扱い，それぞれ異なる挙動をとることを示唆した.

本稿は，アウトカムに探索的試験で採用されやすい 2 値分類変数を設定し，解析的な計算が可能な状況における標本サイズ計算方法を SAS プログラムで示すことにより，ベイズ流検出力を基準とした 2 つの事前分布に基づくアプローチの特徴を再考するとともに，離散データに対する動作特性を詳述することを目的とする．さらに，頻度論による計算結果も踏まえつつ，2 つの事前分布それぞれの選択が必要例数に及ぼす影響を詳述し，試験設計時の留意点を考察する.

2. 方法

2.1 2 つの事前分布を用いたベイズ流例数設計

検出力を基準とした予測分布に基づく例数設計において，解析事前分布とデザイン事前分布という 2 種類の事前分布を用いる方法がある (Sambucini 2008). 解析事前分布は，解析時に事後分布を導出するために用いられるものであり，臨床試験においては，客観的な根拠（先行研究のデータなど）に基づいて定義されるのが一般的である．こうした客観的証拠が存在しない場合には，無情報あるいは弱情報の事前分布を採用することで，利害関係者の理解と合意を得やすくなると考えられている（Walley et al., 2015）．一方，デザイン事前分布は，臨床試験の計画段階において，観測データの事前予測分布を構築するために用いられるもの

である．この分布は，頻度主義的枠組みにおける対立仮説のイメージに相当し，試験設計における仮説的なシナリオを定式化する役割を担う．

反応確率を表すパラメータを θ，標本サイズ n の標本を $\boldsymbol{Y}_n$ とする．解析事前分布，デザイン事前分布をそれぞれ $p^{\mathrm{A}}(\theta), p^{\mathrm{D}}(\theta)$ とし，反応確率の目標値を θ^{T} とする．この目標値 θ^{T} は「反応確率がある目標値を超える」という仮説に関する情報である．この仮説の事後確率に対する確率閾値を λ とし，ベイズ流予測検出力に対応する確率閾値を γ とする．ここで便宜的に，厳密な定義を犠牲にして頻度論による例数設計との対応を考えると，目標値 θ^{T} は頻度論の仮説検定における帰無仮説の値に対応し，あるデータが与えられたとき「反応確率がある目標値を超える」という仮説の事後確率がある確率閾値 λ を超えるかどうかを意思決定の基準とするとき，頻度論の第一種の過誤確率 α との関係は $\alpha = 1 - \lambda$ と考えられる．

2 値分類変数をアウトカムとする臨床試験を想定し，本稿では手良向 (2023) に倣い，次のような表記を行う．標本サイズ n において，反応例数（例えば「有効」という結果が得られた例数）を R_n，その観測値 r_n とする．R_n が反応確率 θ をパラメータとする二項分布に従うとすると，

$$p(\boldsymbol{y}_n|\theta) \sim \mathrm{Binomial}(n, \theta).$$

ここで，θ の解析事前分布としてベータ分布を導入する．

$$p^{\mathrm{A}}(\theta) \sim \mathrm{Beta}(a^{\mathrm{A}}, b^{\mathrm{A}}).$$

ベータ分布と二項分布の共役性により，観測データで更新した事後分布は以下のベータ分布となる．

$$p(\theta|y) \propto p(y|\theta)p^{\mathrm{A}}(\theta) \sim \mathrm{Beta}(a^{\mathrm{A}} + r_n,\ b^{\mathrm{A}} + n - r_n)$$

このとき「反応確率がある目標値を超える」という仮説の事後確率に対する確率閾値を λ とすることは，

$$\boldsymbol{p(\theta > \theta^{\mathrm{T}}|R_n = r_n) = \int_{\theta^*}^{1} \mathrm{Beta}(a^{\mathrm{A}} + r_n,\ b^{\mathrm{A}} + n - r_n) > \lambda} \tag{1}$$

に対応する．

また，デザイン事前分布を以下の単峰ベータ分布の形式で設定する．

$$p^{\mathrm{D}}(\theta) \sim \mathrm{Beta}(a^{\mathrm{D}}, b^{\mathrm{D}})$$

$$a^{\mathrm{D}} = n^{\mathrm{D}} * \theta^{\mathrm{D}} + 1 \text{ および } b^{\mathrm{D}} = n^{\mathrm{D}} * (1 - \theta^{\mathrm{D}}) + 1$$

ここで、θ^{D} は分布のモード，n^{D} は分布の分散を調整するパラメータである．この n^{D} によって，デザイン事前分布の不確実性を考慮に入れることが可能である．$n^{\mathrm{D}} = \infty$ のとき，デザイン事前分布はモードが θ^{D} の退化分布となる．デザイン事前分布 $p^{\mathrm{D}}(\theta)$ に基づき，仮想的に n 例を観測した後の事前予測分布は以下のようなベータ二項分布となる．

$$p(r_n) = \int_0^1 p(r_n|\theta)\, p^{\mathrm{D}}(\theta) d\theta = \text{ Beta-Binomial}(n, a^{\mathrm{D}}, b^{\mathrm{D}})$$

$$= \binom{n}{r_n} \frac{\mathrm{B}(a^{\mathrm{D}} + r_n, b^{\mathrm{D}} + n - r_n)}{\mathrm{B}(a^{\mathrm{D}}, b^{\mathrm{D}})}.$$

ここで $\mathrm{B}(\cdot, \cdot)$ はベータ関数である．したがって，ベイズ流検出力 γ を用いて，以下の式を満たす最小の n が標本サイズとして求められる．

$$\boldsymbol{\sum_{r_{n_{min}}}^{n} p(r_n) \geq \gamma} \tag{2}$$

2.2 SAS による計算手順

ここでは，手良向 (2017) に示された R コードを参考に， 2 つの事前分布に基づく例数設計法の SAS による実装方法を示す．計算アルゴリズムを踏まえ，2 つの事前分布に基づく例数設計の手順を再考すると，次のように表現できる．なお，プログラム全体は Appendix A に記載した．

(1) 必要なパラメータ，基準値を設定する
(2) 特定の標本サイズおよび反応例数 r_n において「反応確率が目標値を超える」事後確率を計算する
(3) その事後確率が λ を超える r_n を特定する
(4) デザイン事前分布に基づき，仮想的に観測した n 例のデータに基づく事前予測分布を得る
(5) 事前予測分布に基づき，事後確率が λ を超える r_n を得られる確率が γ を超えるかどうか判定する
(6) n を変えて計算を続け，(5) を満たす n のうち，最小の値を標本サイズとする．

(1) において，本稿のプログラムでは，解析事前分布のパラメータを&alpha.と&beta.，デザイン事前分布のパラメータを&mode.と&nd.，仮説の目標値を&Target.，ベイズ流検出力を&lambda.とし，計算上必要な最大例数を&n_max.で設定した．プログラム 2.1 に，手順 (2) から (4) に対応するステップを示す．

プログラム 2.1: 計算手順 (2) – (4) に相当する SAS プログラム

```
data beta_cdf;
        do n=1 to &n_max.;
        min_x=-1;
                do until (cdf>&lambda. or min_x=n);
                        min_x=min_x+1;
                        alpha01=&alpha.; beta01=&beta.;
                        cdf=1-cdf('beta',&Target.,(alpha01+min_x),(beta01+n-min_x));
                end;
                output;
        end;
run;
data distribution;
        set beta_cdf;
        sum_dest=0;
        do x=min_x to n;
                a=&mode.*&nd.+1; b=(1-&mode.)*&nd.+1;
                destiny= comb(n,x) * ( beta(x+a,n-x+b) / beta(a,b) );
                sum_dest=sum_dest+destiny;
        end;
        seq=-1*n;
run;
```

次に，プログラム 2.2 に，手順 (5) から (6) に対応するステップを示す．反応者数 X は離散的な確率変数であり，事前予測確率：$\sum_{r_{n_{min}}}^{n} p(x)$ は n の増加に対して鋸歯状の挙動を示す．このため γ を超える最小の n を定義する際には注意が必要である．ここでいう「最小の n」とは，「それ以上のすべての n に対して，式 (2) を満たすような最小の n」であることを指している．

プログラム 2.2: 計算手順 (5) – (6) に相当する SAS プログラム

```
data Result;
        set distribution;
        where sum_dest>&gamma.;
        retain seq2 %eval(&n_max.+1);
        seq2=seq2-1;
        if n=seq2;
run;
data Result_min;
        set Result end=last;
        if last;
        call symputx('n_Result',n);
        call symputx('n_min_Result',min_x);
        keep sum_dest min_x cdf n;
run;
```

SAS の COMB 関数の仕様により，デザイン事前分布の&nd.や最大例数&n_max がある程度大きい値になると計算できない場合があるが，少なくとも 800 程度までは計算可能である．したがって，頻度論の対立仮説の考え方に沿って&nd.を非常に大きな値に設定したい場合は，場合分けを行い 2 項分布の累積分布関数を計算するようにプログラミングする必要がある (Appendix A).

2.3 数値例の算出

本稿では，前項の SAS プログラムを用いて，目標値 θ^{T}，デザイン事前分布のパラメータ θ^{D}，n^{D} に関する複数の状況を検討した．すなわち，θ^{T} を 0.05 ずつ変化させた場合に対して，デザイン事前分布は θ^{D} を θ^{T}+0.2 とし，n^{D} を 50、∞とした．また，一般的に「無情報」とされる事前分布の中でも Beta(1,1), Beta(0.5,0.5) 及び Beta(0,0) の 3 種類を採用した場合の検出力及び標本サイズをそれぞれ記述した．また，頻度論の場合に採用されやすい基準と比較しやすくするため，$\lambda = 0.975, \gamma = 0.80$ については固定した．

また，頻度論の場合の例数設計には POWER プロシジャの ONESAMPLEFREQ ステートメントを使用した．使用したオプションとして，デフォルトである TEST=EXACT に加え，TEST=Z 及び TEST=ADJZ に対してはいずれも METHOD=EXACT 及び METHOD=NORMAL を選択できることから，合わせて 5 通りのオプションを使用した．TEST=Z 及び TEST=ADJZ に対して METHOD=EXACT を使用するオプションについて，筆者の知る限りでは使用機会があまりないと思われるが，棄却限界値の設定までを近似的に行い，それ

以降の検出力の計算等においては TEST=EXACT とした場合と同様に F 分布を用いて計算される (SAS, 2025).

なお，本稿の例数設計に関する計算はすべて Base SAS 9.4_M8 及び SAS/STAT 15.3 を用いて実施した．グラフによる可視化には部分的に python 3.12.4 を用いた.

3. 結果

3.1 例数設計過程の数値例

ここでは，2.1 項の式 (1)，(2) に対応する数値例をプログラム 2.1, 2.2 を用いて示すことにより，2 つの事前分布に基づくベイズ流例数設計方法の手順を詳述する.

まず，式 (1) の判定に対応する数値例を示す．$\theta^{\mathrm{T}} = 0.3,\ \lambda = 0.975,\ \gamma = 0.80$ において，解析事前分布を Bata(1,1)，Bata(0.5,0.5)とした場合を考える。後述する検出力の検討も踏まえ，n = 54 とした場合を例にとる. プログラム 2.1 の過程では, 反応例数 r_n が 54 例中 1, 2, …54 例であった場合の事後確率 $\boldsymbol{p}(\boldsymbol{\theta} > \theta^{\mathrm{T}}|R_n = r_n)$ が順次算出される．ここで，$r_n = 23$ のとき，Figure 1 に示すように，いずれの事前分布に対しても事後確率が 97.5 %以上となる値が得られた．ここで，オレンジ色で塗りつぶされた領域が $p(\theta > \theta^{\mathrm{T}}|y)$ に該当し，それぞれ 97.8 %，97.6 %であった.

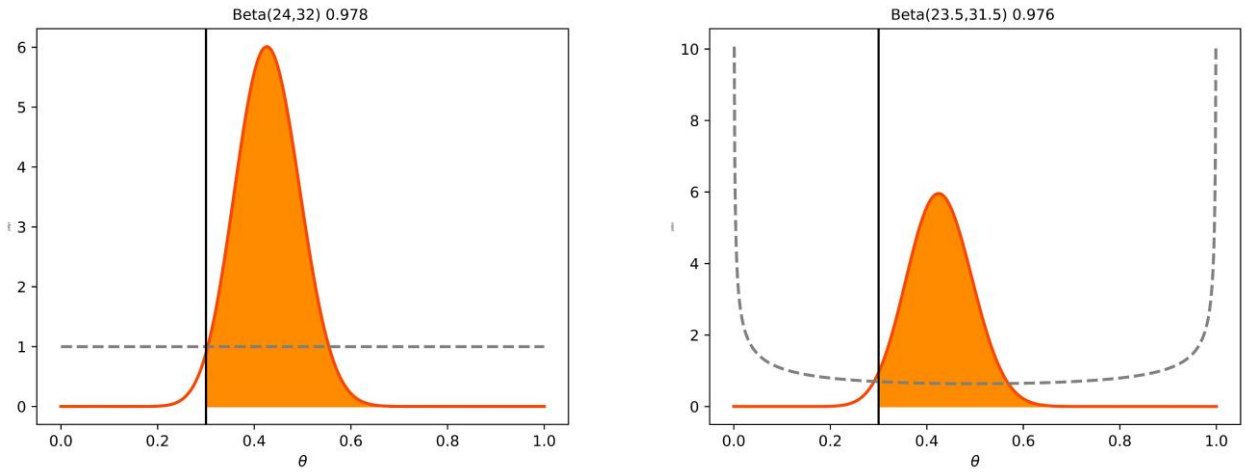


Figure 1 仮説 $\theta > 0.3$ に対する事後確率が閾値 $\lambda = 0.975$ を超える例

解析事前分布 (点線) を(左) Bata(1,1)，(右) Bata(0.5,0.5)としたときの，標本サイズ 54 例，反応例数 23 例における事後確率．実線 (黒) は目標値 (30 %) を示す.

次に，デザイン事前分布を $p_0^{\mathrm{D}} = 0.5,\ n^D = 50$ として，式 (2) の判定に対応する数値例を示す．$a^{\mathrm{D}} = n^{\mathrm{D}} * p_0^{\mathrm{D}} + 1 = 50 * 0.5 + 1 = 26$ および $b^{\mathrm{D}} = n^{\mathrm{D}} * (1 - \theta^{\mathrm{D}}) + 1 = 50 * 0.5 + 1 = 26$ より，ベータ-二項分布は Figure 2 のように表せる．ここで，前述の $r_n = 23$ を上回るデータが得られる事前予測確率が 80.4 % となり，検出力の基準 80 % を超えていることが分かる.

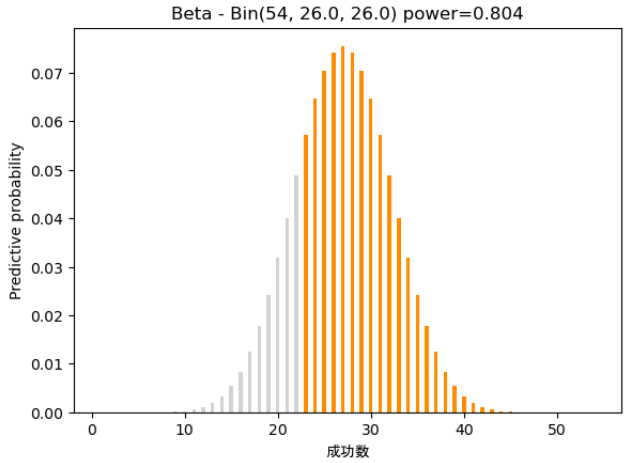


Figure 2 $r_n = 23$以上となる事前予測確率が閾値 $\gamma = 0.8$ を超える例

デザイン事前分布のパラメータを $\theta^{\mathrm{D}} = 0.5$, $n^{D} = 50$ としたときの，標本サイズ 54 例における事前予測確率．反応例数 $r_n = 23$ 以上の部分をオレンジ色で，それ未満の部分を灰色で示す．

実際には，前述の通り事前予測確率は n の増加に対して単調に増加しないため，これより大きな n でも判定を続ける必要がある．

ここで，ベイズ流の方法における Figure 1, 2 に対応する，頻度論の場合の判定基準を Figure 3 に示す．

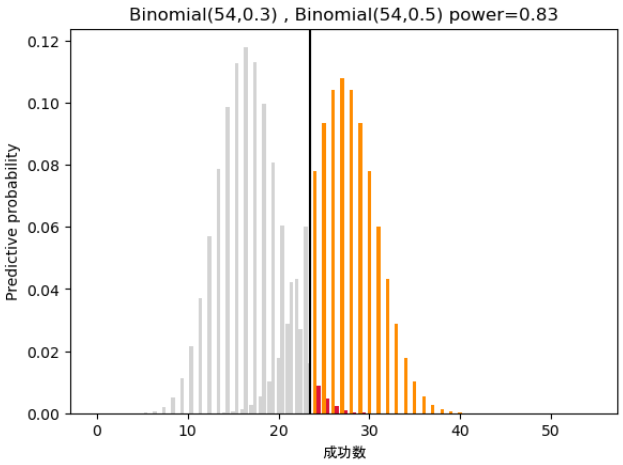


Figure 3 頻度論に基づく exact な方法における 2 つの 2 項分布

頻度論において類似の状況を検討するためには，帰無仮説として目標値 $\theta^{\mathrm{T}} = 0.3$ をとる仮説検定を考える必要がある．また，例数設計時には対立仮説として期待される特定の値を定めるが，ここでは $\theta^{\mathrm{D}} = 0.5$ がそれに対応する．このとき，例数のパラメータを n = 54 として，θ^{T}, θ^{D} それぞれを確率パラメータとしてもつ 2 つの二項分布に基づいて判定が行われる．Figure 3 は，帰無仮説を表す二項分布が左側に，対立仮説を表す 2 項分布が右側に描かれ，帰無仮説の下で $r_n = 23$ を上回るデータが得られる赤色の部分の確率が 1-0.975 未満に，対立仮説の下で $r_n = 23$ を上回るデータが得られるオレンジ色の部分の確率が 0.8 以上になることを示している．

3.2　標本サイズ及び検出力

2.2 項で示したプログラムを用い，2.3 項の設定に従って 2 つの事前分布に基づくベイズ流例数設計を実行し，対応する頻度論による算出結果とともに Table 1 に示した．概観して分かる通り，いずれの方法でも，$\boldsymbol{\theta}^*$ の設定に依存した例数のピークは$\boldsymbol{\theta}^* = 0.4$ または $\boldsymbol{\theta}^* = 0.45$ であり，$\boldsymbol{\theta}^*$と $\boldsymbol{\theta}^D$ の双方が 0.5 に近い場合に必要例数が大きくなる傾向が確認された．また，本稿で設定した小サンプルのシナリオの範囲では，頻度論及びベイズ流のアプローチ $(n^D = \infty,\ 50)$ は各方法内でいずれもオプションの設定あるいは事前分布の選択に依存して 10 例程度の差があった．

ベイズ流で算出された標本サイズに注目すると，$n^D = \infty$ の場合と $n^D = 50$ の比較において，$\boldsymbol{\theta}^* = 0.2, 0.5, 0.7$ のときそれぞれ 11 例，21 例，5 例程度 $n^D = 50$ の場合に必要例数が大きかった．また，事前分布の選択に注目すると，θ^* が 0.5 未満の場合，標本サイズは，Beta(1,1) < Beta(0.5,0.5) < Beta(0,0) の順に大きかったが，θ^* が 0.5 を超える場合にはこの傾向が逆転した．これは Beta (a, b) の事後分布の平均が $a/(a+b)$ であることを踏まえると，一様分布としての事前情報が 1/2 の 50 % の方向に影響したものと考えられる．

また，$\boldsymbol{\theta}^* = 0.3$ または $\boldsymbol{\theta}^* = 0.5$ の場合の検出力曲線を Figure 4，対応する検出力の値は Appendix B-E に示した．ここでは，2 項分布を用いた場合の鋸歯状の曲線が，n が大きくなるにつれて徐々に正規近似を用いた場合のなめらかな 2 本の曲線の間に収まっていく様子が確認された．また，鋸歯状の曲線は重なっている部分が多いが，2 本の曲線が完全に一致するものはなかった．例示として，Table 1 においても必要例数が非常に類似していた 2 つの方法，すなわち頻度論の方法において POWER プロシジャのオプションを (TEST=Z, METHOD=EXACT) と指定した場合とベイズ流の解析事前分布に Beta(1,1) を採用した場合を抜粋した図を Figure 4 (下) に示した．

Table 1 頻度論及び 2 つの事前分布に基づくベイズ流の標本サイズ

	頻度流（proc power）					ベイズ流					
	test					n^D					
	z		adjz		exact	∞			50		
	method					$p^A(\theta)$					
θ^T	normal	exact	normal	exact	exact	beta(1,1)	beta(0.5,0.5)	beta(0,0)	beta(1,1)	beta(0.5,0.5)	beta(0,0)
0.2	36	38	41	44	44	38	41	44	49	52	55
0.25	41	43	45	48	48	43	46	46	56	56	58
0.3	44	45	49	52	54	45	47	52	59	61	66
0.35	46	47	51	52	52	47	47	48	63	63	65
0.4	48	48	52	55	55	48	48	48	71	71	71
0.45	48	49	53	52	57	49	49	49	70	70	70
0.5	47	48	52	54	54	48	48	48	69	69	66
0.55	45	46	50	53	53	46	46	46	67	67	67
0.6	43	44	47	48	48	44	43	43	65	58	58
0.65	39	40	44	46	45	40	40	39	58	54	53
0.7	34	35	38	41	36	35	34	28	48	47	46
0.75	27	25	32	34	33	32	24	21	37	35	34

$\theta^D = \theta^T + 0.2$ 、$\gamma = 1 - \beta = 0.8$、$\lambda = 1 - \alpha = 0.975$とした

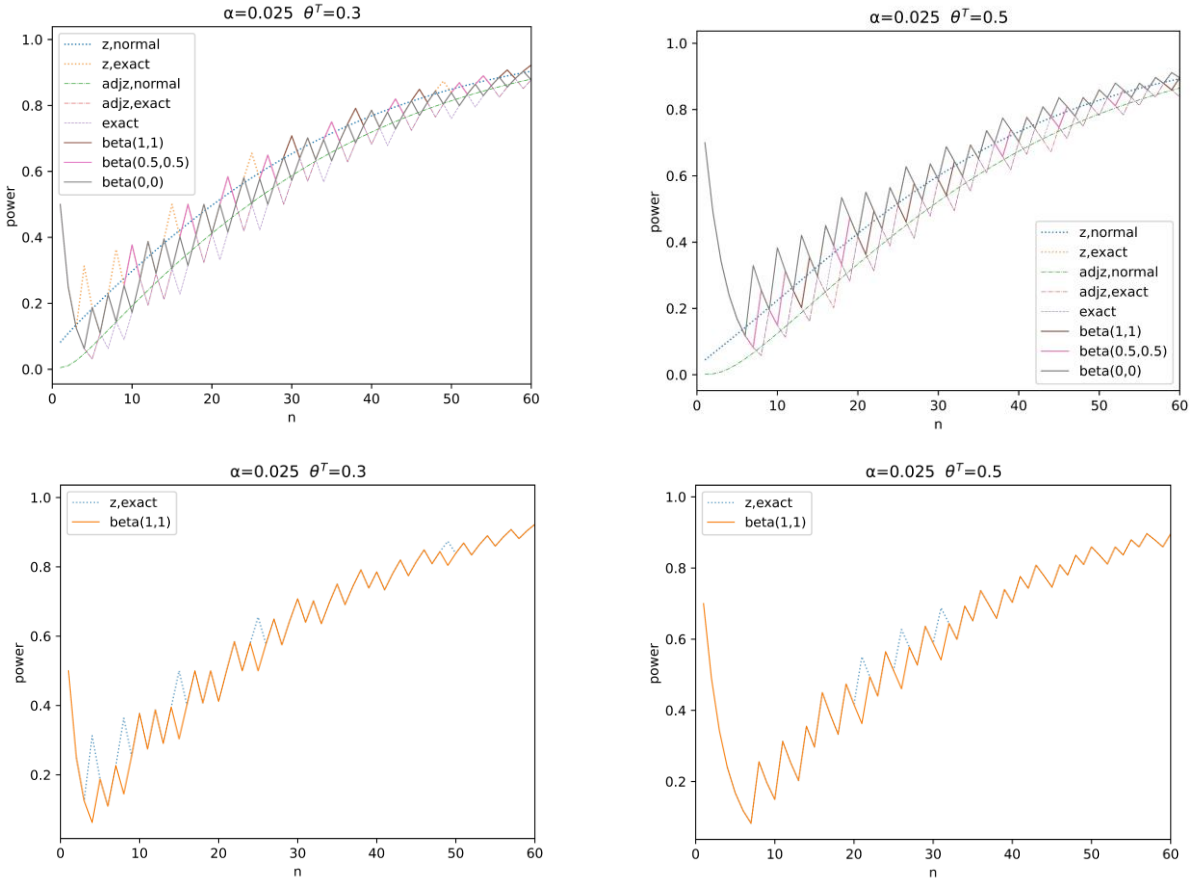


Figure 4 検出力曲線

(上)　目標値 $\theta^* = 0.3$ (左), 0.5 (右) のときの頻度論・ベイズ流それぞれの方法で計算された検出力曲線．ベイズ流の方法においては，$n^D = \infty$ の場合のみを示した．

(下)　上の検出力曲線のうち，POWER プロシジャの ONESAMPLEFREQ ステートメントにおけるオプション (TEST=Z, METHOD=EXACT) の場合とベイズ流の beta(1,1) の場合を抜粋したもの．

4. おわりに

本稿では，アウトカムに 2 値分類変数を設定した単群臨床試験のベイズ流例数設計法のうち， 2 つの事前分布に基づく方法に焦点を当て，その計算手順を詳述し，SAS による実装方法を紹介するとともに，頻度論の場合の結果と比較しながら例数設計の結果を示した．

3.1 項では，2 つの事前分布に基づくベイズ流例数設計方法の過程で鍵となる 2 つの判定基準に対応する確率分布を Figure 1, 2 に，頻度論的なアプローチにおける対応する判定基準を Figure 3 に示した．Figure 2 のベータ-二項分布は，対応する Figure 3 の対立仮説における二項分布と比較して，有限の n^D をもつ不確実性が反映され，分布がばらつくことになる．すなわち，ベイズ流のアプローチにおいて $n^D = \infty$ とすることが，頻度論的なアプローチに対応する条件となる．

一方で，Table 1，Figure 4 に示すように，$n^D = \infty$ の条件においても，頻度論とベイズ流の計算結果が，様々な目標値に対する幅広い条件下で完全に一致する方法はない．これは，Figure 2 で示すように．ベイズ流の方法においては確率閾値に対する判定を行うときに連続的なベータ分布を用いていることが特に影響している

かもしれない．このことは，Table 1 において，頻度論の方法において POWER プロシジャのオプションを (TEST=Z, METHOD=EXACT) と指定した場合と，ベイズ流の解析事前分布に Beta(1,1) を採用した場合の必要例数が非常に類似していたこととも関連付けられる．POWER プロシジャの当該オプションを指定した場合，閾値の算出までが近似による正規分布を用いて特定され，検出力の計算においては 2 項分布が用いられる．

本稿で用いた 3 つの事前分布の選択については，より詳細な検討が必要であるが，一様分布においても完全な「無情報」分布にはならないことが，Table 1 において目標値が 0.5 より大きいかどうかによって挙動が変わることに顕れている．これらの選択においては当然ながら標本サイズの大きさによってではなく，解析に関連する動作特性による評価が必要であり，De and Lecoutre (2010)らの記述は参考になるかもしれない．

しかしながら，本稿で比較した条件のうち，最も大きな影響を及ぼすのはデザイン事前分布のn^Dの選択であった．n^D に関する考察は，頻度論の例数設計においては，期待する治療効果に関して不確実性を含めないことを示唆するものとなる．矢田ら (2017) は「ヒストリカルデータを，そのデータが持つ情報量に応じて新規試験におけるデザインと解析内容に柔軟に組み入れることが可能なことが，ベイズ流例数設計の大きな魅力ではないだろうか」と述べている．手良向 (2023) には，デザイン事前分布を主に念頭に置いて，専門家らの意見を集約するためのフレームワークも要約されている．これらの知見をもとに，臨床試験の試験デザイン手法が洗練されていくことが期待される．

Appendix

Appendix A

```
%macro find_size(alpha=, beta=, mode=, nd=, Target=, lambda=, gamma=, n_max=200);
	data beta_cdf;
		do n=1 to &n_max.;
			min_x=-1;
			do until (cdf>&lambda. or min_x=n);
				min_x=min_x+1;
				alpha01=&alpha.; beta01=&beta.;
				cdf=1-
cdf('beta',&Target.,(alpha01+min_x),(beta01+n-min_x));
			end;
			output;
		end;
	run;

	%if &nd.=-1 %then %do;
		data distribution;
			set beta_cdf;
			do x=min_x to n;
				sum_dest = 1 - probbnml(&mode., n, min_x-1);
			end;
			seq=-1*n;
			proc sort ;
				by seq;
		run;
	%end;

	%else %do;
		data distribution;
			set beta_cdf;
			sum_dest=0;
			do x=min_x to n;
				a=&mode.*&nd.+1; b=(1-&mode.)*&nd.+1;
				destiny= comb(n,x) * ( beta(x+a,n-x+b) /
beta(a,b) );
				sum_dest=sum_dest+destiny;
			end;
```

```
                    seq=-1*n;
                    proc sort ;
                        by seq;
            run;
    %end;

    data Result;
            set distribution;
            where sum_dest>&gamma.;
            retain seq2 %eval(&n_max.+1);
            seq2=seq2-1;
            if n=seq2;
    run;

    data Result_min;
            set Result end=last;
            if last;
            call symputx('n_Result',n);
            call symputx('n_min_Result',min_x);
            keep sum_dest min_x cdf n %if    &nd.^=-1 %then %do; a b %end;;
    run;
%mend find_size;
```

Appendix B

	頻度論（proc power）					ベイズ流					
	test					n^D					
	z		adjz		exact	∞			50		
	method					$p^A(\theta)$					
n	normal	exact	normal	exact	exact	beta(1,1)	beta(0.5,0.5)	beta(0,0)	beta(1,1)	beta(0.5,0.5)	beta(0,0)
30	0.653	0.708	0.587	0.572	0.572	0.708	0.572	0.572	0.668	0.557	0.557
31	0.667	0.640	0.602	0.640	0.640	0.640	0.640	0.640	0.611	0.611	0.611
32	0.680	0.702	0.617	0.570	0.570	0.702	0.702	0.702	0.661	0.661	0.661
33	0.692	0.636	0.631	0.636	0.636	0.636	0.636	0.636	0.607	0.607	0.607
34	0.704	0.696	0.645	0.696	0.568	0.696	0.696	0.696	0.655	0.655	0.655
35	0.716	0.750	0.658	0.632	0.632	0.750	0.750	0.632	0.699	0.699	0.602
36	0.727	0.691	0.671	0.691	0.691	0.691	0.691	0.691	0.649	0.649	0.649
37	0.738	0.744	0.684	0.629	0.629	0.744	0.744	0.744	0.692	0.692	0.692
38	0.748	0.791	0.696	0.686	0.686	0.791	0.686	0.686	0.730	0.644	0.644
39	0.759	0.739	0.708	0.739	0.739	0.739	0.739	0.739	0.685	0.685	0.685
40	0.768	0.785	0.719	0.682	0.682	0.785	0.785	0.785	0.723	0.723	0.723
41	0.778	0.734	0.730	0.734	0.734	0.734	0.734	0.734	0.679	0.679	0.679
42	0.787	0.780	0.741	0.780	0.678	0.780	0.780	0.780	0.716	0.716	0.716
43	0.796	0.820	0.751	0.729	0.729	0.820	0.820	0.729	0.750	0.750	0.673
44	0.804	0.774	0.761	0.774	0.774	0.774	0.774	0.774	0.710	0.710	0.710
45	0.812	0.814	0.771	0.724	0.724	0.814	0.814	0.814	0.743	0.743	0.743
46	0.820	0.849	0.780	0.769	0.769	0.849	0.769	0.769	0.773	0.704	0.704
47	0.828	0.809	0.789	0.809	0.809	0.809	0.809	0.809	0.736	0.736	0.736
48	0.835	0.844	0.798	0.765	0.765	0.844	0.844	0.844	0.766	0.766	0.766
49	0.842	0.874	0.806	0.804	0.804	0.804	0.804	0.804	0.730	0.730	0.730
50	0.849	0.839	0.814	0.839	0.760	0.839	0.839	0.839	0.760	0.760	0.760
51	0.855	0.869	0.822	0.799	0.799	0.869	0.869	0.799	0.786	0.786	0.724
52	0.862	0.834	0.829	0.834	0.834	0.834	0.834	0.834	0.753	0.753	0.753
53	0.868	0.864	0.837	0.864	0.795	0.864	0.864	0.864	0.780	0.780	0.780
54	0.873	0.890	0.843	0.830	0.830	0.890	0.890	0.830	0.804	0.804	0.747
55	0.879	0.860	0.850	0.860	0.860	0.860	0.860	0.860	0.773	0.773	0.773
56	0.884	0.886	0.857	0.825	0.825	0.886	0.886	0.886	0.798	0.798	0.798
57	0.889	0.908	0.863	0.855	0.855	0.908	0.855	0.855	0.820	0.767	0.767
58	0.894	0.881	0.869	0.881	0.881	0.881	0.881	0.881	0.791	0.791	0.791
59	0.899	0.904	0.874	0.851	0.851	0.904	0.904	0.904	0.813	0.813	0.813
60	0.904	0.922	0.880	0.877	0.877	0.922	0.877	0.877	0.833	0.785	0.785

$p^D(\theta)=\theta^T+0.2$ 、$\gamma=1-\beta=0.8$、 $\lambda=1-\alpha=0.975$とした

Appendix C

	頻度論（proc power）					ベイズ流					
	test					n^D					
	z		adjz		exact	∞			50		
	method					$p^A(\theta)$					
n	normal	exact	normal	exact	exact	beta(1,1)	beta(0.5,0.5)	beta(0,0)	beta(1,1)	beta(0.5,0.5)	beta(0,0)
30	0.599	0.589	0.523	0.589	0.589	0.589	0.589	0.589	0.551	0.551	0.551
31	0.615	0.688	0.540	0.542	0.542	0.542	0.688	0.688	0.513	0.631	0.631
32	0.629	0.644	0.556	0.495	0.495	0.644	0.644	0.644	0.594	0.594	0.594
33	0.644	0.599	0.572	0.599	0.599	0.599	0.599	0.599	0.557	0.557	0.557
34	0.658	0.693	0.588	0.554	0.554	0.693	0.693	0.693	0.632	0.632	0.632
35	0.671	0.652	0.603	0.652	0.652	0.652	0.652	0.652	0.597	0.597	0.597
36	0.684	0.737	0.618	0.609	0.609	0.737	0.737	0.737	0.666	0.666	0.666
37	0.697	0.698	0.633	0.698	0.698	0.698	0.698	0.698	0.633	0.633	0.633
38	0.709	0.659	0.647	0.659	0.659	0.659	0.659	0.774	0.601	0.601	0.696
39	0.721	0.740	0.661	0.618	0.618	0.740	0.740	0.740	0.665	0.665	0.665
40	0.733	0.703	0.674	0.703	0.703	0.703	0.703	0.703	0.634	0.634	0.634
41	0.744	0.776	0.687	0.665	0.665	0.776	0.776	0.776	0.694	0.694	0.694
42	0.755	0.743	0.700	0.743	0.743	0.743	0.743	0.743	0.665	0.665	0.665
43	0.765	0.808	0.712	0.708	0.708	0.808	0.808	0.808	0.720	0.720	0.720
44	0.775	0.778	0.724	0.672	0.778	0.778	0.778	0.778	0.692	0.692	0.692
45	0.785	0.746	0.735	0.746	0.746	0.746	0.746	0.836	0.664	0.664	0.743
46	0.794	0.809	0.746	0.713	0.713	0.809	0.809	0.809	0.717	0.717	0.717
47	0.803	0.780	0.757	0.780	0.780	0.780	0.780	0.780	0.691	0.691	0.691
48	0.812	0.836	0.767	0.749	0.749	0.836	0.836	0.836	0.739	0.739	0.739
49	0.820	0.810	0.777	0.810	0.810	0.810	0.810	0.810	0.714	0.714	0.714
50	0.828	0.859	0.787	0.782	0.782	0.859	0.859	0.859	0.759	0.759	0.759
51	0.836	0.836	0.796	0.836	0.836	0.836	0.836	0.836	0.736	0.736	0.736
52	0.843	0.811	0.805	0.811	0.811	0.811	0.811	0.880	0.712	0.712	0.777
53	0.851	0.859	0.813	0.784	0.784	0.859	0.859	0.859	0.755	0.755	0.755
54	0.857	0.837	0.821	0.837	0.837	0.837	0.837	0.837	0.732	0.732	0.732
55	0.864	0.879	0.829	0.813	0.813	0.879	0.879	0.879	0.772	0.772	0.772
56	0.870	0.859	0.837	0.859	0.859	0.859	0.859	0.859	0.751	0.751	0.751
57	0.876	0.897	0.844	0.837	0.837	0.897	0.897	0.897	0.788	0.788	0.788
58	0.882	0.879	0.851	0.879	0.879	0.879	0.879	0.879	0.768	0.768	0.768
59	0.888	0.859	0.858	0.859	0.859	0.859	0.912	0.912	0.748	0.803	0.803
60	0.893	0.896	0.865	0.838	0.838	0.896	0.896	0.896	0.784	0.784	0.784

$p^D(\theta)=\theta^T+0.2$ 、$\gamma=1-\beta=0.8$、$\lambda=1-\alpha=0.975$

Appendix D

	頻度論（proc power）					ベイズ流					
	test					n^D					
	z		adjz		exact	∞			50		
	method					$p^A(\theta)$					
n	normal	exact	normal	exact	exact	beta(1,1)	beta(0.5,0.5)	beta(0,0)	beta(1,1)	beta(0.5,0.5)	beta(0,0)
30	0.237	0.285	0.178	0.175	0.175	0.285	0.175	0.175	0.335	0.237	0.237
31	0.243	0.219	0.185	0.219	0.219	0.219	0.219	0.219	0.279	0.279	0.279
32	0.249	0.268	0.191	0.165	0.165	0.268	0.268	0.268	0.322	0.322	0.322
33	0.254	0.206	0.197	0.206	0.206	0.206	0.206	0.206	0.270	0.270	0.270
34	0.260	0.251	0.203	0.251	0.155	0.251	0.251	0.251	0.311	0.311	0.311
35	0.266	0.300	0.209	0.193	0.193	0.300	0.300	0.193	0.352	0.352	0.261
36	0.271	0.236	0.215	0.236	0.236	0.236	0.236	0.236	0.300	0.300	0.300
37	0.277	0.282	0.221	0.182	0.182	0.282	0.282	0.282	0.340	0.340	0.340
38	0.283	0.330	0.227	0.222	0.222	0.330	0.222	0.222	0.380	0.290	0.290
39	0.288	0.265	0.233	0.265	0.265	0.265	0.265	0.265	0.328	0.328	0.328
40	0.294	0.311	0.239	0.209	0.209	0.311	0.311	0.311	0.366	0.366	0.366
41	0.299	0.250	0.245	0.250	0.250	0.250	0.250	0.250	0.317	0.317	0.317
42	0.305	0.294	0.251	0.294	0.197	0.294	0.294	0.294	0.354	0.354	0.354
43	0.310	0.340	0.257	0.236	0.236	0.340	0.340	0.236	0.391	0.391	0.307
44	0.316	0.277	0.262	0.277	0.277	0.277	0.277	0.277	0.342	0.342	0.342
45	0.321	0.321	0.268	0.222	0.222	0.321	0.321	0.321	0.378	0.378	0.378
46	0.327	0.367	0.274	0.262	0.262	0.367	0.262	0.262	0.413	0.332	0.332
47	0.332	0.304	0.280	0.304	0.304	0.304	0.304	0.304	0.366	0.366	0.366
48	0.338	0.348	0.286	0.247	0.247	0.348	0.348	0.348	0.400	0.400	0.400
49	0.343	0.393	0.291	0.288	0.288	0.288	0.288	0.288	0.355	0.355	0.355
50	0.348	0.330	0.297	0.330	0.234	0.330	0.330	0.330	0.388	0.388	0.388
51	0.354	0.374	0.303	0.272	0.272	0.374	0.374	0.272	0.420	0.420	0.344
52	0.359	0.313	0.308	0.313	0.313	0.313	0.313	0.313	0.376	0.376	0.376
53	0.364	0.355	0.314	0.355	0.258	0.355	0.355	0.355	0.408	0.408	0.408
54	0.369	0.398	0.319	0.297	0.297	0.398	0.398	0.297	0.439	0.439	0.365
55	0.375	0.337	0.325	0.337	0.337	0.337	0.337	0.337	0.396	0.396	0.396
56	0.380	0.379	0.330	0.281	0.281	0.379	0.379	0.379	0.427	0.427	0.427
57	0.385	0.422	0.336	0.321	0.321	0.422	0.321	0.321	0.457	0.385	0.385
58	0.390	0.361	0.341	0.361	0.361	0.361	0.361	0.361	0.415	0.415	0.415
59	0.395	0.402	0.347	0.305	0.305	0.402	0.402	0.402	0.444	0.444	0.444
60	0.400	0.444	0.352	0.344	0.344	0.444	0.344	0.344	0.473	0.403	0.403

$p^D(\theta)=\theta^T+0.1$ 、$\gamma=1-\beta=0.8$、$\lambda=1-\alpha=0.975$

Appendix E

	頻度論（proc power）					ベイズ流					
	test					n^D					
	z		adjz		exact	∞			50		
	method					$p^A(\theta)$					
n	normal	exact	normal	exact	exact	beta(1,1)	beta(0.5,0.5)	beta(0,0)	beta(1,1)	beta(0.5,0.5)	beta(0,0)
30	0.189	0.176	0.135	0.176	0.176	0.176	0.176	0.176	0.222	0.222	0.222
31	0.194	0.245	0.141	0.143	0.143	0.143	0.245	0.245	0.192	0.285	0.285
32	0.199	0.205	0.146	0.116	0.116	0.205	0.205	0.205	0.250	0.250	0.250
33	0.204	0.169	0.151	0.169	0.169	0.169	0.169	0.169	0.219	0.219	0.219
34	0.209	0.233	0.156	0.138	0.138	0.233	0.233	0.233	0.277	0.277	0.277
35	0.214	0.195	0.162	0.195	0.195	0.195	0.195	0.195	0.245	0.245	0.245
36	0.219	0.262	0.167	0.162	0.162	0.262	0.262	0.262	0.303	0.303	0.303
37	0.224	0.222	0.172	0.222	0.222	0.222	0.222	0.222	0.271	0.271	0.271
38	0.229	0.186	0.177	0.186	0.186	0.186	0.186	0.290	0.241	0.241	0.328
39	0.234	0.248	0.183	0.155	0.155	0.248	0.248	0.248	0.295	0.295	0.295
40	0.239	0.211	0.188	0.211	0.211	0.211	0.211	0.211	0.265	0.265	0.265
41	0.244	0.275	0.193	0.178	0.178	0.275	0.275	0.275	0.318	0.318	0.318
42	0.249	0.236	0.198	0.236	0.236	0.236	0.236	0.236	0.288	0.288	0.288
43	0.254	0.301	0.204	0.201	0.201	0.301	0.301	0.301	0.341	0.341	0.341
44	0.259	0.261	0.209	0.170	0.261	0.261	0.261	0.261	0.310	0.310	0.310
45	0.264	0.225	0.214	0.225	0.225	0.225	0.225	0.327	0.281	0.281	0.362
46	0.269	0.286	0.219	0.192	0.192	0.286	0.286	0.286	0.331	0.331	0.331
47	0.274	0.249	0.224	0.249	0.249	0.249	0.249	0.249	0.302	0.302	0.302
48	0.279	0.311	0.229	0.214	0.214	0.311	0.311	0.311	0.351	0.351	0.351
49	0.284	0.272	0.235	0.272	0.272	0.272	0.272	0.272	0.322	0.322	0.322
50	0.289	0.336	0.240	0.237	0.237	0.336	0.336	0.336	0.371	0.371	0.371
51	0.294	0.296	0.245	0.296	0.296	0.296	0.296	0.296	0.342	0.342	0.342
52	0.299	0.259	0.250	0.259	0.259	0.259	0.259	0.360	0.315	0.315	0.390
53	0.304	0.320	0.255	0.226	0.226	0.320	0.320	0.320	0.361	0.361	0.361
54	0.308	0.282	0.260	0.282	0.282	0.282	0.282	0.282	0.333	0.333	0.333
55	0.313	0.343	0.265	0.247	0.247	0.343	0.343	0.343	0.379	0.379	0.379
56	0.318	0.305	0.270	0.305	0.305	0.305	0.305	0.305	0.351	0.351	0.351
57	0.323	0.366	0.275	0.269	0.269	0.366	0.366	0.366	0.396	0.396	0.396
58	0.328	0.327	0.280	0.327	0.327	0.327	0.327	0.327	0.369	0.369	0.369
59	0.333	0.291	0.286	0.291	0.291	0.291	0.388	0.388	0.343	0.412	0.412
60	0.338	0.349	0.291	0.257	0.257	0.349	0.349	0.349	0.385	0.385	0.385

$p^D(\theta)=\theta^T+0.1$ 、$\gamma=1-\beta=0.8$、$\lambda=1-\alpha=0.975$

5. 参考文献

C. R. De and Bruno Lecoutre. (2010). A new criterion for prior probabilities. 2010. hal-00911175.

Sambucini, V. (2008). A Bayesian predictive two-stage design for phase II clinical trials. *Statistics in Medicine*, *27*(8), 1199-1224.

SAS Institute Inc. (2025) “PROC POWER” SAS Help Center, Cary, NC, USA. SAS Institute Inc. https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/graphproc/n0rxu6ziykwmvyn1qatw3ytn36l1.htm. Accessed 22 Sept. 2025.

Simon, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials*, *10*(1), 1-10.

Teramukai, S., Daimon, T., & Zohar, S. (2012). A Bayesian predictive sample size selection design for single-arm exploratory clinical trials. *Statistics in Medicine*, *31*(30), 4243-4254.

Walley, R. J., Smith, C. L., Gale, J. D., & Woodward, P. (2015). Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study. *Pharmaceutical Statistics*, *14*(3), 205-215.

坂巻顕太郎, 兼 道雄, 大和田章一, 松浦健太郎, 柿爪智行, 高橋文博, 高沢翔, 萩原駿祐, 森田智視. (2020). ベイズ流決定理論を用いる臨床試験：効用とサンプルサイズ設計. 計量生物学, 2020-2021, 41 巻, 1 号, p. 55-91

手良向聡. (2017). なぜベイズを使わないのか!? -臨床試験デザインのために- . 金芳堂.

手良向聡. ベイズ流臨床試験における標本サイズ設定：2 つの事前分布を用いた推論上の性能に基づく接近法. 計量生物学, 2023-2024, 44 巻, 1 号, p. 35-51

矢田真城, 魚住龍史, 浜田知久馬. （2016）. 生存時間データに対するベイズ流例数設計. SAS ユーザー会 2016 論文集, 268-278.

用量反応試験におけるnormal dynamic linear model及びベイズ流階層モデルのシミュレーションによる動作特性の検討

○畑山知慶[1]、○松田裕也[1,2]、栗城正朗[1,3]、伊庭克拓[1,4]

（[1]日本製薬工業協会 医薬品評価委員会 データサイエンス部会、[2]中外製薬株式会社 デジタル戦略企画部、[3]協和キリン株式会社 開発本部 バイオメトリックス部、[4]大塚製薬株式会社 バイオメトリックス部）

Simulation study of operational characteristics of normal dynamic linear models and Bayesian hierarchical models in dose-response studies

Tomoyoshi Hatayama[1], Yuya Matsuda[1,2], Masaaki Kuriki[1,3], Katsuhiro Iba[1,4]

[1]Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association, [2]Digital Strategic Planning Department, Chugai Pharmaceutical Co., Ltd., [3]Biometrics Department, Development Division, Kyowa Kirin Co., Ltd., [4]Department of Biometrics, Otsuka Pharmaceutical Co., Ltd.

要旨

Smith et al. (2006) は用量反応曲線の推定と proof of concept（POC）の確認を目的とし、比較的多い実薬用量である 7 用量を設定した臨床試験において、用量反応曲線の形状に対する強い仮定を必要としない normal dynamic linear model（NDLM）によってベイズ流の解析を行った。また、無益性の中間解析によって無効な治療群への組み入れを停止する計画としていた。用量反応試験における他のベイズ流の解析事例として、用量反応に関する事前の仮定を置くことなく、観察された用量群間の治療効果の類似性に基づいて、用量群間で動的に情報を借用するベイズ流階層モデル（Bayesian hierarchical model : BHM）もある。本論文では、Smith et al. (2006) の試験設定を参考に、ベイズ流の非階層モデル（non-hierarchical model : NHM）を比較対象として、用量反応試験で NDLM 及び BHM を用いる場合の動作特性をシミュレーションで評価した。

キーワード：proof of concept, 用量反応試験, NDLM, ベイズ流階層モデル

1. はじめに

本稿では、用量反応試験におけるベイズ流試験デザインの活用例として、Smith et al. (2006) の試験設定を参考に、ベイズ流試験デザインのシミュレーションによる動作特性を検討した[1]。本試験は、帯状疱疹後神経痛（postherpetic neuralgia：PHN）に対する新規治療の用量反応関係の評価及び proof of concept（POC）の確認を目的とした二重盲検プラセボ対照ランダム化比較試験として実施された。用量反応関係の事前情報が不十分だったため、一般的な用量反応試験と比較して多い 7 用量（50, 100, 150, 200, 300, 450, 600 mg）を設定し、用量反応曲線の形状に強い仮定を必要としない normal dynamic linear model（NDLM）で推定を行った

[2]。また、無益性の中間解析を設定し、無効な治療群への組み入れを停止した。投与 4 週後の Pain score の週平均値（average daily pain score：ADPS）のベースラインからの変化量が主要評価項目として評価された。

他のベイズ流試験デザインとしては、ベイズ流階層モデル（Bayesian hierarchical model：BHM）があり、用量群間の治療効果の類似性に基づき動的に情報を借用することで推定精度を高める特徴がある。Smith et al. (2006) では NDLM が用いられたが、本稿では同試験設定で BHM の適用も検討した。用量反応試験における BHM の適用例としては、全身性エリテマトーデスの第 IIb 相試験がある[3]。

本稿では、Smith et al. (2006) の試験計画に基づき、用量反応曲線推定と POC 確認を目的とした試験を想定し、NDLM・BHM に加え、比較対象としてベイズ流非階層モデル（non-hierarchical model：NHM）も含めた動作特性をシミュレーションで評価した。シミュレーション設定は簡略化のため実際の試験とは一部異なるが、詳細は Smith et al. (2006) を参照されたい。

本稿は日本製薬工業協会医薬品評価委員会データサイエンス部会から公表を予定している報告書「ベイズ流試験デザインにおけるシミュレーションの利用法」（以下、製薬協報告書）[4]の 4 章を要約したものであり、詳細については製薬協報告書を参照されたい。また、本稿の付録では製薬協報告書から新たに NDLM 及び BHM を実行するための MCMC プロシジャの SAS コードを示した。

2. Normal dynamic linear model

NDLM は時系列データの分析のために開発された統計的方法である[2]。Smith et al (2006) では、NDLM を用量反応関係の推定に用いることで単調性を仮定しないモデリングを行った。NDLM のより詳細については West et al. (1997)を参照されたい[2]。

用量j $(j = 0\ (Placebo), 1, \dots, J)$における$k\left(k = 1, \dots, n_j\right)$番目の被験者のアウトカムを$Y_{jk}$とし、$\theta_j$を用量$Z_j$におけるアウトカムの平均とし、以下のようにモデル化する。

$$Y_{jk} = \theta_j + v_{jk}, \qquad v_{jk} \sim N(0, \sigma^2)$$

ここで、θ_j $(j = 1, \dots, J)$は以下のようにモデル化する。

$$\theta_j = \theta_{j-1} + \delta_{j-1} + \omega_j \qquad \omega_j \sim N(0, W\sigma^2)$$
$$\delta_j = \delta_{j-1} + \varepsilon_j \qquad \varepsilon_j \sim N(0, W\sigma^2)$$

ただし、$j = 0\ (Placebo)$の場合は以下の通りとする。

$$\theta_0, \delta_0 \sim N(0, 1000)$$

また，σ^2とWの事前分布は以下の通りとする。

$$\sigma^2 \sim InvGamma(0.001, 0.001)$$
$$W \sim uniform(0.001, 100)$$

3. ベイズ流階層モデル

BHM は、複数のレベルで記述された階層形式の統計モデルのことである。ここで用いる BHM では、実薬群の各用量の平均を表すパラメータが、特定の同一分布から生成されるという仮定を置く。用量$j(j = 0\ (Placebo), 1, \dots, J)$の$k\left(k = 1, \dots, n_j\right)$番目の被験者のアウトカムを$Y_{jk}$としたとき、$Y_{jk}$は平均$\theta_j$、分散$\sigma^2$の正規分布から生成され、更に実薬群の各用量の平均$\theta_j (j = 1, \dots, J)$が平均$\mu$、分散$\tau^2$の正規分布から生成される

という BHM によってモデル化する。これを式で表すとY_{jk}~$N(\theta_j, \sigma^2), \theta_j$~$N(\mu, \tau^2)$ $(j = 1, \ldots, J)$となる。また、プラセボ群の平均θ_0、アウトカムの分散σ^2、ハイパーパラメータであるμ及びτ^2の事前分布には無情報のフラット事前分布を用いる。

NHM は、パラメータの事前分布が単一レベルで構成される統計モデルのことである。ここで用いる NHM では、上記の BHM の実薬群の各用量の平均$\theta_j (j = 1, \ldots, J)$の事前分布である$N(\mu, \tau^2)$を、プラセボ群の平均$\theta_0$の事前分布と同じ無情報のフラット事前分布に変更したモデルである。

4. シミュレーションによる動作特性の検討

4.1 想定する試験計画

想定する試験計画を以下に示す。シミュレーションの目的は，中間解析の時期と回数、arm-dropping のための無益性及び有効性の判断基準に関連した動作特性の評価を行うことである。

試験標題	PHN 患者を対象とした新規治療の二重盲検プラセボ対照ランダム化比較第 II 相試験
目的	PHN に対する新規治療の用量反応関係の評価及び POC の確認
試験期間	6週間（ベースライン 1週間、二重盲検 4週間、フォローアップ 1週間）
投与群	プラセボ群、試験薬群 7用量（50, 100, 150, 200, 300, 450, 600 mg）
割付比	均等割付
主要評価項目	投与4週後の ADPS のベースラインからの変化量
最大被験者数	280例（各群35例）
効果判定	臨床的に意味のある有効性（プラセボ群に対して ADPS 変化量1.5の改善）が得られる事後確率に基づき、当該用量の効果有無を判断する。なお、下記θ_jは NDLM 又は BHM 並びに NHM に基づく。 中間解析： 最大被験者数のうち50%のアウトカムが利用可能な時点（情報量50%時）で1回の中間解析を行う。中間解析時点で以下の無効の基準に該当した用量jへの新規被験者の組み入れを停止し、全用量が該当する場合は試験を中止する。 • 無効：$P(\theta_j - \theta_0 < 1.5 \mid Data) > 0.8$ 最終解析： 最終解析時点で以下の有効の基準に該当した用量jについて有効な用量と判定する。 • 有効：$P(\theta_j - \theta_0 > 1.5 \mid Data) > 0.8$ 中間解析時点で無効と判定された用量に割り付けられた被験者のデータは、最終解析においてモデルの推定には用いられるが、当該用量は有効性の判定は行われない。

4.2 データ発生及びシナリオの概要

Smith et al. (2006)を参考に、用量Z_jにおけるk番目の被験者のアウトカムをY_{jk}とし、平均θ_j及び分散2.3^2の正規分布に従う疑似乱数を発生させる。シナリオ1（Null）は、実薬の効果が全くなく、プラセボと同等という

仮定であり、シナリオ2a~2d、3a~3d、4a~4d は、最大の効果をそれぞれ、1.45、2.20、2.9とし、用量反応モデルとして、linear (a)、Emax (b)、betaMod (c)、logistic (d)を仮定している（図1）。各シナリオの用量jの平均θ_jは製薬協報告書に示した。それぞれのシナリオにおいて中間解析あり、なしの場合を仮定したシナリオでデータを発生させた。各シナリオについて、シミュレーションの繰り返し数は1,000回とした。

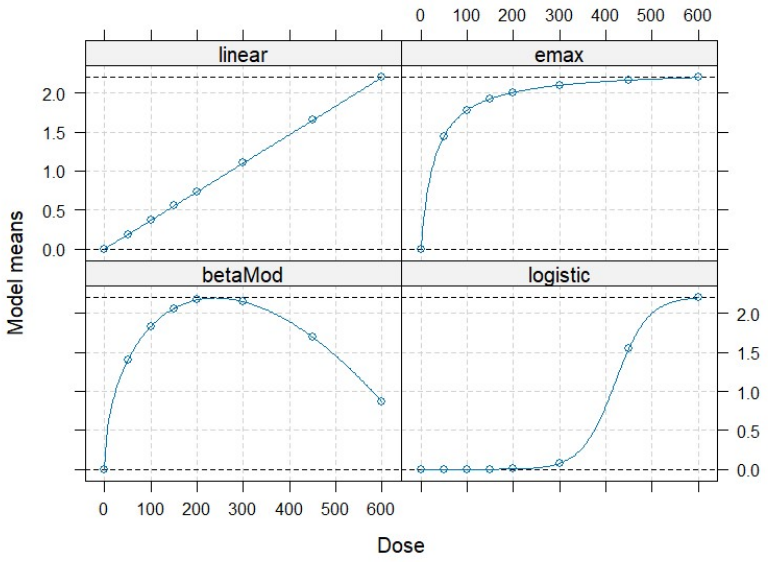


図 1 シナリオ 3a~3d の用反応曲線

4.3 評価指標

動作特性評価のため、下記の指標について検討した。

指標	内容
検出力	少なくとも1用量で効果ありと判定される確率を検出力とする。 全用量においてプラセボ群との差がない仮定（帰無仮説、シナリオ1）のもと少なくとも1つの用量で臨床的に意味のある有効性（プラセボ群との群間差1.5）を示すことを第一種の過誤とする。
被験者数	試験全体及び各用量群の平均組み入れ被験者数を確認する。
最大効果を持つ用量を正しく特定できる確率	最大効果を持つ用量である確率$P(\theta_j - \theta_0 = \max_j(\theta_j - \theta_0)\,\vert Data)(j = 1, \dots, J)$に基づいて、真の最大効果を持つ用量を正しく特定できる確率を確認する
治療効果の推定値、バイアス、RMSE	治療効果の推定値（プラセボ群との群間差）のバイアス及び root mean squared error（RMSE）を確認する。

4.4 シミュレーション結果の概要

シミュレーション結果のうち、シナリオ 1、3a～3d の概要を抜粋して示す。すべてのシミュレーション結果については，製薬協報告書を参照されたい[4]。なお，以下のシミュレーション結果から考察はすべてのシ

ミュレーション結果に基づいている。

表 1 に各モデルの検出力、平均組み入れ被験者数、最大効果を持つ用量を正しく特定できる確率を示す。本設定では、第一種の過誤確率（シナリオ 1 における検出力）は、いずれのモデルでも低く抑えられており、モデル間の違いは見られなかった。シナリオを通して，BHM の検出力は NHM より低く、NDLM は NHM と同程度か低い場合があった。NDLM は他の 2 つのモデルと比較して、平均組み入れ被験者数が小さかった。最大効果を持つ用量を正しく特定できる確率は、BHM と NHM で大きな違いはないが、NDLM は他の 2 つのモデルと比較して高かった。

表1 各モデルの検出力、平均組み入れ被験者数、最大効果を持つ用量を正しく特定できる確率

中間解析	シナリオ	検出力（少なくとも1用量で効果ありと判定される確率）			平均組み入れ被験者数			最大効果を持つ用量を正しく特定できる確率		
		NDLM	BHM	NHM	NDLM	BHM	NHM	NDLM	BHM	NHM
なし	1	0.00	0.00	0.00	280	280	280	NA	NA	NA
	3a	0.70	0.50	0.71	280	280	280	0.96	0.82	0.82
	3b	0.79	0.77	0.89	280	280	280	0.45	0.27	0.27
	3c	0.70	0.71	0.88	280	280	280	0.40	0.32	0.32
	3d	0.68	0.51	0.68	280	280	280	0.96	0.89	0.89
あり	1	0.00	0.00	0.00	151.0	152.1	166.3	NA	NA	NA
	3a	0.74	0.50	0.69	215.2	220.9	222.6	0.95	0.80	0.81
	3b	0.79	0.76	0.90	261.3	272.9	269.7	0.49	0.28	0.28
	3c	0.73	0.71	0.86	253.4	265.3	261.1	0.41	0.35	0.35
	3d	0.68	0.52	0.67	194.7	197.4	202.0	0.94	0.87	0.87

図 2 にシナリオ 1、3a～3d のバイアスと RMSE を示す。無情報事前分布を設定していたため、NHM ではバイアスが生じなかった。BHM では、用量反応がある場合（シナリオ 1 以外）に縮小推定により低用量では正のバイアス、高用量では負のバイアスが生じる傾向があった。NDLM では、プラセボを含め 1 つ前の用量に推定値が影響を受けるため、一概ではないが、負のバイアスが生じる傾向があった。NDLM と BHM の比較では、相対的に NDLM のバイアスが小さい傾向があった。BHM（高用量を除いて）及び NDLM では、NHM と比べて全体的に RMSE が小さかった。BHM よりも NDLM の方がより RMSE が小さい傾向があった。

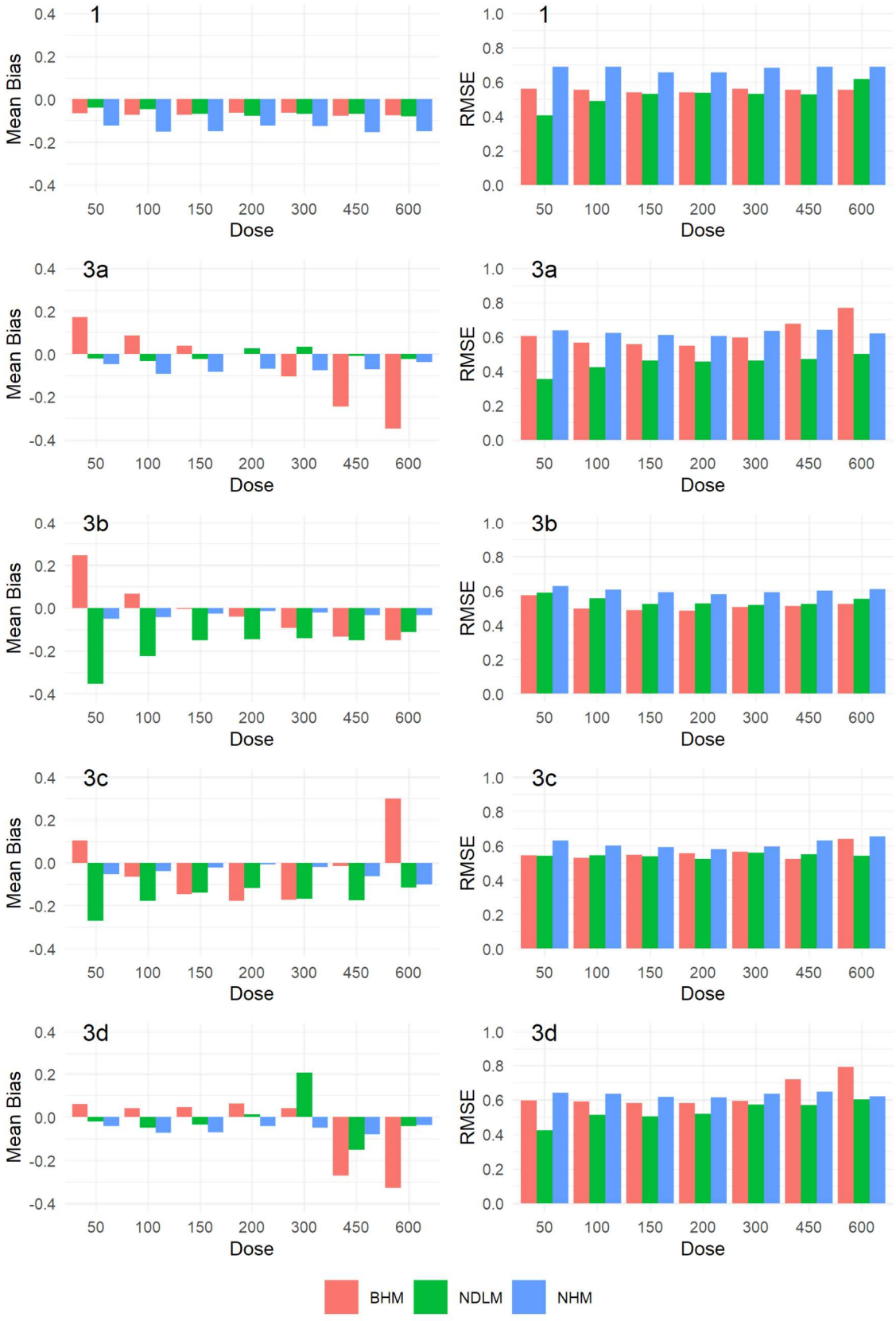


図 2 シナリオ 1、3a～3d のバイアスと RMSE

5. まとめ

本稿では、Smith et al. (2006) の試験設定を参考に、用量反応関係の事前情報が不十分な状況におけるベイズ流試験デザイン（NDLM、BHM）の適用の概略を説明するとともに、これらの動作特性をシミュレーションにより検討した。本稿で想定した用量反応試験では、用量反応関係の事前情報の不足から通常の用量反応試験より多い 7 用量の有効性を評価し、中間解析に基づく arm-dropping を適用した。ベイズ流試験デザインを用いることで、本稿で示した様な柔軟性のある試験デザインの計画を視野に入れることができる。

本シミュレーションで用いた R コードは製薬協報告書を[4]、NDLM、BHM の SAS コードは本稿の付録を参照されたい。本稿がベイズ流用量反応試験の計画及び SAS による解析の一助になれば幸いである。

参考文献

[1] Smith, M. K., Jones, I., Morris, M. F., Grieve, A. P., & Tan, K. (2006). Implementation of a Bayesian adaptive design in a proof of concept study. Pharmaceutical statistics, 5(1), 39–50.

[2] West, M., & Harrison, J. (2006). Bayesian forecasting and dynamic models. Springer Science & Business Media.

[3] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. （2024）. Complex Innovative Trial Design の現状と事例の整理. https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202404_CID.html

[4] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. （2025）. ベイズ流試験デザインにおけるシミュレーションの利用法.

付録

製薬協報告書では NDLM 及び BHM を Stan で実装した[4]。本付録では、NDLM 及び BHM を SAS の MCMC プロシジャで実装する方法を説明する。SAS コード中のデータセット EXAMPLE には数値型変数として，アウトカム Y 及び治療群 TRT（1：プラセボ群、2~8：実薬群の用量 1~7）を含んでいるとする。

NDLM を実行するための MCMC プロシジャの SAS コードを以下に示す。MCMC プロシジャで実装するにあたり，本文の NDLM のモデル式から以下のように表現の変更を行っている。

本文の表現		付録の表現
$Y_{jk} = \theta_j + v_{jk},$	$v_{jk} \sim N(0, \sigma^2)$	$Y_{jk} \sim N(\theta_j, \sigma^2)$
$\theta_j = \theta_{j-1} + \delta_{j-1} + \omega_j,$	$\omega_j \sim N(0, W\sigma^2)$	$\theta_j \sim N(\theta_{j-1} + \delta_{j-1}, W\sigma^2)$
$\delta_j = \delta_{j-1} + \varepsilon_j,$	$\varepsilon_j \sim N(0, W\sigma^2)$	$\delta_j \sim N(\delta_{j-1}, W\sigma^2)$

```
proc mcmc data=EXAMPLE seed=1234 nbi=1000 nmc=5000 thin=1 outpost=POST1;
  parms W;
  parms sigma;
  parms delta0;
  parms delta1;
```

```
parms delta2;
parms delta3;
parms delta4;
parms delta5;
parms delta6;
parms delta7;
parms theta0;
parms theta1;
parms theta2;
parms theta3;
parms theta4;
parms theta5;
parms theta6;
parms theta7;

array a_theta[8] theta0-theta7;

prior W ~ uniform(left=0.001, right=100);
prior sigma ~ igamma(shape=0.001, scale=0.001);
prior delta0 theta0 ~ normal(0, var=1000);
prior delta1 ~ normal(delta0, var=W*sigma);
prior delta2 ~ normal(delta1, var=W*sigma);
prior delta3 ~ normal(delta2, var=W*sigma);
prior delta4 ~ normal(delta3, var=W*sigma);
prior delta5 ~ normal(delta4, var=W*sigma);
prior delta6 ~ normal(delta5, var=W*sigma);
prior delta7 ~ normal(delta6, var=W*sigma);
prior theta1 ~ normal(theta0 + delta1, var=W*sigma);
prior theta2 ~ normal(theta1 + delta2, var=W*sigma);
prior theta3 ~ normal(theta2 + delta3, var=W*sigma);
prior theta4 ~ normal(theta3 + delta4, var=W*sigma);
prior theta5 ~ normal(theta4 + delta5, var=W*sigma);
prior theta6 ~ normal(theta5 + delta6, var=W*sigma);
prior theta7 ~ normal(theta6 + delta7, var=W*sigma);

do j=1 to 8;
  if TRT=j then theta=a_theta[j];
end;
```

```
  model Y ~ normal(theta, var=sigma);
run;
```

BHM を実行するための MCMC プロシジャの SAS コードを以下に示す。本文ではアウトカムの分散σ^2、ハイパーパラメータであるμ及びτ^2の事前分布は，Stan で事前分布を指定しなかった場合に用いられる無情報のフラット事前分布としていた。MCMC プロシジャでは，共役事前分布が指定された場合、共役事前分布を利用して効率的にサンプリングを行うことができることから，μの事前分布は$N(0,1000)$、σ^2及びτ^2の事前分布は$InvGamma(0.001, 0.001)$に変更している。

```
proc mcmc data=EXAMPLE seed=1234 nbi=1000 nmc=5000 thin=1 outpost=POST2;
  parms mu;
  parms tau;
  parms sigma;
  parms theta0;
  parms theta1;
  parms theta2;
  parms theta3;
  parms theta4;
  parms theta5;
  parms theta6;
  parms theta7;

  array a_theta[8] theta0-theta7;

  prior mu theta0 ~ normal(0, var=1000);
  prior tau sigma ~ igamma(shape=0.001, scale=0.001);
  prior theta1-theta7 ~ normal(mu, var=tau);

  do j=1 to 8;
    if TRT=j then theta=a_theta[j];
  end;

  model Y ~ normal(theta, var=sigma);
run;
```

～実例と演習で学ぶ～

Transpose Procedure / Array Statement

○牛尾　英孝

（株式会社ピープロ・ジャパン）

Learning through Practical Examples and Exercises: Transpose Procedure / Array Statement

Hidetaka Ushio

Ppro-Japan Co.,Ltd.

要旨

臨床試験におけるデータ解析において，横方向に収集された繰り返し測定データや複数項目データを縦方向に変換する処理は頻繁に求められ，SAS では Proc Transpose や Array Statement を用いてデータの再構築を行うことができる．本稿では，実業務で遭遇する典型的な事例を題材に，Proc transpose や Array Statement による柔軟なデータの転置方法について，PDV（Program Data Vector）の働きにも着目し，内部処理の流れを具体例と演習を通して視覚的なイメージと共に解説する．

キーワード：Proc Transpose，Array Statement，PDV（Program Data Vector）

1. 緒言

臨床試験では，観測値が繰り返し測定される場合や多数の項目を同時に扱う場合が多く，そのデータは必ずしも統計解析に適した形式で得られるとは限らない．特に横方向に配置された測定値を縦方向に変換する処理は，解析を進めるうえで避けて通れないステップであり，効率的かつ正確な実装が求められる．SAS はこのようなニーズに応えるため，Proc Transpose や Array Statement といった強力な機能を提供している．Proc Transpose は自動的にデータの縦横変換を行う手段として広く利用される一方，Array Statement は複数の変数を一括で操作する柔軟な枠組みを与える．両者はいずれもデータハンドリングにおいて欠かせない存在であるが，その内部挙動を十分に理解していないと，意図しない結果が生じる原因となる．本稿では，これらの基本原理を整理するとともに，視覚的なイメージ及び SAS 内部での PDV（Program Data Vector）の役割に焦点を当てて，実例と演習を通じて理解を深めることを目的とする．これにより，初学者がデータハンドリングに直結する実践的なノウハウを体系的に習得できることを期待する．

2. Transpose Procedure の基本構文と使用例

本 Procedure では，入力データセット（input-data-set）と出力データセット（output-data-set）を指定し，さらに必要に応じてオプションを付与することができる．基本的な記述は以下の通りである．

```
PROC TRANSPOSE   DATA=[input-data-set]   OUT=[output-data-set]    [OPTION];
    BY [variable];          /* グループごとに転置するkey変数を指定 */
    VAR [variable];             /* 転置する変数を指定 */
    ID [variable];           /* 転置後の列名を指定 */
RUN;
```

ここで，BY Statement は指定した変数ごとに独立して転置を行うために用いられ，ID Statement は転置後の列名を与える役割を持つ．VAR Statement を指定した場合には，その変数のみが転置対象となり，指定がない場合には全ての変数が対象となる．このように，Proc Transpose は少数の指定で多様なデータ変換を柔軟に実現できる点に特徴がある．

VIEWTABLE: Work.Lb1

	USUBJID	HB	ALT	AST
1	001	13.5	25	22
2	002	13.8	23	20
3	003	13.0	40	35

図 2-1：Lb1

図 2-1 のように横方向に配置されている変数 HB，ALT，AST を縦方向に再構築するため，図 2-1：Lb1 に対して Proc Transpose を実行する．USUBJID ごとに転置処理を行うため，BY Statement には USUBJID，VAR Statement には転置の対象となる HB，ALT，AST をそれぞれ指定する．なお，BY Statement に指定する変数（USUBJID）については事前に sort 処理を行っておく必要がある．

```
proc sort data=Lb1;
    by USUBJID;
run;

proc transpose   data=Lb1              /* 入力データセットを指定 */
                 out=Lb1_trans         /* 出力データセットを指定 */
                 name=PARAMCD;         /* _NAME_をPARAMCD に変更するオプション */
      by USUBJID;                      /* グループごとに転置するkey変数 */
      var HB ALT AST;                  /* 転置する変数を指定 */
run;

data Lb1_trans;
    set Lb1_trans(rename=(col1=LBORRES));     /* COL1をLBORRES にリネーム */
    label PARAMCD = PARAMCD;                  /* ラベルを設定を変更 */
run;
```

VIEWTABLE: Work.Lb1_trans

	USUBJID	PARAMCD	LBORRES
1	001	HB	13.5
2	001	ALT	25
3	001	AST	22
4	002	HB	13.8
5	002	ALT	23
6	002	AST	20
7	003	HB	13.0
8	003	ALT	40
9	003	AST	35

図 2-2：Lb1_trans

図 2-2 を再び図 2-1 へ戻す際も Proc Transpose を実行することで可能である．

```
proc transpose data=Lb1_trans
               out=Lb1_trans_back(drop=_NAME_);      /* _NAME_ は不要なので削除 */
    by USUBJID;                                      /* USUBJIDごとに横方向へ戻す */
    var LBORRES;                                     /* 値を横方向にする対象変数を指定 */
    id PARAMCD;                                      /* 縦方向に展開された項目名を列名に戻す */
run;
```

3. Proc Transpose の視覚的なイメージ

図 2-1：Lb1 に対し Proc Transpose を適用した際のデータ構造の変換イメージを視覚的に表したものを図 3-1 に示した．横方向に展開されていた複数の測定項目が縦方向に整列される．

VIEWTABLE: Work.Lb1

	USUBJID	HB	ALT	AST
1	001	13.5	25	22
2	002	13.8	23	20
3	003	13.0	40	35

Step 1

	Lb1_Trans
	USUBJID
HB	001
ALT	001
AST	001
	002
	002
	002
	003
	003
	003

Step 2

Lb1_Trans	
USUBJID	_NAME_
001	HB
001	ALT
001	AST
002	HB
002	ALT
002	AST
003	HB
003	ALT
003	AST

Step 3

Lb1_Trans		
USUBJID	_NAME_	COL1
001	HB	13.5
001	ALT	25
001	AST	22
002	HB	13.8
002	ALT	23
002	AST	20
003	HB	13.0
003	ALT	40
003	AST	35

図 3-1：Lb1 に proc Transpose を実行した際の視覚的なイメージ

データ構造の再構築の過程を 3 つのステップで解説する．

【Step1】では BY Statement に指定する変数（USUBJID）は出力データセット Lb1_Trans に残る変数となる．VAR Statement に指定した転置の対象となる変数は HB，ALT，AST の 3 変数であるため，BY Statement に指定した USUBJID は出力データセット（Lb_trans）に 3 行ずつ繰り返し出力される．

【Step 2】では VAR Statement に指定した転置の対象となる変数（Hb，ALT，AST）がデフォルトでは _NAME_という変数に BY Statement に指定した変数（USUBJID）ごとに格納される．

【Step 3】では_ NAME _へ格納された HB，ALT，AST に対応する変数値がそれぞれデフォルトでは COL1 という変数名で縦方向へ積まれる．_NAME _及び COL1 を Rename Option を用いて後続の解析に適した変数名に変更することも可能である．

4. 演習問題～Proc transpose～

ここまでの内容で演習問題を 1 つ紹介する．図 4-1：Vs データセットに対して Proc Transpose を適応して図 4-2：Vs_trans の構造へ変換する．

VIEWTABLE: Work.Vs

	USUBJID	VISITNUM	SYSBP	SYSBP_UNIT	DIABP	DIABP_UNIT	PULSE	PULSE_UNIT
1	001	1	123	mmHg	80	mmHg	72	BEATS/MIN
2	001	2	119	mmHg	78	mmHg	71	BEATS/MIN
3	002	1	124	mmHg	79	mmHg	75	BEATS/MIN
4	002	2	118	mmHg	81	mmHg	73	BEATS/MIN

図 4-1：Vs

VIEWTABLE: Work.Vs_trans

	USUBJID	VISITNUM	PARAMCD	VSORRES	VSORRESU
1	001	1	SYSBP	123	mmHg
2	001	1	DIABP	80	mmHg
3	001	1	PULSE	72	BEATS/MIN
4	001	2	SYSBP	119	mmHg
5	001	2	DIABP	78	mmHg
6	001	2	PULSE	71	BEATS/MIN
7	002	1	SYSBP	124	mmHg
8	002	1	DIABP	79	mmHg
9	002	1	PULSE	75	BEATS/MIN
10	002	2	SYSBP	118	mmHg
11	002	2	DIABP	81	mmHg
12	002	2	PULSE	73	BEATS/MIN

図 4-2：Vs_trans

ここでは図 4-1 に示すバイタルサインデータ（SYSBP，DIABP，PULSE）と，それぞれに対応する単位（SYSBP_UNIT，DIABP_UNIT，PULSE_UNIT）が横方向に収集されたデータを題材とする．図 4-1：Vs に対して，Proc Transpose を実行し，収集データを PARAMCD と VSORRES に整理する．さらに単位も別途縦方向に変換し VSORRESU と対応付けたデータ（図 4-2：Vs_trans）を再構築するコードについて検討することを演習問題とした．演習問題のコード例の概要を下記に示す．収集値（SYSBP，DIABP，PULSE）と単位（_UNIT）が別列で対になっているため，変数値及び単位それぞれに対して Proc Transpose を実行する．その後，USUBJID，VISITNUM，SORT を key にして結合することで目的の出力データセット（Vs_trans）を生成することができる．

```
/* ① 値及び単位をそれぞれ縦方向へ変換 */
proc transpose data=Vs out=Vs_val(rename=(col1=VSORRES _NAME_=PARAMCD));
    by USUBJID VISITNUM;
    var SYSBP DIABP PLUSE;
run;
proc transpose data=Vs out=Vs_unit(rename=(col1=VSORRESU _NAME_=PARAMCD_U));
    by USUBJID VISITNUM;
    var SYSBP_UNIT DIABP_UNIT PLUSE_UNIT;
run;
/* ② SORT keyを作成　Vs_unitを対象にPARAMCD_Uからも同様にSORTT keyを作成する */
data Vs_val;
    set Vs_val;
    if PARAMCD="SYSBP" then SORT=1;
    else if PARAMCD="DIABP" then SORT=2;
    else SORT=3;
run;
proc sort data=Vs_val;
    by USUBJID VISITNUM SORT;
run;
/* ③ 値と単位をmerge */
data Vs_trans;
    merge Vs_val Vs_unit;
    by USUBJID VISITNUM SORT;
    label PARAMCD="PARAMCD";
    drop PARAMCD_U SORT;
run;
```

5. Array Statement の基本構文と使用例

本 Statement では配列名（Array name），配列の要素数（Number of elements），配列の要素（Array elements）を定義する．基本的な記述は以下の通りである．また，配列の要素について対応関係を表 5-1 に示した．

```
Array Array name[Number of elements] Array elements; /* 配列名[要素数] 配列の要素 を定義 */
/* 具体例 */
Array QS[3] Q1-Q3;    /*[具体例] 配列名：QS，要素数：3，配列の要素：Q1，Q2，Q3*/
```

表 5-1：配列の対応関係

QS[1]	QS[2]	QS[3]	QS[*]
Q1	Q2	Q3	Q1, Q2, Q3

VIEWTABLE: Work.Qs

	USUBJID	Q1	Q2	Q3
1	001	1	3	5
2	002	2	4	6
3	003	3	6	9

図 5-1：Qs

図 5-1 において，各 USUBJID における Q1～Q3 の平均スコアを求めるコード例を下記に示した．MeanScore = (Q1 + Q2 + Q3) / 3 として算出することもできるが，配列を用いることで変数をまとめて処理することが可能となる．

```
data Qs_array;
   set Qs;
   array QS[3] Q1-Q3;   /* 配列名，要素数，要素を定義 */
        /* 配列を用いて平均を計算 */
   MeanScore = mean(of QS[*]);
run;
```

6. Array Statement の内部挙動（Program Data Vector）

図 2-1 で示した Lb1 に対して Array Statement を用いて図 6-1 のような構造に再構築することを考える．ここでは下記のコード例について，PDV（Program Data Vector）に触れながら内部挙動を視覚的なイメージとともに解説する．

VIEWTABLE: Work.Lb1_array

	USUBJID	PARAMN	LBORRES
1	ABC-001	1	13.5
2	ABC-001	2	25
3	ABC-001	3	22
4	ABC-002	1	13.8
5	ABC-002	2	23
6	ABC-002	3	20
7	ABC-003	1	13
8	ABC-003	2	40
9	ABC-003	3	35

図 6-1：Lb1_array

```
data Lb_array;
      set Lb1;
        array _lborres[3] HB ALT AST;
           do i = 1 to 3;
              PARAMN = i;
              LBORRES = _lborres(i);
              output;
           end;
      drop i HB ALT AST;
run;
```

コンパイルフェイズにおいて，PDV（_N_及び_ERROR_は省略）が作成される．ここで，USUBJID，HB，ALT，AST は入力データセット（Lb1）由来の変数，i，PARAMN，LBORRES は割り当て Statement にて作成された新規変数である．HB，ALT，AST 及び i は Drop Statement により削除フラグ[Drop]が付与され，出力データセット（Lb1_array）には出力されない．

PDV（Program Data Vector）						
USUBJID	HB [Drop]	ALT [Drop]	AST [Drop]	i [Drop]	PARAMN	LBORRES

図 6-2：コンパイルフェイズ終了時点の PDV のイメージ

実行フェイズにおいて，ここでは Do loop 開始時に PDV の全変数が初期化され，続いて入力データセット（Lb1）の 1 行目が PDV へ上書きされる．Do loop 内の_lborres(1)，_lborres(2)，_lborres(3)はそれぞれ，HB，ALT，AST であるから，i=1 の場合，PARAMN は 1，LBORRES は HB の値が PDV へコピーされ，Output Statement により PDV から出力データセット（Lb1_array）へ 1 行分が出力される．

PDV（Program Data Vector）						
USUBJID	HB [Drop]	ALT [Drop]	AST [Drop]	i [Drop]	PARAMN	LBORRES
001	13.5	25	22	1	1	13.5

図 6-3-1：入力データセット 1 行目参照時の i=1 における出力前 PDV

Lb1_array		
USUBJID	PARAMN	LBORRES
001	1	13.5

図 6-3-2：入力データセット 1 行目参照時の i=1 における出力後データセット

この段階では End Statement へは到達せず，2 回目（i=2）の Do loop 処理が実行される．i，PARAMN，LBORRES は PDV 上で上書きされ，同様に Output Statement により PDV から出力データセット（Lb1_array）へ 1 行分が出力される．

PDV（Program Data Vector）						
USUBJID	HB [Drop]	ALT [Drop]	AST [Drop]	i [Drop]	PARAMN	LBORRES
001	13.5	25	22	2	2	25

図 6-4-1：入力データセット 1 行目参照時の i=2 における出力前 PDV

Lb1_array		
USUBJID	PARAMN	LBORRES
001	1	13.5
001	2	25

図 6-4-2：入力データセット 1 行目参照時の i=2 における出力後データセット

同様に 3 回目（i=3）の Do loop 処理にて，i，PARAMN，LBORRES が PDV 上で上書きされ，出力データセット（Lb1_array）へ 1 行分が出力される．

PDV（Program Data Vector）						
USUBJID	HB [Drop]	ALT [Drop]	AST [Drop]	i [Drop]	PARAMN	LBORRES
001	13.5	25	22	3	3	22

図 6-5-1：入力データセット 1 行目参照時の i=3 における出力前 PDV

Lb1_array		
USUBJID	PARAMN	LBORRES
001	1	13.5
001	2	25
001	3	22

図 6-5-2：入力データセット 1 行目参照時の i=3 における出力後データセット

i=4 では Do loop の条件（i=1 to 3）を満たさないので，これ以降では Do loop は実行されず，End Statement へ到達する．その後，Data Step の先頭へ処理が戻り，PDV の割り当て Statement が再初期化され，続いて入力データセット（Lb1）の 2 行目が PDV へ上書きされる．

PDV（Program Data Vector）						
USUBJID	HB [Drop]	ALT [Drop]	AST [Drop]	i [Drop]	PARAMN	LBORRES
002	13.8	23	20	.	.	

図 6-6：入力データセット 2 行目参照時の PDV

以降，同様に Do loop が実行され Output Statement にて出力データセットへ 1 行分出力され，入力データセット（Lb1）の最終行が処理されるまで繰り返される．

PDV（Program Data Vector）						
USUBJID	HB [Drop]	ALT [Drop]	AST [Drop]	i [Drop]	PARAMN	LBORRES
002	13.8	23	20	1	1	13.8

図 6-7-1：入力データセット 2 行目参照時の i=1 における出力前 PDV

Lb1_array		
USUBJID	PARAMN	LBORRES
001	1	13.5
001	2	25
001	3	22
002	1	13.8

図 6-7-2：入力データセット 2 行目参照時の i=1 における出力後データセット

7. 演習問題～Array Statement～

最後に演習問題をもう 1 つ紹介する．USUBJID ごとに測定値を週別に収集したもの（図 7-1：Lb2）を題材とし，1 症例（USUBJID）ごと，1 検査時点（VISIT：Week1 又は Week2）ごとに整理し，各 VISIT 内では PARAMCD（HB，ALT，AST）ごとに測定値（LBORRES）を収集した構造（図 7-2：Lb2_array）へ変換するコードを検討するものを演習問題とした．

VIEWTABLE: Work.Lb2

	USUBJID	HB_week1	ALT_week1	AST_week1	HB_week2	ALT_week2	AST_week2
1	001	13.5	25	22	12.2	23	22
2	002	13.8	23	20	12.8	21	22
3	003	13.0	40	35	12.7	38	33

図 7-1：Lb2

VIEWTABLE: Work.Lb2_array

	USUBJID	VISIT	PARAMN	PARAMCD	LBORRES
1	001	Week1	1	HB	13.5
2	001	Week1	2	ALT	25
3	001	Week1	3	AST	22
4	001	Week2	1	HB	12.2
5	001	Week2	2	ALT	23
6	001	Week2	3	AST	22
7	002	Week1	1	HB	13.8
8	002	Week1	2	ALT	23
9	002	Week1	3	AST	20
10	002	Week2	1	HB	12.8
11	002	Week2	2	ALT	21
12	002	Week2	3	AST	22
13	003	Week1	1	HB	13.0
14	003	Week1	2	ALT	40
15	003	Week1	3	AST	35
16	003	Week2	1	HB	12.7
17	003	Week2	2	ALT	38
18	003	Week2	3	AST	33

図 7-2：Lb2_array

Proc transposeを用いたコードの概要を下記に示した。

```
/* HB，ALT，AST を縦方向へ */
proc transpose data=Lb2 out=Lb2_trans(rename=(COL1=LBORRES));
    by USUBJID;
    var HB_week1 HB_week2 ALT_week1 ALT_week2 AST_week1 AST_week2;
run;
data Lb2_trans_1;
    length USUBJID $8.VISIT $8. PARAMN 8. PARAMCD $20. LBORRES $200.; /* データセット出力順 */
    set Lb2_trans;
    if _NAME_ in ('HB_week1','ALT_week1','AST_week1') then VISIT='Week1';   /* Week2についても同様*/
    if _NAME_ in ('HB_week1','HB_week2') then do;
        PARAMN=1; PARAMCD='HB';
    end;
    /* [ALT，AST についても同様の処理を行う]*/
    keep USUBJID VISIT PARAMN PARAMCD LBORRES;
run;
proc sort data= Lb2_trans_1; out= Lb2_array;
    by USUBJID VISIT PARAMN;
run;
```

続いて Array Statement を用いた演習問題のコード例を下記に示す.

```
data Lb2_array;
    set Lb2;
    /* 各週ごとの配列 */
    array week1[3] HB_week1 ALT_week1 AST_week1;
    array week2[3] HB_week2 ALT_week2 AST_week2;
    /* 項目名（検査コード）を保持する一時配列 */
    array paramcds[3] $20 _temporary_ ('HB' 'ALT' 'AST');
    length VISIT $8. PARAMN 8. PARAMCD $20.;   /* 文字変数の長さを指定 */
    do visitnum = 1 to 2;                /* 検査時点のloop */
        do PARAMN = 1 to 3;       /* 検査項目のループ */
            if visitnum = 1 then do;
                LBORRES = week1(PARAMN);
                VISIT = "Week1";
            end;
            if visitnum = 2 then do;
                LBORRES = week2(PARAMN);
                VISIT = "Week2";
            end;
            PARAMCD = paramcds(PARAMN);
            output;
        end;
    end;
  keep USUBJID VISIT PARAMN PARAMCD LBORRES;
run;
```

8. まとめ

データ解析においては，横方向から縦方向，あるいはその逆といった構造の変換を行う場面は少なくない．その際に用いられる代表的な手法が Proc Transpose 及び Array Statement である．いずれも使用頻度が高く，効率的なデータハンドリングに不可欠な手法といえる．これらを適切に活用するためには，PDV の挙動を理解することが重要であり，PDV の仕組みを把握することで効率的なプログラミングが可能となる．

9. 参考文献

1. matui -SAS 忘備録 DO ループ入門-
   https://sas-boubi.blogspot.com/2016/01/do-loop.html（最終アクセス 2025/09/01）
2. 高浪 洋平/舟尾 暢男.『改訂版 統計解析ソフト「SAS」』カットシステム（2015）
3. SAS®認定スペシャリスト Base Programming Using SAS® 9.4 完全ガイド
4. SAS®認定スペシャリスト Advanced Programming Using SAS® 9.4 完全ガイド

がん第II相バスケット試験での
ベイズ流階層モデルの適用と動作特性評価

○豊泉　樹一郎[1]、松田　裕也[1, 2]
（[1]日本製薬工業協会 医薬品評価委員会 データサイエンス部会、[2]中外製薬株式会社）

Application of Bayesian Hierarchical Model and Operating Characteristics in a Phase II Basket Trial in Cancer

Kiichiro Toyoizumi[1], Yuya Matsuda[1, 2]
[1] Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association.,
[2] Chugai Pharmaceutical Co., Ltd.

要旨

特定の遺伝子異常が発現した悪性腫瘍を標的とした新薬の開発においては、その遺伝子異常が発現した複数の固形がんに対する有効性をまとめて評価するバスケット試験が実施される場合がある。しかし、各がん種での有効性評価に通常の頻度論流の手法を用いた場合、各がん種での例数が限られ、個々のがん種での有効性の判断が難しくなる可能性がある。この解決法として共通の遺伝子異常を持っていることから、各がん種における奏効率が類似している可能性を考慮できるベイズ流階層モデルを使用し、がん種間で情報借用を行い、より精度の高い推定を行うことが考えられる。本発表では、Pant et al. (2023)で報告された臨床試験を参考とし、がん第 II 相バスケット試験で中間解析及び最終解析で各がん種における無益性及び有効性を評価にベイズ流階層モデルを適用した場合の動作特性をシミュレーションにより評価した。シミュレーションを実施する際のコードのポイント、動作特性で評価するべきポイント及びその結果について発表する。

キーワード：Bayesian Hierarchical Model、Basket Trial

---本文開始点---

1　はじめに

特定の遺伝子異常が発現した悪性腫瘍を標的とした新薬の開発においては、その遺伝子異常が発現した複数の固形がんに対する有効性をまとめて評価するバスケット試験が実施される場合がある。しかし、中間解析及び最終解析時の各がん種での有効性評価に通常の頻度論流の手法を用いた場合、各がん種での症例数が限られるため、各がん種での有効性の判断が難しくなる可能性がある。この解決法として共通の遺伝子異常を持っていることから、各がん種における奏効率が類似している可能性を考慮できるベイズ流階層モデルの適用が考えられる。ベイズ流階層モデルを用いることで、がん種間で情報借用を行い、各がん種でのより精度の高い推定を行うことが考えられる。

本発表では、Pant et al. (2023)[1] を参考にして、共通の遺伝子異常を発現した複数の固形がんに対する臨床試験を計画する状況を想定し、シミュレーションによりベイズ流階層モデルに基づく中間解析の動作特性を評価した。一部、シミュレーションを簡便に実行するために、Pant et al. (2023)[1] とは異なる試験デザインを仮定しているため、実際のデザインは原著を参照されたい。

2　想定する臨床試験

2.1　試験デザイン

本発表で想定する臨床試験は、共通のバイオマーカーを発現した複数がん種の進行がんに対する新規治療の有効性の評価を目的とした単群第 II 相試験である。主要評価項目は奏効率とし、帰無仮説下での奏効率は 15%、対立仮説下での奏効率は 35%である。被験者数はすべてのがん種を合わせて 100 例である。この症例数を設定することで、頻度論流の解析計画を立てた際に、有意水準片側 0.025、新規薬剤での奏効率を 35%とした際に、単群試験で閾値 15%に対する優越性を 99.7%の確率で検出できる。一般的な臨床試験の検出力より高い値であるが、今回は各がん種での活性の有無も検討するため、100 例と設定した。

2.2　中間解析

本試験では、共通のバイオマーカーを発現した患者のみを登録するが、すべてのがん種で治療が有効かどうかは臨床試験のデータから確認する必要がある。無効ながん種が含まれていた場合、試験の成功確率が低下する。そのため、無効ながん種が含まれていないかを確認するために、中間解析を 50 例の患者で奏効データが確定した時点で無益性解析を 1 回実施する。中間解析では奏効率の交換可能性を仮定したベイズ流階層モデルを用いる。ベイズ流階層モデルの詳細は後述する。中間解析で各がん種における奏効率が 25%以上である事後確率を算出し、40%未満である場合には当該がん種で無益性の判断を行う。

- 試験継続：$P(\pi_j > 0.25|Data) > 0.4$
- 無効中止：$P(\pi_j > 0.25|Data) \leq 0.4$

2.3　主要解析

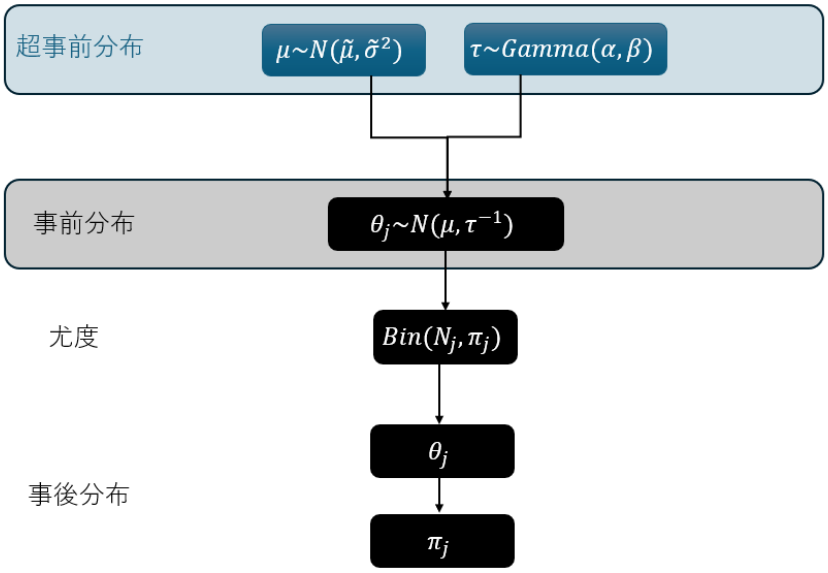
主要解析では中間解析で無効中止とならなかったがん種を併合して頻度論流の解析を行い、奏効率の 95%信頼区間の下限が閾値(15%)を超えているかどうか判定する。また、各がん種での有効無効の判断をするために、ベイズ流階層モデルを適用し、奏効率が 25%以上である事後確率が 80%以上である場合、当該がん種での有効性が認められたとする。

3　適用するベイズ流階層モデル

図1に本試験で適用したベイズ流階層モデルの構造を示す。がん種jにおける奏効率をπ_jとし、$\theta_j = logit(\pi_j)$とした際に、θ_jは、平均μ、分散τ^{-1}の正規分布に従っているとし、超事前分布として$\mu \sim N(-2.19, 2^2)$及び$\tau \sim Gamma(\alpha, \beta)$を仮定した。$\mu$の事前平均-2.19 は奏効率 10%の logit 変換後の値に対応し、μの事前分布の範囲は奏効率 0%から 99.7%に相当する。ガンマ分布のパラメータα及びβについては、本試験では事前分布が動作特性に及ぼす影響を評価するために以下の二つの超事前分布で比較した。Prior 2 の方ががん種間の奏効率に関する情報借用の程度が大きくなる。

Prior 1: $\tau \sim Gamma(2,20)$、Prior 2: $\tau \sim Gamma(2,2)$

図 1　ベイズ流階層モデルの構造



具体的にベイズ流階層モデルを適用した場合の結果を表 1 に示す。事後平均 1 は Prior 1 を、事後平均 2 は Prior 2 を用いた場合の結果である。がん種間の奏効率に関する情報借用の程度がより大きくなる Prior 2 を用いた場合の方が、事前分布及び他のがん種の奏効率により引っ張られていること、95%信用区間の区間幅が狭く、各がん種間での奏効率の情報を借用し、より効率の良い推定が行うことができていることがわかる。

表 1 ベイズ流階層モデルを適用した場合の結果例

	被験者数	奏効数	奏効率	事後平均 1 (95%信用区間)	事後平均 2 (95%信用区間)
がん種 1	30	10	33.3%	33% (18%, 50%)	31% (17%, 47%)
がん種 2	30	9	30%	29% (15%, 47%)	28% (15%, 44%)
がん種 3	30	12	40%	39% (23%, 57%)	36% (22%, 53%)
がん種 4	5	1	20%	19% (1%, 56%)	19% (3%, 49%)
がん種 5	5	0	0%	6% (0%, 27%)	12% (1%, 34%)

今回適用するベイズ流階層モデルは R の trialr パッケージに含まれる stan_hierarchical_response_thall 関数を用いることで二項分布の下でのベイズ流階層モデルの適用が可能となる。上記の具体例でのコードとその結果を示す。prob_response が各がん種で推定された奏効率及びその 95%信用区間を示している。

```
library(trialr)

#Prior 1 の場合
fith <- stan_hierarchical_response_thall(
   group_responses = c(10, 9, 12, 1, 0), #各がん種での奏効例数
   group_sizes = c(30, 30, 30, 5, 5),      #各がん種での被験者数
   mu_mean = -2.19,                          #μの超事前分布の平均
   mu_sd = 2,                    #μの超事前分布の標準偏差
   tau_alpha = 2,                          #τの超事前分布のパラメータ
   tau_beta = 20)

> fith
Inference for Stan model: ThallHierarchicalBinary.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                    mean se_mean     sd     2.5%      25%      50%      75%    97.5% n_eff Rhat
mu                 -1.71       0.02 1.11   -4.06   -2.40   -1.67   -0.96     0.38   3724       1
sigma2              8.36       0.12 6.16     2.68     4.79     6.74     9.86   24.15   2702       1
rho[1]             -0.74       0.01 0.39   -1.53   -1.00   -0.73   -0.47     0.00   5282       1
rho[2]             -0.89       0.01 0.42   -1.77   -1.17   -0.88   -0.61   -0.12   6331       1
rho[3]             -0.44       0.00 0.36   -1.19   -0.69   -0.44   -0.20     0.25   6198       1
rho[4]             -1.71       0.02 1.21   -4.44   -2.43   -1.60   -0.87     0.40   4815       1
rho[5]             -3.90       0.05 2.12   -9.05   -5.02   -3.58   -2.41   -0.74   1905       1
sigma               2.76       0.02 0.85     1.64     2.19     2.60     3.14     4.91   2865       1
prob_response[1]    0.33       0.00 0.08     0.18     0.27     0.33     0.39     0.50   5207       1
prob_response[2]    0.30       0.00 0.08     0.15     0.24     0.29     0.35     0.47   6495       1
prob_response[3]    0.39       0.00 0.08     0.23     0.33     0.39     0.45     0.56   6041       1
prob_response[4]    0.21       0.00 0.16     0.01     0.08     0.17     0.29     0.60   5995       1
prob_response[5]    0.06       0.00 0.09     0.00     0.01     0.03     0.08     0.32   4673       1
lp__              -75.91       0.05 2.09 -81.11 -77.04 -75.53 -74.37 -73.02   1509       1
```

4　シミュレーションの設定

シミュレーションにおける各がん種における登録例数、奏効率のシナリオを表 2 のように設定した。各がん種での登録割合について、どのがん種も同程度の場合（シナリオ 1-1~3）、特定のがん種に偏っている場合（シナリオ 2-1~3）、いくつかの希少ながん種がある場合（シナリオ 3-1~3）を仮定した。また

奏効率については各がん種で活性がある場合、一部のがん種でのみ活性がある場合（シナリオ 2-2 では最も登録例数が多いがん種でのみ活性がある場合)、いずれのがん種でも活性がない場合を検討した。活性がないがん種の奏効率に関しては灰色で記載した。

主要評価項目に二項分布を適用し、欠測値や脱落例はノンレスポンダーとして、データを生成した。各シナリオについて、繰り返し数は 10,000 回とした。

表2 各シナリオでの各がん種における登録例数及び奏効率

	がん1		がん2		がん3		がん4		がん5	
シナリオ	登録例数	奏効率	登録例数	奏効率	登録例数	奏効率	登録例数	奏効率	登録例数	奏効率
1-1	20	0.35	20	0.35	20	0.35	20	0.35	20	0.35
1-2		0.35		0.35		0.35		0.15		0.15
1-3		0.15		0.15		0.15		0.15		0.15
2-1	60	0.35	10	0.35	10	0.35	10	0.35	10	0.35
2-2		0.35		0.15		0.15		0.15		0.15
2-3		0.15		0.15		0.15		0.15		0.15
3-1	30	0.35	30	0.35	30	0.35	5	0.35	5	0.35
3-2		0.35		0.35		0.35		0.15		0.15
3-3		0.15		0.15		0.15		0.15		0.15

シミュレーションの動作特性評価のため、下記の指標について検討した。

指標	内容
Selection Power	活性があるがん種のうち 80%以上が主要解析に含まれる確率
Proportion of Success	主要解析のベイズ流階層モデルによって一つ以上のがん種で成功と判定される確率（成功判定基準: $\Pr(\pi_j > 25\%) > 80\%$）
Family-wise Error Rate (FWER)	活性がないがん種で一つ以上偽陽性により有効性を主張してしまう確率。Proportion of Success と同様のベイズ流階層モデル判定基準を用いる
Power	主要解析の併合解析で有効と判定される確率
Mixing Nonactive	一つ以上活性がないがん種が含まれる中で、主要解析の併合解析で有効と判定される確率

5 シミュレーション結果

シミュレーション結果を表 3 に示す。

がん種間の登録割合が同じ状況（シナリオ 1-1~1-3）

いずれのがん種でも活性がある状況では（シナリオ 1-1)、Prior の設定によらず、いずれの指標でも性能はほぼ同じであった。

一部のがん種で活性がない状況（シナリオ 1-2）では、より情報借用する Prior 2 で Proportion of Success が低くなった。これは、シナリオ 1-1 と比較して奏効率が低いがん種が多くなり、ベイズ流階層モデルを用いることで活性があるがん種でも奏効率が低く推定されることだと考えられた。

すべてのがん種で活性がない場合では(1-3)、Power が 0.05 を超えており、erroneous conclusion rate が適切に制御できていなかった。これは中間解析で活性がないにも関わらず、偶然奏効率が高くなったがん種のみを残して試験が続行され、最終解析に各がん種で併合した解析を行うことで、治療効果が過大推定されるためだと考えられる。一方、最終解析でベイズ流階層モデルを用いた評価を行った Proportion of Success は 0.05 以下となっていた。これは，各がん種の結果を用いた階層ベイズモデルにより治療効果を評価したため，奏効率が低いがん種があった場合にその影響を受けたためと考えられる。

<u>がん種間の登録割合が異なる状況（シナリオ 2-1~2-3）</u>

いずれのがん種でも活性がある状況では（シナリオ 2-1）、Prior の設定によらず、いずれの指標でも性能はほぼ同じであった。しかし Selection Power はシナリオ 1-1 よりも低くなった。これはがん種 2-5 での被験者数が少ないため、偶然いずれかのがん種で治療効果が低くなり、無効中止されてしまったものがあったためと考えられる。

一部のがん種で活性がない状況（シナリオ 2-2）では、活性があるがん種は、最も発現割合が高いがん種 1 のみのシナリオなので、最終解析に含められる確率は高くなり、シナリオ 1 と比較して Selection Power が高くなった。その一方で、がん種間の情報借用をすることで、活性がないがん種も含めて活性があると判断してしまう、FWER 及び mixing nonactive が高くなった。

すべてのがん種で活性がない状況（シナリオ 2-3）では、Prior 2 を使い、よりがん種間の情報借用をすることで、偶然奏効率が高かったがん種があった場合でも、他の奏効率が低いがん種の結果に引っ張られてしまうため、Prior 1 を用いた場合と比較して Proportion of Success 及び FWER が低くなった。

<u>一部のがん種の登録割合が非常に低い状況（シナリオ 3-1~3-3）</u>

いずれのがん種でも活性がある状況では（シナリオ 3-1）、Prior の設定によらず、いずれの指標でも性能はほぼ同じであった。

希少ながん種で活性がない状況（シナリオ 3-2）では、Prior 2 を使い、よりがん種間の情報借用をすることで、活性がないがん種の結果に引っ張られてしまうため、Prior 1 を用いた場合と比較して Selection power が低くなった。

すべてのがん種で活性がない状況（シナリオ 3-3）での動作特性は、シナリオ 1-3 の場合とほぼ同様であった。

表3　各シナリオでのシミュレーション結果

	Selection Power		Proportion of Success		Family-wise Error Rate (FWER)		Power		Mixing Nonactive	
シナリオ /Prior	1	2	1	2	1	2	1	2	1	2

1-1	0.61	0.61	0.92	0.87	-	-	0.99	0.98	-	-
1-2	0.40	0.40	0.82	0.74	0.06	0.04	0.96	0.90	0.32	0.31
1-3	-	-	0.05	0.03	0.05	0.03	0.16	0.12	0.16	0.12
2-1	0.39	0.39	0.91	0.84	-	-	0.98	0.98	-	-
2-2	0.88	0.84	0.77	0.70	0.13	0.06	0.88	0.84	0.46	0.46
2-3	-	-	0.04	0.02	0.04	0.02	0.13	0.13	0.13	0.13
3-1	0.72	0.62	0.91	0.85	-	-	0.99	0.97	-	-
3-2	0.56	0.49	0.87	0.81	0.11	0.06	0.98	0.96	0.62	0.40
3-3	-	-	0.05	0.01	0.05	0.01	0.14	0.10	0.14	0.10

6　まとめ

本発表では特定の遺伝子異常が発現した複数の固形がんに対する有効性をまとめて評価するバスケット試験を実施する場合での、ベイズ流階層モデルの適用について検討した。ベイズ流階層モデルを用いることで各がん種間での奏効率の情報を借用し、より効率の良い推定が行うことができることが示された。

一方で、ベイズ流階層モデルを用い、無益性中止しなかったがん種のみを最終解析で頻度論流での併合解析に含めるため、治療効果の過大推定が生じ、erroneous conclusion rate が増える場合があった。各がん種での活性の有無についてベイズ流階層モデルを適用して評価することで、erroneous conclusion rate を減らすことが可能であった。事前分布の設定でより情報を借用するものを利用した場合、一部のがん種で奏効率が低い場合には、その結果に引っ張られ、無情報に近い事前分布を仮定した場合にはその結果に影響をより受ける可能性がある。がん種横断的な適応を狙いバスケット試験を行う場合には、一部のがん種で活性が認められなかった場合、その妥当性は議論になる可能性はあると考えられる。

7　参考文献

1. Pant, S., Schuler, M., Iyer, G., et al. RAGNAR Investigators (2023). Erdafitinib in patients with advanced solid tumours with FGFR alterations (RAGNAR): an international, single-arm, phase 2 study. The Lancet. Oncology, 24(8), 925–935.
2. 日本製薬工業協会 医薬品評価委員会の報告書「ベイズ流試験デザインにおけるシミュレーションの利用法」

BASE SAS環境におけるk近傍法を用いた欠測値補完の実装と精度評価

○小林　邦世

（イーピーエス株式会社）

Implementation and evaluation of missing value imputation method “k-nearest neighbor” in BASE SAS

Kuniyo Kobayashi
EPS corporation

要旨

欠測値補完はデータ分析における重要な前処理の一つであり，手法の選択や実装方法により結果に大きな影響を与える．k 近傍法（k-Nearest Neighbor: KNN）は，シンプルかつ汎用性が高く，数値・カテゴリ変数を問わず適用可能な方法として知られている．本発表では，BASE SAS 環境において KNN 法を実装するための方法を検討する．具体的には，数値変数向けのユークリッド距離計算と，数値とカテゴリ変数を組み合わせた場合においても利用できる距離計算の仕組みをデータステップを用いて構築する．また，これらを用いた欠測値補完の流れを，小規模なサンプルデータを通じて示す予定である．BASE SAS の標準機能のみで柔軟な補完処理が可能であることを，実装例を交えて紹介したい．

キーワード：欠測値補完，k 近傍法，k-Nearest Neighbor

緒言

実データを用いた解析においては, 欠測値の存在が避けられず, その処理方法は結果の信頼性に直結する．単純な除外法ではサンプルサイズの減少や推定の偏りを招くため，近年は統計的推定や機械学習に基づく欠測値補完手法が用いられている．本論文では，その中でも k 近傍法（k-Nearest Neighbor: KNN）を用いた欠測値補完法(KNN 補完)に着目した．KNN 補完は，欠測値を含む観測に対して類似度の高い近傍を探索し，その情報を基に推定値を与える手法であり，分布仮定に依存せず柔軟に適用できる特徴をもつとされている[1][2]. 本論文では BASE SAS 環境において KNN 補完を実装し，その補完精度を検証した．

k 近傍法を用いた欠測値補完について

k 近傍法とは

パターン分類とは，複数の特徴量で表現された対象を，あらかじめ定義された有限個のクラスのいずれかに割り当てる問題である．この課題は医療を含む多くの分野で重要であり，未知のデータを既知のカテゴリに正しく分類することが求められる．このような分類問題に対する代表的な手法の一つに，1967 年に Cover と Hart によって提案された k 近傍法（KNN）がある[3]．KNN はノンパラメトリックな分類器であり，デー

タの分布を仮定する必要がない．データ内の各観測値は多次元空間上の点として表現され，未知の観測値を分類する際には，まず未知の観測値とデータ中のすべての観測値との距離を計算する．次に，距離が最も近い k 件の観測値（近傍）を抽出し，その所属クラスの多数決に基づいて，未知の観測値のクラスが決定される．例えば，k=1 であれば最も近い 1 点のクラスをそのまま割り当て，k=5 であれば近傍 5 点のうち多数を占めるクラスが選ばれる(Figure 1)[4].

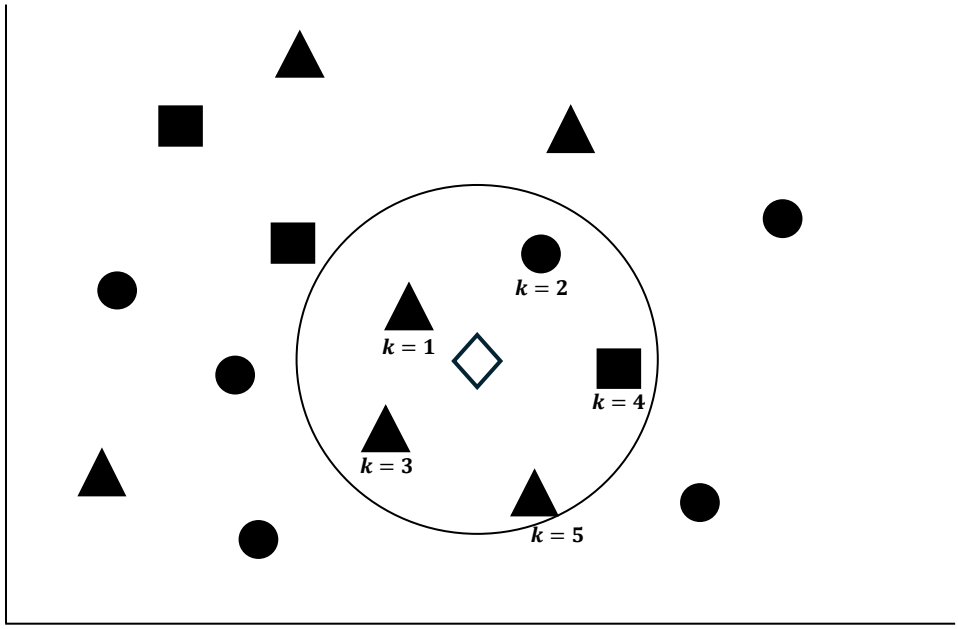


Figure 1 近傍法によるクラス予想．クラスの種類は●，▲，■であり，未知の観測値◇のクラスは，$k=5$のとき，多数を占める▲になる．

このとき，k の値はクラスの割り当てに影響を与える．k が小さい場合はクラスの局所的な特徴を反映するが外れ値に影響されやすく，k が大きい場合は外れ値に頑健な一方でクラスの特徴がぼやけることがある[2].

さらに，KNN の性能は未知の観測値とデータ内のその他の観測値の「距離」の定義に大きく依存する，連続変数のみで構成されるデータの観測値同士の距離の算出にはいくつかの手法が提案されているが，その中でも代表的な 3 つの距離の計算方法を以下に示す．なお，変数の数はm個とし，未知の観測値を$A=(x_1,x_2,x_3,\cdots,x_m)$，その他の任意の観測値を$B=(y_1,y_2,y_3,\cdots,y_m)$と表す．

ユークリッド距離

最も基本的な距離尺度で，数値データに広く用いられる一方，外れ値に弱いとされる．

$$dist_Euclid(A,B)=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2}$$

マンハッタン距離

外れ値に対して頑健である．

$$dist_Manhattan(A,B)=\sum_{i=1}^{m}|x_i-y_i|$$

ミンコフスキー距離

ユークリッド距離やマンハッタン距離を一般化したもの．パラメータqにより柔軟に調整できる．

$$dist_Minkowsky(A,B) = \left(\frac{\sum_{i=1}^{m} |x_i - y_i|^q}{m} \sum_{i=1}^{m} |x_i - y_i|^q \right)^{\frac{1}{q}}$$

$q = 1$でマンハッタン距離，$q = 2$でユークリッド距離となる[5].

なお，これらの距離尺度を用いる場合，変数間でスケール（単位や範囲の大きさ）が大きく異なると，距離計算において一部の変数の影響が大きくなり，結果が歪むことがある．例えば,「身長 (m)」と「体重 (kg)」を同時に扱う場合，値の変動幅が異なるため，このままでは体重の寄与が大きくなりやすい．この問題を回避するため，実務上はスケーリングが行われる．代表的な方法は以下の通りである．

① Z スコア標準化（Standardization）

任意の観測値iの変数kの値x_{ik}を，変数k全体の値の平均μ_k，標準偏σ_kを用いて変換する．

$$z_{ik} = \frac{x_{ik} - \mu_k}{\sigma_k}$$

このとき，変換後の変数z_{ik}は平均 0，分散 1 となり，すべての変数が同等の重みで距離計算に寄与できるようになる．

② 最小最大正規化（Min-Max Normalization）

任意の観測値iの変数kの値x_{ik}を，変数kの全値における最小値$min(X_k)$，最大値$max(X_k)$を用いて次のように変換する．

$$x'_{ki} = \frac{x_{ik} - min(X_k)}{max(X_k) - min(X_k)}$$

これにより変数x'_{ki}の値は$[0,1]$に収められ，尺度の違いによる不均衡が解消される[6].

このようにスケーリングは，ユークリッド距離やマンハッタン距離といった距離尺度を適切に適用するための前処理として欠かせない．特に多次元データにおいては，スケーリングの有無が分類性能に大きな差を生むため，KNN を実務に適用する際には考慮すべき工程であるとされている[7].

また，混合型データ（数値変数とカテゴリ変数が混在しているデータ）に対しては，Gower 距離が広く利用される．これは項目ごとにスケール（単位や範囲の大きさ）を揃えて，その項目ごとの違いを比べ，最後に，すべての項目の違いを平均して，全体の距離を算出する方法である．

Gower 距離

Gower（1971）は，混合型データ（連続変数・カテゴリ変数・順序変数・バイナリ変数）に対して，統一的に近さ（類似度）と距離を定義する手法を提案した．この手法は，変数ごとに 0〜1 に正規化された一致度を定め，それらの加重平均から観測間の総合的な類似度を構成する点に特徴がある．

2 つの観測値 A とBの類似度S_{AB}を次式で定義する．

$$S_{AB} = \frac{\sum_{k=1}^{p} \omega_{ABk} s_{ABk}}{\sum_{k=1}^{p} \omega_{ABk}}$$

ここで，pは変数の数，s_{ABk}は変数kに基づく類似度スコア，ω_{ABk}は重みを示し，通常は変数 k に観測値 A, B ともに値が存在する場合に 1，欠測している場合には 0 と設定される．

変数kに基づく類似度s_{ABk}については以下のように算出する．

① 連続変数（数値）

任意の観測値 A, B の変数kの値x_{Ak}, x_{Bk}に対して，変数kの全観測値における最小値，最大値をそれぞれ$min(x_k)$，$max(x_k)$と表したときの範囲$R_k = max(x_k) - min(x_k)$を用いて次のように定義する．

$$S_{ABk} = 1 - \frac{|x_{Ak} - x_{Bk}|}{R_k}$$

② カテゴリ変数（名義尺度）

任意の観測値 A, B の変数kの値x_{Ak}, x_{Bk}に対して，値x_{Ak}, x_{Bk}が一致/不一致での場合分けをして 0 もしくは 1 で定める．

$$R_k = \begin{cases} 1, & x_{Ak} = x_{Bk} \quad (\text{一致する}) \\ 0, & x_{Ak} \neq x_{Bk} \quad (\text{一致しない}) \end{cases}$$

③ 順序変数

カテゴリを 0 から 1 の区間に等間隔で割り当て，連続変数と同様に扱う．

④ バイナリ変数

バイナリ変数の取り扱いにおいては，対称型と非対称型の二種類を区別する必要がある．

a, 対称型バイナリ変数

値「0」と「1」の両方に同等の意味がある場合（例：性別），一致すれば$s_{ABk} = 1$，不一致なら$s_{ABk} = 0$と定義する．

b, 非対称型バイナリ変数

例えば，値「1」が「特性あり」，「0」が「特性なし」を意味する場合（例：疾患有無），両者が「0」で一致しても，これは類似度に含めないことがある．すなわち，両方が「特性なし」である場合は，情報量がないと見なし類似度から除外するのが妥当であるとされている．これを一般化すると，非対称型の場合の類似度は次のように定義できる．

$$s_{ijk} = \begin{cases} 1, & x_{Ak} = x_{Bk} = 1 \\ 0, & x_{Ak} \neq x_{Bk} \\ excluded, & x_{Ak} = x_{Bk} = 0 \end{cases}$$

したがって，最終的な距離D_{AB}は以下のように定義される．

$$D_{AB} = 1 - S_{AB}$$

すなわち，2 つのサンプル間の一致度が高いほど類似度は大きくなり，距離は小さくなる．逆に，不一致が多い場合は類似度が小さくなり，距離が大きくなる[7].

KNN における距離の定義は，補完結果に大きな影響を与える．数値データに対してはユークリッド距離が一般的に用いられるが，異なるスケールの変数を含む場合には標準化が必須となる．また，マンハッタン距離は外れ値の影響を受けにくい性質を持つ．さらに，数値とカテゴリが混在する場合には Gower 距離が有効であり，変数タイプに応じた柔軟な類似度計算が可能となる[5].

KNN を用いた欠測値補完法

KNN の考え方は，欠測値補完にも拡張できる．基本方針は「欠測を含む観測値に最も近いk件のサンプルを探索し，その情報から欠測成分を推定する」ことである[9].

欠測値を含み補完を必要とする観察値を補完対象値と呼ぶことにする．さらに，補完を行う際，データ内の他の観察値を候補値と呼ぶことにする．このとき, KNN を利用した欠測値補完の手順は以下の通りになる．

① 対象観測値と候補観測値との距離を算出する．

補完対象値内の任意の変数と各候補値内の対応する変数との距離を算出する．ただし，距離計算には両方の観測値で観測されている変数のみを用い，いずれかで欠測が発生している変数は距離計算から除外する．補完対象値Aの値を$A = (x_1, x_2, x_3, \cdots, x_m)$，これに対応する候補値$B$の値を$B = (y_1, y_2, y_3, \cdots, y_m)$とし，補完対象値 A あるいは候補値のいずれかで欠測している変数kの値をx_kあるいはy_kと表すとき, A と B の距離$d(A, B)$は以下の通りに求められる．

$$d(A,B) = f\left(\{(x_i, y_i)\} \middle| x_i \neq \text{欠測}, \quad y_i \neq \text{欠測},\right), \qquad i = 1,2,3,\cdots,m \ (0 \leq k < m \text{かつ} k \text{は除外})$$

ここで$f(\cdot)$はユークリッド距離や Gower 距離を算出するための距離関数である．

② ①で算出した距離に基づき距離が最も近い k 件の観測値を抽出する．

ユークリッドや Gower 距離計算式で算出された距離の大きさに基づき，値が小さい順にk件の観測値を抽出する．kの値に関して，先行研究では $k = 10$ や $k = 15$が実際に用いられており，本研究でもこれを参考に$k = 10$にて実施する[1][8].

③ ②で抽出した距離が最も近い k 件の観測値から補完値を推定する．

補完対象変数が連続変数の場合は，距離が最も近い k 件(本論文では$k = 10$)の観測値における該当変数の値の平均または距離の逆数を重みとした加重平均を割り当てる．補完対象変数がカテゴリ変数の場合においては，距離が最も近い k 件の観測値の該当変数の最頻カテゴリを割り当てる[9][10].

なお，KNN 補完では「近傍を算出するための比較可能な変数」が必要である．本手法では欠測値が存在するデータにおいて，補完対象値と候補値と両者で観測された変数のみを利用して距離を計算する設計としているが，補完対象変数以外の変数全体で欠測が多すぎる場合には，有効な近傍が見つからず正確な補完結果が得られない可能性がある点に注意すべきとの指摘もされている[8][9].

KNN を利用した欠測値補完法の BASE SAS での実装

本論文では，BASE SAS のみで KNN 補完を実装するマクロを作成した．本マクロは数値変数に対してユークリッド距離，数値・カテゴリ両変数に対して対応可能な Gower 距離を選択可能とし，補完対象変数の値が数値の場合は平均，カテゴリの場合は近傍内カテゴリの最頻値によって決定するようにした．マクロの詳細については Appendix 1 を参照のこと．本マクロは複数のパラメータを指定できる設計となっている．主要な指定項目は以下の通りである．

- **data, out**：入力データセットおよび出力データセットの指定.
- **id**：識別子変数を指定.
- **target, target_type**：補完対象変数とその型（数値：NUM，カテゴリ：CHAR）.
- **inputs_num, inputs_cat**：距離計算に用いる数値変数およびカテゴリ変数をそれぞれ指定.

- **k**：近傍数の指定（既定値は 3 とした）.
- **distance**：距離関数の選択（Gower 距離であれば”GOWER”，ユークリッド距離であれば”EUCLID”を指定）.
- **scale**：ユークリッド距離使用時のスケーリング方法（Z スコア標準化の場合は「Z」，MINMAX 正規化の場合は「MINMAX」，スケーリングなしの場合は「NONE」を指定．指定のない場合，Z スコア標準化を実施したうえでユークリッド距離を計算する.）

これらのパラメータにより，データ特性や解析目的に応じて補完方法を柔軟に調整できるようにした.

実装例

本論文では，SAS 標準データセットである SASHELP.HEART を利用して KNN 補完マクロの実装例を提示する. SASHELP.HEART は 10 の数値変数と 7 のカテゴリ変数からなるデータセットである. SASHELP.HEART の各変数について欠測割合を確認した後, 以下の手順で KNN 欠測値補完の実装を実行した. SASHELP.HEART の欠測割合の確認結果は Appendix 2 に記載している.

① 欠測データの人工的な作成（Missing Completely at Random に基づく欠測導入）

補完対象変数として数値変数 Systolic，カテゴリ変数 BP_Status の 2 変数を選択した．Systolic，BP_Status ともに欠測値の存在しない変数であるため，真値を保持した上で Missing Completely at Random(MCAR)メカニズムの下，人工的に欠測を発生させた.

② KNN マクロを用いた欠測値補完

提案マクロ%knn_impute を適用し，欠測を補完した．Systolic，BP_Status の両者ともに，数値変数のみを利用するユークリッド距離，数値変数とカテゴリ変数の両方を利用する Gower 距離の 2 種類の距離計算により補完を実施した．近傍についてはすべて k=10 として取り扱った．なお，ユークリッド距離は Z スコア標準化を実施したうえで算出することとする.

③ 補完後データと真値の比較による精度評価

数値変数 Systolic の補完精度の評価には，平均絶対誤差（Mean Absolute Error: MAE）および二乗平均平方根誤差（Root Mean Squared Error: RMSE）を利用した．MAE は補完値と真値との差の絶対値を平均した指標であり，補完値と真値の誤差の大きさを直感的に把握できる利点がある．一方，RMSE は誤差を二乗して平均した後に平方根をとるため，誤差の大きさを相対的に強調する性質を持つ．両者を併用することにより，欠測値補完法の性能を「補完値の真値からの誤差の大きさ」として検討することを目的とした.

MAE, RMSE の計算式については下記に定義する．補完後の変数における補完値を$\hat{x}_i$，真値をx_i，対象となる欠測観測値の総数をnと表す.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_i - x_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}$$

これらの指標は欠測補完に関する先行研究でも広く用いられており，補完精度の比較において標準的な評価方法であると主張されている[2][5].

カテゴリ変数 BP_Status の補完精度の評価には Accuracy（正解率）を使用した．Accuracy とは，補完した値が真の値と一致した割合を示す指標であり，正しく推測できた「当たり」の割合を意味する．数式では次のように表される．

$$Accuracy = \frac{(\text{真値と一致した補完値の数})}{(\text{全補完値の数})}$$

Accuracy は，連続変数に対する RMSE や MAE に相当し，カテゴリ変数の代表的な評価基準として広く用いられている．Accuracy の値が大きいほど，補完手法が真値をより適切に再現していることを意味する[4][9].

結果

① 欠測データの人工的な作成

KNN 欠測値補完法の評価を行う前段階として, SAS 標準データセットである SASHELP.HEART 内の Systolic 変数または BP_Status 変数において MCAR20%のメカニズムで欠測を人工的に発生させた 2 種類のデータセットを取得した．欠測割合は各変数ともにおおよそ 20%と設定した．

例として，Systolic 変数欠測を発生させるためのプログラムを以下に記載する．

```
%let MISS_PROP = 0.20;          /*  欠測導入割合（MCAR）  */
%let SEED      = 12345;         /*  乱数シード  */

/*  データ作成：行 ID と真値の保持（SYSTOLIC を対象に欠測を導入）  */
data heart_mcar;
   set sashelp.heart;
   rowid = _n_;                              /*  自作 ID（数値）  */
   Systolic_true = Systolic;                 /*  真値を保持  */
   length masked 8; masked = 0;

   /* MCAR で欠測導入（対象がもともと欠測でない場合のみ）  */
   if not missing(Systolic) then do;
      call streaminit(&SEED);
      if rand('uniform') < &MISS_PROP then do;
         Systolic = .;
         masked = 1;
      end;
   end;
run;
```

Systolic 変数，BP_Status 変数に欠測を発生させたそれぞれのデータセットにおいてそれぞれの変数を確認したところ，ともに 20%の欠測が発生していることを確認した(Figure 2, 3).

Systolicの欠測割合と平均値・標準偏差

変数名	N	欠測	欠測割合	平均	標準偏差
Systolic	5,209	1,046	20.08%	136.92	23.58

Figure 2　欠測発生後の Systolic 変数の欠測割合と平均値・標準偏差

BP_Statusのカテゴリ別頻度の集計結果

		N = 5209	
変数	カテゴリ	件数	割合
BP_Status	(Missing)	1046	20.08%
	High	1824	35.02%
	Normal	1701	32.66%
	Optimal	638	12.25%

Figure 3　欠測発生後の BP_Status 変数の各カテゴリの欠測割合

② KNN を用いた欠測値の補完

①において 20%MCAR 欠測を発生させた Systolic, BP_Status ともに KNN を用いた欠測値補完を実施した．補完にあたり Systolic，BP_Status 以外に年齢・身長・体重・拡張期血圧・コレステロール値の数値変数，および性別・喫煙習慣のカテゴリ変数(Gower 距離でのみ使用)を距離計算に採用した．

例として，Systolic におけるユークリッド距離を用いた KNN 補完マクロの実行プログラムを以下に記載する．

```
%knn_impute(
   data=Systolic_mcar,
   out =Systolic_imp_euc,
   id    =rowid,
   target=Systolic,
   target_type=NUM,              /* Systolic は数値変数のため NUM を指定*/
   inputs_num=AgeAtStart Height Weight Diastolic Cholesterol,
   inputs_cat=,                  /*  ユークリッド距離計算のためカテゴリ変数は指定しない  */
   k=10,
   distance=EUCLID,           /* ← 距離計算の方法をユークリッドに指定  */
   id_type=NUM
);
data Systolic_imp_euc2;
   set Systolic_imp_euc;
   rename Systolic = Systolic_KNN;
run;
```

また，BP_Status における Gower 距離を用いた KNN 補完マクロの実行プログラムも以下に記載する．

```
%knn_impute(
   data=BPS_mcar,
   out =BPS_imp_gower,
   id    =rowid,
   target=BP_Status,
   target_type=CHAR,             /* BP_Status はカテゴリ変数のため CHAR を指定*/
   inputs_num=AgeAtStart Height Weight Diastolic Cholesterol,
   inputs_cat=Sex Smoking Status,
   k=10,
   distance=GOWER,               /* ← 距離計算の方法を Gower に指定  */
   id_type=NUM
);
data BPS_imp_gower2;
   set BPS_imp_gower;
   rename BP_Status = BPS_KNN;
run;
```

Systolic，BP_Status のユークリッド距離，Gower 距離それぞれを用いた KNN 補完結果の集計結果を以下 Figure 4~7 に示す．

KNN補完(ユークリッド)におけるSystolicの欠測割合と平均値・標準偏差

変数名	N	欠測	欠測割合	平均	標準偏差
Systolic_KNN	5,209	0	0.00%	136.83	22.11

Figure 4　KNN 補完(ユークリッド距離)後の Systolic 変数の欠測割合と平均値・標準偏差

KNN補完(Gower)におけるSystolicの欠測割合と平均値・標準偏差

変数名	N	欠測	欠測割合	平均	標準偏差
Systolic_KNN	5,209	0	0.00%	136.64	21.75

Figure 5　KNN 補完(Gower 距離)後の Systolic 変数の欠測割合と平均値・標準偏差

KNN補完(ユークリッド)におけるBP_Statusのカテゴリ別頻度の集計結果

		N = 5209	
変数	カテゴリ	件数	割合
BPS_KNN	(Missing)	0	0%
	High	2202	42.27%
	Normal	2318	44.50%
	Optimal	689	13.23%

Figure 6　KNN 補完(ユークリッド距離)後の BP_Status 変数の各カテゴリの欠測割合

KNN補完(Gower)におけるBP_Statusのカテゴリ別頻度の集計結果

		N = 5209	
変数	カテゴリ	件数	割合
BPS_KNN	(Missing)	0	0%
	High	2224	42.70%
	Normal	2301	44.17%
	Optimal	684	13.13%

Figure 7　KNN 補完(Gower 距離)後の BP_Status 変数の各カテゴリの欠測割合

Systolic，BP_Status のユークリッド距離，Gower 距離それぞれを用いた KNN 補完結果すべてにおいて，欠測値はすべて補完されていることを確認した. 真値における集計結果と比較し, Systolic のユークリッド距離, Gower 距離それぞれを用いた KNN 補完結果においては概ね真値を再現できていた．一方，BP_Status のユークリッド距離，Gower 距離を用いた KNN 補完結果においては真値と比較して各カテゴリにおける割合の再現がやや不正確であった.

③　補完精度の評価

Systolic のユークリッド距離，Gower 距離それぞれを用いた KNN 補完結果に対して MAE，RMSE を，BP_Status のユークリッド距離，Gower 距離を用いた KNN 補完結果に対しては Accuracy を算出した．それぞれの算出に用いたプログラムは Appendix 3 にて記載する.

Systolic のユークリッド距離，Gower 距離それぞれに対する MAE，RMSE，そして BP_Status のユークリッド距離，Gower 距離に対する Accuracy については以下 Figure 8~11 の通りの結果となった.

SystolicにおけるKNN補完（ユークリッド）の評価

impute_obs	MAE	RMSE
1046	11.4607	16.4653

Figure 8　Systolic 変数における KNN 補完(ユークリッド距離)の MAE, RMSE の結果

SystolicにおけるKNN補完（Gower）の評価

impute_obs	MAE	RMSE
1046	14.0257	19.8836

Figure 9　Systolic 変数における KNN 補完 Gower 距離)の MAE, RMSE の結果

BP_StatusにおけるKNN補完（ユークリッド）の評価

Accuracy
0.706

Figure 10　BP_Status 変数における KNN 補完(ユークリッド距離)の Accuracy の結果

BP_StatusにおけるKNN補完（Gower）の評価

Accuracy
0.587

Figure 11　BP_Status 変数における KNN 補完(Gower 距離)の Accuracy の結果

Systolic の補完結果の MAE，RMSE を比較すると，Gower 距離において，ユークリッド距離の MAE，RMSE よりも大きい傾向がみられた．また，BP_Status における補完結果では，Gower 距離の Accuracy はユークリッド距離のものに対して小さい結果となった．

考察

KNN を利用した欠測値補完マクロを利用し，標準データセット（SASHELP.HEART）を用いて KNN 補完が実装できること，そしてその精度の 2 点の検討を実施した．数値変数（Systolic），カテゴリ変数（BP_Status）ともにユークリッド距離および Gower 距離の両方で KNN 補完の実装が可能であることが示された．精度については，Systolic に対する補完においては，補完後の平均値・標準偏差，そして真値からの誤差指標（MAE, RMSE）のいずれの検討においても真値を概ね再現できたと考えられた．特にユークリッド距離を用いた場合において誤差指標が小さくなる傾向が確認され，その有効性が示唆された．一方，BP_Status に対する補完においては，ユークリッド距離，Gower 距離ともに欠測発生前の各カテゴリの割合の再現がやや不正確であったこと，真値再現率(Accuracy)が両距離計算式においても 0.8 を下回る結果となったことなどから，補完自体は可能であるものの精度自体には疑問が残る結果となった．特に Gower 距離を用いた補完では，Accuracy が 0.6 未満に留まり，十分な精度が得られなかった．これらの要因として，BP_Status の分布が偏っており近傍が有効に働きにくかった可能性，すなわち近傍数 k=10 という設定が変数特性に対して最適ではなかった可能性があげられる．したがって，カテゴリ変数の補完においては距離尺度の選択，あるいは k 値の設定をより慎重に行う必要があると考えられる．

まとめ

BASE SAS において KNN を利用した欠測値補完マクロを実装し，標準データセット（SASHELP.HEART）を用いてその精度を検討した．KNN 補完は「近傍を計算するための共通変数が存在する」ことを前提とするため，欠測が広範囲に及ぶ場合や高次元データに対しては有効な近傍探索が困難となることや計算処理が膨大になり処理機に過負荷がかかるリスクがあることが指摘されている[10]．本論文では MCAR 20%欠測の条

件下で検証したが，実データでは欠測メカニズムが Missing at Random や Missing Not at Random である場合や欠測率がより高い場合も考えられる．また，今回作成した KNN 補完マクロは BASE SAS 標準機能のみで柔軟に欠測補完を実行できる点で有用性が高いと考えられるものの，カテゴリ変数における精度低下や，欠測パターン・メカニズムに精度が依存する点で課題が残る結果となった．今後は，カテゴリ変数に対する改良的な KNN 補完法の実装検討や，シミュレーション条件を変えてのマクロの精度評価，他の補完手法（多重代入法、ランダムフォレスト補完等）と KNN 補完法の比較などにも着手していきたい．

参考文献

[1] Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R.& Altman, R. B. Missing value estimation methods for DNA microarrays. Bioinformatics, 17(6), 520–525 (2001).

[2] Beretta, L., & Santaniello, A. Nearest neighbor imputation algorithms: A critical evaluation. BMC Medical Informatics and Decision Making, 16(Suppl 3), 74 (2016).

[3] Cover, T., & Hart, P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27 (1967).

[4] El Badisy, I., Graffeo, N., Khalis, M. *et al.* Multi-metric comparison of machine learning imputation methods with application to breast cancer survival. BMC Med Res Methodol 24, 191 (2024).

[5] Hu, LY., Huang, MW., Ke, SW. et al. The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus 5, 1304 (2016).

[6] Chawla, Shagun and Kumar, Rajat and Aggarwal, Ekansh and Swain, Sarthak, Breast Cancer Detection Using K-Nearest Neighbour Algorithm. International Journal of Computational Intelligence & IoT, Vol. 2, 4 (2018)

[7] Gower, J. C. A general coefficient of similarity and some of its properties. Biometrics, 27(4), 857–871 (1971).

[8] Batista, G. E. A. P. A., & Monard, M. C. A study of k-nearest neighbour as an imputation method. HIS, 87–96 (2002).

[9] Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. BMC Med Inform Decision Making. 16(3) 74 (2016).

[10] KNN Imputation: The Complete Guide - KNN Imputation: A Complete Guide to Handling Missing Data with Precision and Accuracy.
(https://medium.com/@tarangds/knn-imputation-the-complete-guide-146f932870a7, 最終アクセス: 2025/08/25)

Appendix 1　KNN を利用した欠測値補完マクロ：%knn_impute

```
/*------------------------------------------------------------
   ヘルパ：接尾辞つき名前リスト生成
------------------------------------------------------------*/
%macro suffix_list(list, suf);
   %local n inx tok out;
   %let n=%sysfunc(countw(&list,%str( )));
   %let out=;
   %do inx=1 %to &n;
      %let tok=%scan(&list,&inx,%str( ));
      %let out=&out &tok.&suf;
   %end;
   &out
```

```
%mend;

/*------------------------------------------------------------
   ヘルパ：リストから特定トークンを除外
   例）%list_remove(A B C, B) → A C
------------------------------------------------------------*/
%macro list_remove(list, drop);
   %local n i tok out;
   %let n=%sysfunc(countw(&list,%str( )));
   %let out=;
   %do i=1 %to &n;
      %let tok=%scan(&list,&i,%str( ));
      %if %upcase(&tok) ne %upcase(&drop) %then %let out=&out &tok;
   %end;
   &out
%mend;

/*============================================================
   KNN 補完マクロ（完全ケース縛り撤廃 + target 自動除外）
============================================================*/
%macro knn_impute(
   data=, out=, id=ID,
   target=, target_type=NUM,              /* NUM | CHAR */
   inputs_num=, inputs_cat=,              /* Gower=両方 OK，Euclid=数値必須 */
   k=3,
   distance=GOWER,                        /* GOWER | EUCLID */
   scale=Z,                               /* EUCLID 時：Z | MINMAX | NONE */
   id_type=CHAR,                          /* CHAR | NUM */
   id_len=200                             /* 文字 ID の長さ統一用 */
);

   %local DIST SCALE IDTYPE has_num has_cat inputs_num2 inputs_cat2;
   %let DIST    = %upcase(&distance);
   %let SCALE   = %upcase(&scale);
   %let IDTYPE = %upcase(&id_type);

   /* ---- target が紛れても距離から自動除外 ---- */
   %let inputs_num2 = %list_remove(%superq(inputs_num), &target);
   %let inputs_cat2 = %list_remove(%superq(inputs_cat), &target);

   /* 入力リストの空対応 */
   %if %length(%superq(inputs_num2)) %then
      %let has_num = %sysfunc(countw(%superq(inputs_num2)));
   %else %let has_num = 0;

   %if %length(%superq(inputs_cat2)) %then
      %let has_cat = %sysfunc(countw(%superq(inputs_cat2)));
   %else %let has_cat = 0;

   /* チェック：EUCLID は数値入力が必須 */
   %if &DIST = EUCLID and &has_num = 0 %then %do;
      %put ERROR: EUCLID には数値入力（inputs_num=）が 1 つ以上必要です。;
      %return;
   %end;

   /* --- 統計量（EUCLID の標準化 / GOWER の range 用） --- */
   %if &DIST = GOWER and &has_num > 0 %then %do;
      proc means data=&data noprint;
         output out=_stat_
            min(&inputs_num2)= %suffix_list(&inputs_num2,_min)
            max(&inputs_num2)= %suffix_list(&inputs_num2,_max);
      run;
   %end;

   %if &DIST = EUCLID and &has_num > 0 %then %do;
      %if &SCALE=Z %then %do;
         proc means data=&data noprint;
            output out=_stat_
               mean(&inputs_num2)= %suffix_list(&inputs_num2,_mean)
```

```
            std (&inputs_num2)= %suffix_list(&inputs_num2,_std);
        run;
      %end;
      %else %if &SCALE=MINMAX %then %do;
        proc means data=&data noprint;
          output out=_stat_
            min(&inputs_num2)= %suffix_list(&inputs_num2,_min)
            max(&inputs_num2)= %suffix_list(&inputs_num2,_max);
        run;
      %end;
    %end;

    %if (&DIST=GOWER and &has_num>0) or (&DIST=EUCLID and &has_num>0 and &SCALE ne
NONE) %then %do;
      data _stat_;
        set _stat_;
        array __n _numeric_;
        do __i=1 to dim(__n);
          if missing(__n[__i]) then __n[__i]=0;
        end;
        drop __i;
      run;
    %end;

    /* ---- 完全ケース縛りを撤廃：target の有無だけで分割 ---- */
    data _miss _comp;
      set &data;
      if missing(&target) then output _miss;
      else                     output _comp;
    run;

    %local nmiss ncomp;
    proc sql noprint;
      select count(*) into :nmiss from _miss;
      select count(*) into :ncomp from _comp;
    quit;

    %if &nmiss=0 or &ncomp=0 %then %do;
      data &out; set &data; run;
      %put NOTE: 補完対象がないか，参照データがありません。処理をスキップしました。;
      %return;
    %end;

    /* --- 距離計算（各項目ごとにペアワイズに加算） --- */
    data _dist;
      %if (&DIST=GOWER and &has_num>0) or (&DIST=EUCLID and &has_num>0 and &SCALE ne
NONE) %then %do;
        if _n_=1 then set _stat_;
      %end;

      set _miss(rename=(
        %do inum=1 %to &has_num; %let nv=%scan(&inputs_num2,&inum); &nv=&nv._m %end;
        %do icat=1 %to &has_cat; %let cv=%scan(&inputs_cat2,&icat); &cv=&cv._m %end;
        &id=&id._m
      ));

      do _p_=1 to _N_;
        set _comp(rename=(
          %do inum=1 %to &has_num; %let nv=%scan(&inputs_num2,&inum); &nv=&nv._c %end;
          %do icat=1 %to &has_cat; %let cv=%scan(&inputs_cat2,&icat); &cv=&cv._c %end;
          &target=&target._c &id=&id._c
        )) point=_p_ nobs=_N_;

        length _sum 8 _cnt 8; _sum=0; _cnt=0;

        /* 数値（ペアワイズ加算） */
        %if &has_num>0 %then %do inum=1 %to &has_num;
          %let nv=%scan(&inputs_num2,&inum);
          %if &DIST = GOWER %then %do;
```

```
            _range = &nv._max - &nv._min;
            if nmiss(&nv._m, &nv._c)=0 then do;
               if _range>0 then _dnum = abs(&nv._m - &nv._c)/_range;
               else _dnum = 0;
               _sum+_dnum; _cnt+1;
            end;
         %end;
         %else %if &DIST = EUCLID %then %do;
            if nmiss(&nv._m, &nv._c)=0 then do;
               %if &SCALE=Z %then %do;
                  if &nv._std>0 then do
                     ; _z_m = (&nv._m - &nv._mean)/&nv._std
                     ; _z_c = (&nv._c - &nv._mean)/&nv._std
                     ; end;
                  else do; _z_m=0; _z_c=0; end;
                  _sum + (_z_m - _z_c)**2;
               %end;
               %else %if &SCALE=MINMAX %then %do;
                  _range = &nv._max - &nv._min;
                  if _range>0 then do
                     ; _z_m = (&nv._m - &nv._min)/_range
                     ; _z_c = (&nv._c - &nv._min)/_range
                     ; end;
                  else do; _z_m=0; _z_c=0; end;
                  _sum + (_z_m - _z_c)**2;
               %end;
               %else %do;
                  _sum + (&nv._m - &nv._c)**2;
               %end;
               _cnt+1;
            end;
         %end;
      %end;

                  /* カテゴリ（GOWER のみ・ペアワイズ） */
                  %if &has_cat>0 %then %do icat=1 %to &has_cat;
                     %let cv=%scan(&inputs_cat2,&icat);
                     %if &DIST = GOWER %then %do;
                        if cmiss(&cv._m, &cv._c)=0 then do;
                           _dcat = (&cv._m ne &cv._c);
                           _sum+_dcat; _cnt+1;
                        end;
                     %end;
                  %end;

      if _cnt>0 then do;
         %if &DIST = EUCLID %then %do; _dist = sqrt(_sum); %end;
         %else %do; _dist = _sum/_cnt; %end;
         output;
      end;
      /* _cnt=0 は距離が定義できないので出力しない */
      keep &id._m &id._c _dist &target._c;
   end;
run;

/* 近傍 k 件抽出（距離が定義できた組のみ） */
proc sort data=_dist; by &id._m _dist; run;
data _topk;
   set _dist; by &id._m;
   retain _r 0;
   if first.&id._m then _r=0; _r+1;
   if _r<=&k then output;
   drop _r;
run;

/* --- 補完値の作成（IDTYPE に応じて LENGTH 句を出し分け） --- */
%if %upcase(&target_type)=NUM %then %do;
   proc sql;
      %if %upcase(&IDTYPE)=CHAR %then %do;
```

```
            create table _imp as
            select &id._m as &id length=&id_len,
                        mean(&target._c) as &target
            from _topk
            group by &id._m;
        %end;
        %else %do; /* NUM */
            create table _imp as
            select &id._m as &id,
                        mean(&target._c) as &target
            from _topk
            group by &id._m;
        %end;
    quit;
%end;
%else %do; /* CHAR  ターゲット  */
    proc sql;
        %if %upcase(&IDTYPE)=CHAR %then %do;
            create table _vote as
            select &id._m as &id length=&id_len,
                        &target._c as _class length=200,
                        count(*) as _cnt,
                        mean(_dist) as _avgd
            from _topk
            group by &id._m, &target._c;
        %end;
        %else %do; /* NUM */
            create table _vote as
            select &id._m as &id,
                        &target._c as _class length=200,
                        count(*) as _cnt,
                        mean(_dist) as _avgd
            from _topk
            group by &id._m, &target._c;
        %end;
    quit;

    proc sort data=_vote; by &id descending _cnt _avgd; run;
    data _imp; set _vote; by &id;
        if first.&id then do; &target=_class; output; end;
        keep &id &target;
    run;
%end;

/*  書き戻し  */
proc sort data=&data out=_orig; by &id; run;
proc sort data=_imp ;                    by &id; run;

%if &IDTYPE = CHAR %then %do;
    data _orig2; length &id $&id_len; set _orig; run;
    data _imp2 ; length &id $&id_len; set _imp ; run;
    proc sort data=_orig2; by &id; run;
    proc sort data=_imp2 ; by &id; run;

    data &out;
        merge _orig2(in=a)
                _imp2 (in=b rename=(&target=_imp_value));
        by &id;
        if a;
        if missing(&target) and b then &target=_imp_value;
        drop _imp_value;
    run;

    proc datasets lib=work nolist; delete _orig _imp _orig2 _imp2; quit;
%end;
%else %do; /*  数値 ID */
    data &out;
        merge _orig(in=a)
                imp (in=b rename=(&target=  imp  value));
```

```
        by &id;
        if a;
        if missing(&target) and b then &target=_imp_value;
        drop _imp_value;
      run;

      proc datasets lib=work nolist; delete _orig _imp; quit;
   %end;

   /* 後片付け */
   %if %upcase(&target_type)=CHAR %then %do;
      proc datasets lib=work nolist; delete _miss _comp _dist _topk _vote _stat_; quit;
   %end;
   %else %do;
      proc datasets lib=work nolist; delete _miss _comp _dist _topk              _stat_; quit;
   %end;

%mend;
```

Appendix 2　SASHELP.HEART データセット内の数値変数・カテゴリ変数の集計結果

SASHELP.HEARTデータセット内の数値変数の欠測割合と平均値・標準偏差

	N = 5209			
変数名	欠測	欠測割合	平均	標準偏差
AgeAtDeath	3,218	61.78%	70.54	10.56
AgeAtStart	0	0.00%	44.07	8.57
AgeCHDdiag	3,760	72.18%	63.30	10.02
Cholesterol	152	2.92%	227.42	44.94
Diastolic	0	0.00%	85.36	12.97
Height	6	0.12%	64.81	3.58
MRW	6	0.12%	119.96	19.98
Smoking	36	0.69%	9.37	12.03
Systolic	0	0.00%	136.91	23.74
Weight	6	0.12%	153.09	28.92

Figure 12　SASHELP.HEART データセット内の数値変数の集計結果

SASHELP.HEARTデータセット内のカテゴリ変数のカテゴリ別頻度の集計結果

		N = 5209	
変数	カテゴリ	件数	割合
BP_Status	(Missing)	0	0%
	High	2267	43.52%
	Normal	2143	41.14%
	Optimal	799	15.34%
Chol_Status	(Missing)	152	2.92%
	Borderline	1861	35.73%
	Desirable	1405	26.97%
	High	1791	34.38%
DeathCause	(Missing)	3218	61.78%
	Cancer	539	10.35%
	Cerebral Vascular Disease	378	7.26%
	Coronary Heart Disease	605	11.61%
	Other	357	6.85%
	Unknown	112	2.15%
Sex	(Missing)	0	0%
	Female	2873	55.15%
	Male	2336	44.85%
Smoking_Status	(Missing)	36	0.69%
	Heavy (16-25)	1046	20.08%
	Light (1-5)	579	11.12%
	Moderate (6-15)	576	11.06%
	Non-smoker	2501	48.01%
	Very Heavy (> 25)	471	9.04%
Status	(Missing)	0	0%
	Alive	3218	61.78%
	Dead	1991	38.22%
Weight_Status	(Missing)	6	0.12%
	Normal	1472	28.26%
	Overweight	3550	68.15%
	Underweight	181	3.47%

Figure 13　SASHELP.HEART データセット内のカテゴリ変数の集計結果

Appendix 3　Systolic のユークリッド距離 / Gower 距離それぞれを用いた KNN 補完結果に対する MAE / RMSE, BP_Status のユークリッド距離 / Gower 距離それぞれを用いた KNN 補完結果に対する Accuracy 算出プログラム

```
/* 1) Systolic におけるユークリッド-KNN 補完後結果評価 */
data eval_euc;
   merge systolic_mcar(in=a where=(masked=1) keep=rowid Systolic_true masked)
         systolic_imp_euc2(in=c keep=rowid Systolic_knn);
   by rowid;

    /* 欠測導入した行のみ評価 */
   if a and c and not missing(Systolic_true) and not missing(Systolic_knn);

   err_euc = Systolic_knn - Systolic_true;
   abs_euc = abs(err_euc);
   sq_euc   = err_euc**2;
run;

/* MAE / RMSE の算出 */
proc means data=eval_euc n mean;
```

```
   var abs_euc sq_euc err_euc;
   output out=stats_euc n=n_obs
      mean(abs_euc)=MAE_euc
      mean(sq_euc)=MSE_euc
      mean(err_euc)=Bias_euc;
run;

data stats_euc2;
   set stats_euc;
   RMSE_euc = sqrt(MSE_euc);
   keep n_obs MAE_euc RMSE_euc;
run;

data stats_euc2;
   set stats_euc2;
   rename n_obs = impute_obs
                  mae_euc = MAE
                  rmse_euc = RMSE;
run;

/*  結果表示  */
title "Systolic における KNN 補完（ユークリッド）の評価";
proc print data=stats_euc2 noobs; run;


/* 2) Systolic における Gower-kNN 補完結果評価  */
data eval_knn;
   merge Systolic_mcar(in=a keep=rowid Systolic_true masked)
         Systolic_imp_gower2     (in=b keep=rowid Systolic_knn);
   by rowid;
   if a and b and masked=1 and not missing(Systolic_true) and not missing(Systolic_knn);
   err_knn = Systolic_knn - Systolic_true;
   abs_knn = abs(err_knn);
   sq_knn   = err_knn**2;
run;

proc means data=eval_knn n mean;
   var abs_knn sq_knn;
   output out=metric_knn_gower n=n_obs mean=mae_knn mse_knn;
run;

data metric_knn_gower;
   set metric_knn_gower;
   rmse_knn = sqrt(mse_knn);
   keep n_obs mae_knn rmse_knn;
run;

data metric_knn_gower;
   set metric_knn_gower;
   rename n_obs = impute_obs
                  mae_knn = MAE
                  rmse_knn = RMSE;
run;

title "Systolic における KNN 補完（Gower）の評価";
proc print data=metric_knn_gower noobs; run;
title;


/* 3) BP_Status におけるユークリッド-KNN 補完結果評価  */
/*  補完後のデータと真値を突き合わせ  */
data eval_cat;
   merge BPS_mcar(in=a keep=rowid BP_Status_true masked)
         BPS_imp_euc2(in=b keep=rowid BPS_knn);
   by rowid;
   if a and b and masked=1;

   /*  正解/不正解フラグ  */
   if BP_Status_true = BPS_knn then correct=1;
```

```
    else correct=0;
run;

/* Accuracy を算出 */
proc means data=eval_cat n mean;
    var correct;
    output out=acc_stats mean=Accuracy;
run;

title "BP_Status における KNN 補完（ユークリッド）の評価";
proc print data=acc_stats noobs;
    var Accuracy ;
    format Accuracy 6.3;
run;

/* 4) BP_Status における Gower-KNN 補完結果評価 */
/* 補完後のデータと真値を突き合わせ */
data eval_cat;
    merge BPS_mcar(in=a keep=rowid BP_Status_true masked)
          BPS_imp_gower2(in=b keep=rowid BPS_knn);
    by rowid;
    if a and b and masked=1;

    /* 正解/不正解フラグ */
    if BP_Status_true = BPS_knn then correct=1;
    else correct=0;
run;

/* Accuracy を算出 */
proc means data=eval_cat n mean;
    var correct;
    output out=acc_stats mean=Accuracy;
run;

title "BP_Status における KNN 補完（Gower）の評価";
proc print data=acc_stats noobs;
    var Accuracy ;
    format Accuracy 6.3;
run;
```

～ノンパラメトリック検定～
代表的手法及びパラメトリック検定との比較

○牛尾　英孝

（株式会社ピープロ・ジャパン）

Nonparametric Tests: Representative Methods and Comparison with Parametric Tests

Hidetaka Ushio
Ppro-Japan Co.,Ltd.

要旨

ノンパラメトリック検定は，母集団分布に関して特定の分布形を仮定せずに行える検定であり，外れ値に対する影響を受けにくい検定手法であることが知られている．本稿では，代表的なノンパラメトリック検定の手法である「ウィルコクソンの符号付き順位検定」と「ウィルコクソンの順位和検定」について，それぞれの特徴，原理，適応場面及び SAS による解析事例を紹介し，理解を深められるよう丁寧に解説する．またパラメトリック検定との比較についても触れる．

キーワード：パラメトリック検定，ノンパラメトリック検定，ウィルコクソンの符号付き順位検定，ウィルコクソンの順位和検定

1. 緒言

統計解析において，データの分布に基づいて検定方法を選択することは極めて重要である．一般的に用いられるパラメトリック検定は，母集団が正規分布に従うことや等分散性といった強い仮定の上に成り立っており，前提条件を仮定することで，高い検出力が得られる．一方，臨床試験における統計解析では，必ずしもこれらの前提が満たされるとは限らず，サンプルサイズが小規模である場合や，外れ値が存在する場合も少なくない．このような状況下では，パラメトリック検定の適用が困難となり，誤った結論に導かれる可能性がある．この問題を補う手法として，ノンパラメトリック検定が広く用いられており，ノンパラメトリック検定は母集団分布に特定の分布形を仮定せず，順位情報などに基づいて解析を行うため，分布の歪みや外れ値の影響を受けにくい利点がある．本稿では臨床試験における統計解析においても頻繁に利用されている「ウィルコクソンの符号付き順位検定」と「ウィルコクソンの順位和検定」について基礎的な概念や手法の理解を目的とする．

2. パラメトリック検定とノンパラメトリック検定

パラメトリック検定とノンパラメトリック検定の手法例について表 2-1 に示した．母平均の検定や母平均の差の検定は統計的仮説検定の中でも代表的な手法であり，母集団分布を「正規分布」や「二項分布」といった特定の分布を仮定するものをパラメトリック検定と言う．パラメトリック検定では母集団分布の確率分布は事前に定義されるが，母集団分布が仮説検定が仮定する分布から逸脱することもあり，このような状況でパラメトリック検定を用いると，検定の信頼性が損なわれることがある．一方でノンパラメトリック検定は母集団分布に特定の分布を仮定することなしに構築される仮説検定であり，得られた測定値を順位に変換することで分布の形状に依存せず，外れ値の影響を抑制できるという利点がある．なお，ノンパラメトリック検定は特定の確率分布は仮定しないものの，分布の対称性などは仮定する場合がある．

表 2-1：パラメトリック検定とノンパラメトリック検定の手法例

比較対象	パラメトリック検定の例	ノンパラメトリック検定
母平均の検定	対応のある t 検定	ウィルコクソンの符号付き順位検定
母平均の差の検定	対応のない t 検定	ウィルコクソンの順位和検定（Mann-Whitney U 検定）

3　ウィルコクソンの符号付き順位検定

3.1　想定状況事例と具体的な解析計画

具体的な想定事例及び解析計画を下記に示した．

- 抗炎症作用を有する新規開発中の薬剤 A の有効性を評価する．
- 本疾患の患者数は極めて限られており，実施可能性を考慮して，症例数は 4 例とした．
- 薬剤 A 投与前後で炎症マーカーの値を測定し，変化量に差があるかどうかを探索的に評価する．

本想定事例における疾患ではその特性上から，測定値データの分布は正規分布を仮定できない可能性が高く，サンプルサイズが 4 例と小さいことから，ノンパラメトリック検定として，ウィルコクソンの符号付き順位検定を行うこととする．本疾患の患者 i=1,2,3,4 における薬剤 A 投与前の炎症マーカーの測定値を F_i，投与後の測定値を G_i とし，投与前後の差 D_i=(G_i-F_i)の分布の中央値をΔとする．

評価項目：薬剤 A 投与前後の炎症マーカー値の差
帰無仮説 H_0：$\Delta = 0$「薬剤 A の投与前後の差の分布の中央値は 0 である」
対立仮説 H_1：$\Delta \neq 0$「薬剤 A の投与前後の差の分布の中央値は 0 でない」
検定の有意水準は両側 5%とする

ウィルコクソンの符号付き順位検定では検定統計量 W（算出方法については後述）を算出する際の仮定として，薬剤 A 投与前後の差の分布が 0 を中心に左右対称な分布であるという仮定を置く．分布が左右対称な場合，薬剤 A 投与前後の差が 0 の場合，差が正であるものの順位の和と差が負であるものの順位の和は同程度になることが期待でき，統計検定量 W を「差が正であるものの順位の和」とした際，薬剤 A 投与前後の差が 0 の場合に期待される値よりも「どの程度値が大きくなっているのか」を評価する検定手法である．

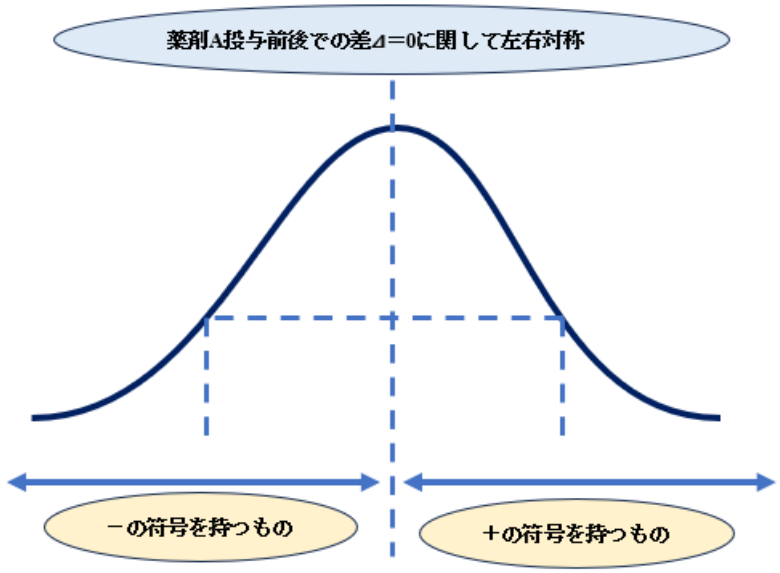


図 3-1：ウィルコクソンの符号付き順位検定における差の分布の仮定

3.2　検定統計量（順位和 W）の算出方法

本想定事例における薬剤 A 投与前後の測定値を表 3-2-1 に，測定値の差（投与後－投与前）と絶対値及び順位を表 3-2-2 に示した．この表をもとにウィルコクソンの符号付き順位検定における検定統計量となる順位和 W の算出過程を解説する．

表 3-2-1：本想定事例における薬剤 A 投与前後の測定値

患者識別番号	薬剤 A 投与前の測定値	薬剤 A 投与後の測定値
001	2.0	1.9
002	3.5	4.7
003	4.0	5.8
004	3.5	4.4

表 3-2-2：薬剤 A 投与前後の測定値の差と順位

患者識別番号	薬剤 A 投与前の測定値	薬剤 A 投与後の測定値	投与前後の変化量	変化量の絶対値	順位
001	2.0	1.9	-1.0	1.0	1
002	3.5	4.7	+1.2	1.2	2
003	4.0	5.8	+1.8	1.8	4
004	3.5	4.4	+1.4	1.4	3

検定統計量 W を「差が正であるものの順位の和」とするため，投与前後の変化量が「＋」であるものの順位の合計値を算出する．本検定においては，W=2+4+3=9 が検定統計量 W（順位和）となる．

3.3　検定統計量 W（順位和）の取り得る値及び検定における p 値の算出

ここでは本想定におけるサンプルサイズ 4 の場合において，検定統計量 W（順位和）の取り得る値をもとにウィルコクソンの符号付き順位検定における p 値を算出する．表 3-3-1 にサンプルサイズ 4 における順位和の取り得る値を示した。順位和の最小値は 0，最大値は 1+2+3+4=10 であり，取り得る値は全部で 11 通りである．

表 3-3-1：サンプルサイズ 4 における順位和の取り得る値

W	1	2	3	4	W	1	2	3	4
0	－	－	－	－	5	－	＋	＋	－
1	＋	－	－	－	6	－	＋	－	＋
2	－	＋	－	－	6	＋	＋	＋	－
3	＋	＋	－	－	7	＋	＋	－	＋
3	－	－	＋	－	7	－	－	＋	＋
4	－	－	－	＋	8	＋	－	＋	＋
4	＋	－	＋	－	9	－	＋	＋	＋
5	＋	－	－	＋	10	＋	＋	＋	＋

※投与前後の変化量（投与後－投与前）が正であるものを＋，負であるものを－と表記

表 3-3-1 で得られた検定統計量が取り得る値から確率を付与し，表 3-3-2 に示した．

表 3-3-2：順位和 W の取り得る値とその確率

W	0	1	2	3	4	5	6	7	8	9	10
P	1/16	1/16	1/16	2/16	2/16	2/16	2/16	2/16	1/16	1/16	1/16

得られた測定値における検定統計量は W=2+4+3=9 であることから，両側 p 値=(1+1+1+1)/16=0.25 となる．本検定における帰無仮説 H_0：$\Delta = 0$「薬剤 A の投与前後の差の分布の中央値は 0 である」は有意水準 5%では棄却されない．本想定では事前に第 1 種の過誤及び第 2 種の過誤を制御してサンプルサイズ設計を行ったものではなく，実施可能性を考慮したものであるから，得られた p 値は名目上のものではあるが，理論的にはサンプルサイズが 4 の場合，ウィルコクソンの符号付き順位検定では有意水準両側 5%で有意差が認められることはない．表 3-3-3 はサンプルサイズが 5 例の場合の順位和 W の取り得る値とその確率を示したものである．順位和 W=0 または 15 の場合でも両側 p 値=(1+1)/32=0.0625 となるので，サンプルサイズが 5 例の場合でも帰無仮説は棄却できない．ウィルコクソンの符号付き順位検定では有意差が認められる最小のサンプルサイズは理論上では 6 例となる．

表 3-3-3：順位和 W の取り得る値とその確率

W	0	1	2	3	4	5	6	7
P	1/32	1/32	1/32	2/32	2/32	3/32	3/32	3/32
W	8	9	10	11	12	13	14	15
P	3/32	3/32	3/32	2/32	2/32	1/32	1/32	1/32

3.4　SAS による解析事例

本想定事例において SAS を用いた解析結果を以下に示す.

```
/*  本想定における薬剤A投与前後の変化量測定  */
data wk1;
   input Subject $ Pre Post;     /*  薬剤A投与前をpre  投与後をpost */
   value = Post - Pre;
   datalines;
001 2.0 1.9
002 3.5 4.7
003 4.0 5.8
004 3.0 4.4
;
run;

/* Wilcoxon符号付き順位和検定  */
proc univariate data=wk1;
    var value;
run;
```

表 3-4-1：Univariate Procedure における出力結果

位置の検定 H0: Mu0=0				
検定	統計量		p 値	
Student の t 検定	t	2.615283	Pr > |t|	0.0793
符号検定	M	1	Pr >= |M|	0.6250
符号付順位検定	S	4	Pr >= |S|	0.2500

表 3-3-2 から算出した両側 p 値は符号付順位検定における p 値（Pr>=|S|）へ出力されていることが確認できる.

4　ウィルコクソンの順位和検定

4.1　想定状況事例と具体的な解析計画

具体的な想定事例及び解析計画を下記に示した．

- 希少な遺伝子変異を有する患者を対象とした治療薬として，既存治療薬の薬剤 X と新規開発中の薬剤 Y の有効性を評価する．
- 対象となる遺伝子変異を保有する患者数は限られており，実施可能性を考慮して，各群への割付は 3 例ずつ合計 6 例とした．
- 観察期間終了時における対象患者の血液中のバイオマーカーZ の濃度を測定し，薬剤 X 群と Y 群とで，母平均に差があるかを探索的に評価する．

本想定事例における疾患ではその特性上から，測定値データの分布は正規分布を仮定できない可能性が高く，サンプルサイズが各群 3 例と小さいことから，ノンパラメトリック検定として，ウィルコクソンの順位和検定を行うこととする．薬剤 X 群 i=1,2,3 における観察期間終了時のバイオマーカーZ の濃度値の分布を P(x)，薬剤 Y 群 j=1,2,3 における観察期間終了時のバイオマーカーZ の濃度値の分布を Q(x)とするとき，分布の関係を「$P(x) = Q(x-\delta)$」と仮定する．分布のイメージを図 4-1 に示した．

評価項目：観察終了時のバイオマーカーZ の濃度測定値
帰無仮説 H_0：$\delta = 0$「薬剤 X 群と薬剤 Y 群の分布の位置シフト $\delta = 0$ である」
対立仮説 H_1：$\delta \neq 0$「薬剤 X 群と薬剤 Y 群の分布の位置シフト $\delta \neq 0$ である」
検定の有意水準は両側 5%とする

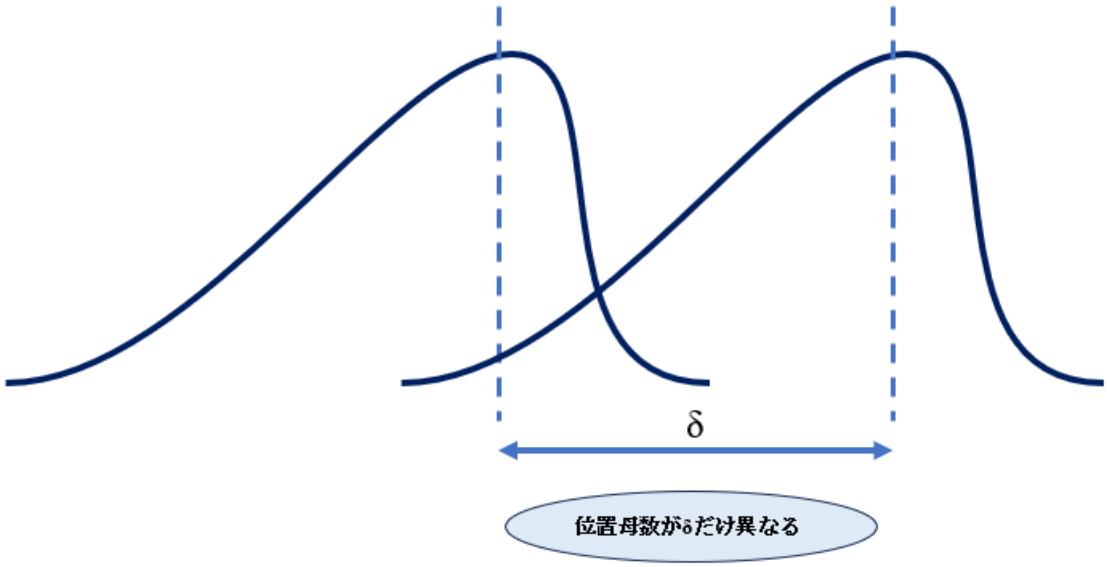


図 4-1：薬剤 X 群と Y 群の観察期間終了時の濃度値 Z の分布のイメージ

4.2　検定統計量 W（順位和）の算出

表 4-2-1：本想定事例における観察期間終了時のバイオマーカーZ の濃度値と順位

患者識別番号	割付群	バイオマーカーZ の濃度値（mg/mL）	順位
001	X	3.9	1
002	X	4.3	3
003	X	4.8	4
004	Y	4.1	2
005	Y	5.2	5
006	Y	5.5	6

本想定事例における観察期間終了時のバイオマーカーZ の濃度値及び両割付群のデータを統合して小さい値から順位を付与したものを表 4-2-1 に示した．Y 群の順位の合計を検定統計量とし，W=2+5+6=13 が算出される．

4.3　検定統計量 W（順位和）の取り得る値及び検定における p 値の算出

表 4-3-1：薬剤 Y 群の順位和の取り得る値及びその確率

W	順位和の取り得る値	通り	確率
⑥	(①，②，③)	1	1/20
⑦	(①，②，④)	1	1/20
⑧	(①，②，⑤)，(①，③，④)	2	2/20
⑨	(①，②，⑥)，(①，③，⑤)，(②，③，④)	3	3/20
⑩	(①，④，⑤)，(②，③，⑤)，(①，③，⑥)	3	3/20
⑪	(①，④，⑥)，(②，③，⑥)，(②，④，⑤)	3	3/20
⑫	(①，⑤，⑥)，(②，④，⑥)，(③，④，⑤)	3	3/20
⑬	(②，⑤，⑥)，(③，④，⑥)	2	2/20
⑭	(③，⑤，⑥)	1	1/20
⑮	(④，⑤，⑥)	1	1/20

得られた測定値における検定統計量は W=2+5+6=13 であることから，Exact な両側 p 値=2×(1+1+2)/20=0.4 となる．本検定における帰無仮説 H_0：δ＝0「薬剤 X 群と薬剤 Y 群の分布の位置シフト δ＝0 である」は有意水準 5%では棄却されない．本想定ではウィルコクソンの符号付き順位検定と同様に得られた p 値は名目上のものではあるが，理論的にはサンプルサイズが各群 3 の場合，ウィルコクソンの順位和検定では有意水準両側 5%で有意差が認められることはない．

4.4　SAS による解析事例

本想定事例においてSASを用いた解析結果を以下に示す.

```
/* 本想定事例におけるデータの作成 */
  data wk2;
     input Subject Group $ value;
     datalines;
  1 X 3.9
  2 X 4.3
  3 X 4.8
  4 Y 4.1
  5 Y 5.2
  6 Y 5.5
  ;
  run;
/* ウィルコクソン順位和検定の実行 */
  proc npar1way data=wk2 wilcoxon;
        class Group;
        var value;
        exact wilcoxon;     /* 連続補正をしない Exact 検定 */
  run;
```

表 4-4-1：Npar1way Procedure における出力結果

変数 value に対する Wilcoxon スコア (順位和)
分類変数：Group

Group	N	スコアの合計	H0 のもとでの期待値	H0 のもとでの標準偏差	平均スコア
X	3	8.0	10.50	2.291288	2.666667
Y	3	13.0	10.50	2.291288	4.333333

Wilcoxon の順位和検定 (2 標本)

統計量 (S)	Z	Pr < Z	Pr > |Z|	t 近似 Pr < Z	t 近似 Pr > |Z|	正確 Pr <= S	正確 Pr >= |S-Mean|
8.0000	-0.8729	0.1914	0.3827	0.2113	0.4227	0.2000	0.4000

Z には 0.5 の連続性の補正が含まれています。

表 4-3-1 から算出した Exact な両側 p 値は Pr>=|S-Mean|へ出力されていることが確認できる.

5　パラメトリック検定とノンパラメトリック検定の検出力の違い

パラメトリック検定として t 検定，ノンパラメトリック検定としてウィルコクソンの順位和検定の両者の検出力を比較する簡易的なシミュレーションプログラムの概要と結果（図 5-1）を下記へ示した.

```
%let n_iter = 100000;     /* シミュレーション回数 */
%let n      = 20;         /* 各群サンプルサイズ */
%let alpha  = 0.05;       /* 有意水準 */

data sim_data;     /* データ生成 */
    call streaminit(20250924);
    do iter = 1 to &n_iter;
        do id = 1 to &n;
            group = "A";
            value = rand("T",2);     /* 群A: 自由度2のt分布から乱数を生成 */
            output;
        end;
        do id = 1 to &n;
            group = "B";
            value = rand("T",3);     /* 群B: 自由度3のt分布から乱数を生成 */
            output;
        end;
    end;
run;

ods output TTests=ttest_p;     /* t検定 */
proc ttest data=sim_data;
    by iter;
    class group;
    var value;
run;
data ttest_p;
    set ttest_p;
    where Method="Satterthwaite";
    keep iter Probt;
    rename Probt=p_value_t;
run;

ods output WilcoxonTest=wilcox_p;     /* Wilcoxon順位和検定 */
proc npar1way data=sim_data wilcoxon;
    by iter;
    class group;
    var value;
run;
```

```
data wilcox_p;
    set wilcox_p;
    keep iter tProb2;
    rename tProb2=p_value_w;
run;
data results;
    merge ttest_p wilcox_p;
    by iter;
    sig_t   = (p_value_t < &alpha);     /* 有意差判定（α=0.05） */
    sig_w   = (p_value_w < &alpha);
run;
proc means data=results noprint;     /* 有意差の回数を集計 */
    var sig_t sig_w;
    output out=sig_summary sum=NumSig_t NumSig_w;
run;
```

母集団分布が正規分布ではない場合（100000 回シミュレーション中）

t検定の有意差回数	Wilcoxon検定の有意差回数
4006	4043

図 5-1：母集団分布が正規分布でない場合のシミュレーション結果

パラメトリック検定（t 検定）は標本平均を検定統計量に用いていること，ノンパラメトリック検定（ウィルコクソンの順位和検定）は検定統計量を順位和として算出することから，外れ値にロバストである．外れ値による影響が補正されている分，本シミュレーションの条件では t 検定よりも検出力が高くなっている．

6　まとめ

ノンパラメトリック検定は，母集団分布を仮定しない検定であり，標本に外れ値が含まれているような場合にロバストな検定法である．一方で，分布に関する情報を用いないので，特定の分布の元での一様最強力検定（正規母集団での小標本に対する t 検定など）に比べると検出力が低下する．

7　参考文献

1. - 晴れ時々SAS Wilcoxon の符号付き順位和検定_proc univariate - https://dengonmemo.blogspot.com/2018/01/wilcoxonproc-univariate.html（最終アクセス 2025/09/01）
2. 竹村彰通.『新装改訂版 現代数理統計学』学術図書出版社（2022）12. ノンパラメトリック検定
3. 久保川達也.『データ解析のための数理統計入門』共立出版（2024）16.5. ノンパラメトリック検定
4. 浜田知久馬.『新版 実用 SAS 生物統計ハンドブック [SAS9.4 R3.2.0 対応]』サイエンス社（2017）

外部データから生成した
Mixture priorを用いたBayesian MMRMのシミュレーションの実装

○東川　正晃[1,3]、遠藤　麻衣[2,3]

([1]EAファーマ株式会社、[2]田辺三菱製薬株式会社、
[3]日本製薬工業協会 医薬品評価委員会 データサイエンス部会)

Implementation of Bayesian MMRM Simulation Using Mixture Prior Formed from External Clinical Trial Data

Masaaki Higashikawa, Mai Endo
[1]EA Pharma Co.,Ltd., [2]Mitsubishi Tanabe Pharma Corporation, [3]Data Science Expert Committee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association

要旨

医薬品の開発において、実施する試験以外のデータソースから得られた外部データの利活用を試みる動きが活発化している。小児医薬品開発においても、成人試験データ等を外挿することにより小児集団への曝露を減らし、かつ効率的な開発を実施できることが期待されている。一方で、外部データの利用においてはいくつか課題も存在する。利用したい参照集団と対象 集団の治療効果が十分に類似していない状況も存在し、データの質や信頼性等の観点から様々な潜在バイアスが懸念される。そのため、外部データを利用する試験計画時にはバイアスや第一種の過誤確率を含む試験の動作特性をシミュレーション実験にて評価する必要がある。本稿では、成人の第 III 相試験の要約データを用いて治療効果の外挿を行った、小児の肺動脈性肺高血圧症患者を対象とした第 III 相ランダム化二重盲検プラセボ対照の並行群間試験における解析方法を参考に、mixture prior を用いたベイズ流の mixed model for repeated measures(MMRM) の動作特性をシミュレーションにて SAS で検討した結果を紹介する。

キーワード：小児用医薬品開発における外挿，Bayesian MMRM，mixture prior，mcmc プロシジャ

1.　はじめに

医薬品の開発において、実際の医療環境下で取得されたデータを含む、実施する試験以外のデータソースから得られたデータの利活用を試みる動きが活発化している[1]。例えば、ICH E8(R1)臨床試験の一般指針にて、過去の臨床研究、全国的な死亡データベース、疾患及び薬剤レジストリ、医療保険請求データ、並びに日常的な医療行為からの医療及び管理の記録を含む外部データの二次利用について記載され、試験計画時に考慮すべき内容が明記された[2]。

外部データの利用は、生命を脅かす希少疾患や、不可逆的な転帰を辿る疾患、小児医薬品開発等、同時対照群の設定が倫理的な観点から困難な場合に薬効を評価する方法の 1 つとして考えられている[3]。

小児医薬品開発はこの外部データの利用が特に注目されている領域の 1 つである。

小児集団は脆弱なサブグループであり、過度の危険から保護する特別な配慮が必要な集団である[4,5]。成人試験データ等を小児試験データに外挿することにより、小児集団全体への曝露を少しでも減らし、かつ効率的な小児医薬品開発が行えることが期待できる。

特に、小児用医薬品開発における外挿に関するガイドラインである ICH-E11A(2025/8 時点 Step 4)では、mixture prior 等のベイズ流アプローチについても言及しており、小児医薬品開発における外部データの利用に対する期待はますます高まっている[6]。

ベイズ流アプローチによる外部データの利用は、既存の情報(専門家の意見、ヒストリカルデータ等)を事前分布として導入し、情報のより効率的な利用を通じてサンプルサイズを大幅に削減、或いは統計的検出力を向上できる可能性を提供する。この結果、試験実施にかかるコストの削減、及び対照群への割付を減らすといった倫理的側面への考慮が期待される。一方で、外部データの利用においてはいくつか課題も存在する。外部データの候補となるデータは、二次利用を前提とせずに設計・運用されていることも多く、実施する試験の目的に十分に合致しないことがある。また外部データとして利用したい参照集団と対象となる集団の治療効果が十分に類似していない状況も存在する。このような場合、データの質や信頼性等の観点から様々な潜在バイアスの存在が懸念される。これらのバイアスに対処するために、外部データを利用する試験計画時には、バイアスや第一種の過誤確率を含む試験の動作特性をシミュレーション実験にて評価する必要がある。

本稿では、成人試験の要約データを用いて治療効果の外挿を行った小児 PAH(pulmonary arterial hypertension：肺動脈性肺高血圧症)患者を対象とした第 III 相ランダム化二重盲検プラセボ対照の並行群間試験を参考に、ベイズ流アプローチを用いる試験の動作特性をシミュレーションにて検討する方法について紹介する。

2. 参考にした小児臨床試験(NCT01824290)の概要

Ivy et al. (2021) [7,8] で示された臨床試験は小児 PAH(pulmonary arterial hypertension：肺動脈性肺高血圧症)患者を対象とした第 III 相ランダム化二重盲検プラセボ対照の並行群間試験(NCT01824290)[9](以下、本試験)である。本試験の概要を表 1 に示す。本試験は 6~17 歳の小児患者を対象としており、tadalafil 群と placebo 群に 1: 1 で割り付けられる。試験開始時の目標被験者数は検出力 80%を確保するために各群 67 例と設定していたが、登録状況が芳しくなかったため試験開始後に各群 17 例に変更されている。主要評価項目は 24 週時の 6 分間歩行距離のベースラインからの変化量である。主解析は mixed model for repeated measures (MMRM)を用いた群間の比較を計画していたが、被験者数不足により十分な検出力が確保できていない。これに伴い、実施済み成人臨床試験において、本試験の小児患者と背景情報が類似している成人患者データの要約を外部データとして利用した Bayesian MMRM を補足的解析として行うように計画を変更している。本稿では補足的解析である Bayesian MMRM の解析方法を参考にしたシミュレーションを実施し、さまざまな条件下における性能評価を行う。

表 1：参考にした臨床試験の概要

試験標題	肺動脈性肺高血圧症の小児第 III 相試験
目的	ERA(エンドセリン受容体拮抗薬)による治療を受けている、肺動脈性肺高血圧症 6~17 歳の小児患者を対象とした tadalafil の有効性の検証

試験期間	30 ヶ月(二重盲検期 6 ヶ月、Open-label 期 24 ヶ月)
投与群	プラセボ群、試験薬群
割付比	1:1
主要評価項目	24 週時の 6 分間歩行距離のベースラインからの変化量
解析方法	主解析： MMRM(検定は実施しない*) 補足的解析*：Bayesian MMRM *計画から変更した箇所

3. 解析方法の詳細

Bayesian MMRM ではすべてのパラメータ(固定効果パラメータおよび共分散パラメータ)は確率変数であり、それらに対する事前分布を設定する必要がある。

Bayesian MMRM で用いる事前分布として本試験では懐疑的な設定の skeptical prior と、実施済みの成人臨床試験結果に基づいて構成された adult prior の混合分布に基づく mixture prior (Ye and Travis (2017))を採用している [10]。このような mixture prior を採用することで、成人臨床試験との類似性が低い場合に、成人試験データに基づいて構成された事前分布を単体で用いる場合よりも type I error rate のインフレを抑えることが期待できる。

解析モデル：

時点、投与群と時点の交互作用を固定効果とし、共分散構造を unstructured と仮定する以下のモデルを考える。

$$\boldsymbol{y_i} = \mathbf{X_i}\boldsymbol{\beta} + \boldsymbol{\varepsilon_i}$$

ここで、$\boldsymbol{y_i}$は被験者iの応答変数ベクトル、$\boldsymbol{X_i}$は被験者iのデザイン行列で、要素にカテゴリカルな説明変数の水準を表す定数か量的な説明変数の測定値を含んでいる。$\boldsymbol{\beta}$はパラメータベクトルであり、治療効果に関連のある$\boldsymbol{\beta_{rt}}$と治療効果に関連のない$\boldsymbol{\beta_{nrt}}$で構成されている。誤差ベクトル$\boldsymbol{\varepsilon_i}$は被験者$i$の各時点の誤差($\boldsymbol{\varepsilon_i} = [\boldsymbol{\varepsilon_{i1}}, \cdots, \boldsymbol{\varepsilon_{iT}}]$)であり、平均 0、共分散行列$\boldsymbol{\Sigma}$をもつ多変量正規分布に従うと仮定する。

事前分布：

共分散行列$\boldsymbol{\Sigma}$の事前分布として自由度T、スケールパラメータ 1 の逆ウィシャート分布$\text{Inv} - \text{Wishart}_T(\mathbf{I})$を用いる。治療効果に関連のない項である$\boldsymbol{\beta_{nrt}}$(本試験では時点の項)については各々独立に適当な無情報事前分布を設定する。

成人データ借用のための治療効果の関連のある項である$\boldsymbol{\beta_{rt}}$(本試験では投与群と時点の交互作用項)については mixture prior を用いる。

Skeptical prior は以下の多変量正規分布で構成される。

$$N(0, \boldsymbol{\Sigma_S}) = N\left(\begin{bmatrix}0\\ \vdots \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_S & \cdots & 0\\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_S\end{bmatrix}\right)$$

Skeptical prior は小児では全く効果がなく、時点間の相関もないことを表している事前分布である。これは type I error rate をコントロールするためのコンポーネントであり、type I error rate を妥当なレベルに抑えるために、シミュレーションに基づいて適切に設定する。本試験ではσ_Sが 9 から 100 の範囲で検討され、以下のシミュレーション実施時には$\sigma_S = 9$と設定されている。

また、adult prior は以下の多変量正規分布で構成される。

$$N(\boldsymbol{\mu_A}, \boldsymbol{\Sigma_A}) = N\left(\begin{bmatrix}\mu_{A1}\\ \vdots \\ \mu_{AT}\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \cdots & \sigma_{1T}\\ \vdots & \ddots & \vdots \\ \sigma_{1T} & \cdots & \sigma_{TT}\end{bmatrix}\right)$$

ここで$\boldsymbol{\mu_A}$及び$\boldsymbol{\Sigma_A}$は実施済みの成人臨床試験結果に基づいて形成される。

上記 2 つのコンポーネントに基づき、$\boldsymbol{\beta_{rt}}$の事前分布を以下のように設定する。

$$\boldsymbol{\beta_{rt}} \sim (1-w)N(0, \boldsymbol{\Sigma_S}) + wN(\mu_A, \boldsymbol{\Sigma_A})$$

混合割合$\boldsymbol{w}$は任意の値を用いる。なお本試験では複数の医学専門家へアンケートを行い、その結果に基づき重み 0.8 を用いている。

上記モデルの理解のため、各モデル及びパラメータの関係を図 1 に示した。

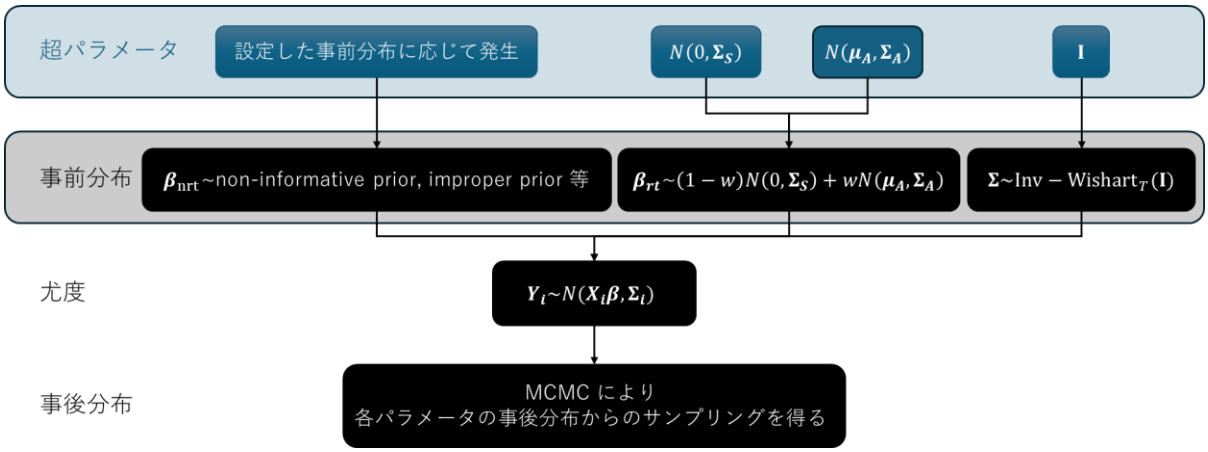


図 1：モデル及びパラメータの関係性

Decision criteria：

本試験では主要評価項目である 24 週時の投与群と時点の交互作用項のパラメータβ_Tについて以下の判定基準を用いている。

$$\Pr(\beta_T > 0) > 0.95$$

4. シミュレーション

4.1. シミュレーションの目的

本シミュレーションでは、mixture prior を用いた Bayesian MMRM の動作特性を確認するため、複数のシナリオを設定し、条件設定に基づき type I error rate、検出力、バイアス、mean squared error(MSE)、effective sample size(ESS)を用いた性能評価を行う。

4.2. シミュレーションの設定

小児の治療効果について以下の 3 つのシナリオを設定した。

1. Null：プラセボと同じ効果であった場合
2. 50% effect of optimistic：成人の半分程度の効果があった場合
3. Optimistic：成人と同程度の効果があった場合

それぞれのシナリオの下で、ベースライン時点を含まない 4 時点分の変化量のデータを多変量正規分布に従う乱数を用いて発生させて小児の仮想データを作成した。Optimistic なシナリオは以下の多変量正規分布に基づき実薬群のシミュレーションデータを発生させている。この分布は、本試験の成人データを基に形成された分布から 3、5 時点目を除いたものである。50% effect of optimistic のシナリオでは optimistic なシナリオの平均ベクトルの各値を半分に、null のシナリオ及びプラセボ群では全て 0 に変更している。各シナリオにおける共分散行列は同一のものを仮定する。

$$N\left(\begin{bmatrix}8\\15\\21\\24\end{bmatrix},\begin{bmatrix}2078 & 1463 & 1075 & 950\\1463 & 2266 & 1543 & 1450\\1075 & 1543 & 2858 & 1950\\950 & 1450 & 1950 & 2858\end{bmatrix}\right)$$

解析モデルとして時点、投与群と時点の交互作用を固定効果とし、共分散構造を unstructured と仮定するモデルを用いる。誤差ベクトル$\boldsymbol{\varepsilon_i}$の共分散行列$\boldsymbol{\Sigma}$の事前分布として自由度 4、スケールパラメータ 1 の逆ウィシャート分布$\text{Inv}-\text{Wishart}_4(\mathbf{I})$を用いる。治療効果に関連のない項(時点)の$\boldsymbol{\beta_{nrt}}$の事前分布は improper prior として区間$[-\infty,\infty]$の一様分布を用いる。治療効果に関連のある項(投与群と時点の交互作用)の$\boldsymbol{\beta_{rt}}$は skeptical prior と adult prior から構成される mixture prior を事前分布として用いる。mixture prior は混合割合 w を 0.5、0.8 の計 2 通りを設定した(図 2)。各事前分布の詳細は以下の通りとする。

Skeptical prior:

$$N(0,\boldsymbol{\Sigma_S}) = N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix},\begin{bmatrix}81 & 0 & 0 & 0\\0 & 81 & 0 & 0\\0 & 0 & 81 & 0\\0 & 0 & 0 & 81\end{bmatrix}\right)$$

Adult prior:

$$N(\boldsymbol{\mu_A},\boldsymbol{\Sigma_A}) = N\left(\begin{bmatrix}8\\15\\21\\24\end{bmatrix},\begin{bmatrix}67 & 69 & 56 & 45\\69 & 78 & 71 & 63\\56 & 71 & 78 & 79\\45 & 63 & 79 & 90\end{bmatrix}\right)$$

Mixture prior:

$$\boldsymbol{\beta_{rt}} \sim (1-w)N(0, \boldsymbol{\Sigma_S}) + wN(\mu_A, \boldsymbol{\Sigma_A})$$

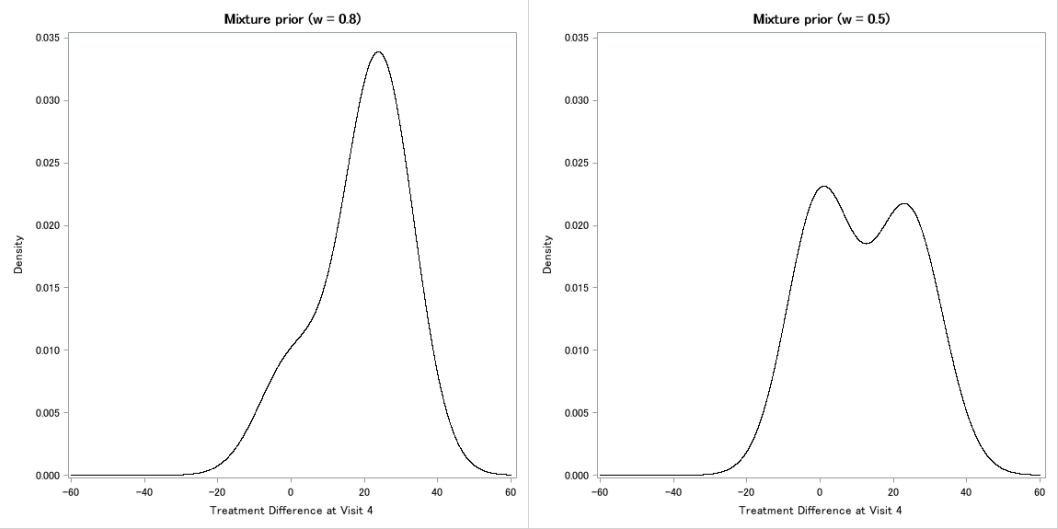


図 2　4 時点目の治療効果の mixture prior (左図：*w*=0.8 右図：*w*=0.5)

$\boldsymbol{\beta_{rt}}$の事前分布として、混合割合 *w* が 0.5 及び 0.8 の mixture prior に加え、成人試験のデータを利用しない想定として improper prior を用いた場合の計 3 通りの条件を設定した。被験者数は各群 17 例、51 例の 2 通り設定し、シナリオ 3 通り×事前分布 3 通り×被験者数 2 通りの合計 18 通りのシミュレーションをそれぞれ 3000 回実施した。得られた事後分布から、最終時点である 4 時点目において実薬群がプラセボ群よりも変化量が大きい確率を求め、decision criteria の 0.95 を超えた割合を type I error rate (シナリオ 1)あるいは検出力(シナリオ 2、3)とする。

ESS は以下の式により得られる [11]。

$$ESS = N\left(\frac{\mathrm{V}_1(\beta_T)}{\mathrm{V}_2(\beta_T)} - 1\right)$$

ここでNは小児臨床試験の被験者数、$\mathrm{V}_1(\beta_T)$は improper prior を用いて算出したβ_Tの分散、$\mathrm{V}_2(\beta_T)$は mixture prior を用いて算出したβ_Tの分散である。

4.3.　SAS を用いたコードのポイント

MCMC プロシジャは、mcmc 法を用いたベイズ流解析を実装するためのプロシジャである。標準的な確率分布が搭載されているだけでなく、任意の確率分布についても MCMC プロシジャ内でプログラミング文を記述することで実装可能であり、汎用的なベイズ流解析が実施可能なプロシジャである。本稿で実装する mixture prior も MCMC プロシジャ内でプログラム文を記述することで実装している。

MCMC プロシジャによる実装例を示す。Bayesian MMRM の基本的な実装については伊庭ら(2023)[12] を参照されたい。ここでは mixture prior(*w* = 0.8)の実装部分についてのみ抜粋して説明する。mixture prior の実装は ARRAY 文及び CALL ルーチンを用いて行っている。なお、mixture prior の指定のためのプログラミング文は最初に一度だけ実行しておけば良いため、BEGINCNST/ENDCNST 文で囲んでいる。事前分布を自身で指定する場合は対数スケールでの記述が必要となるため、mixture prior を対数スケールで記述し(実装例では lp)、治療効果に関連のある項(投与群と時点の交互作用、実装例では beta_tm)の事前分布として設定している。任意

の事前分布を指定する場合は PRIOR ステートメントの確率分布で general を指定することで実装できる。なお、本稿では事前分布の実装に際し MCMC プロシジャ内で直接 CALL ルーチンを用いているが、FCMP プロシジャで関数を定義して MCMC プロシジャで呼び出すことで実装することも可能である。

```
/*adult component of the mixture prior*/
array muA[4] muA1-muA4 (8 15 21 24);
array SA[4, 4] SA1-SA16 (67 69 56 45,
                         69 78 71 63,
                         56 71 78 79,
                        45 63 79 90);

/*skeptical component of the mixture prior*/
array muS[4] muS1-muS4 (0 0 0 0);
array SS[4, 4] SS1-SS16 (81 0 0 0 ,
                         0 81 0 0 ,
                         0 0 81 0 ,
                         0 0 0 81);

/* mcmc プロシジャ内でmixture priorの計算用の配列を用意*/
array DIFA[4];
array tDIFA[1, 4];
array invSA[4, 4];
array XMA[1, 4];
array XMXA[1];
array DIFS[4];
array tDIFS[1, 4];
array invSS[4, 4];
array XMS[1, 4];
array XMXS[1];

/*CALLルーチンを用いて各配列(事前分布の算出に用いる)の初期値を設定*/
begincnst;
 call identity(S);
 call det(SA, dSA);
 call inv(SA, invSA);
 call det(SS, dSS);
 call inv(SS, invSS);
endcnst;
```

```
/*beta_tmはmixture priorを事前分布とするパラメータ群の配列*/
parms beta_tm {8 15 21 24};

/*mixture priorの計算*/
beginnodata;
 n=4;
 w=0.8;

 call subtractmatrix(beta_tm, muA, DIFA);
 call transpose(DIFA, tDIFA);
 call mult(tDIFA, invSA, XMA);
 call mult(XMA, DIFA, XMXA);

 call subtractmatrix(beta_tm, muS, DIFS);
 call transpose(DIFS, tDIFS);
 call mult(tDIFS, invSS, XMS);
 call mult(XMS, DIFS, XMXS);

 const_adlut=1/((CONSTANT('PI')**(n/2))*(dSA**(1/2)));
 adlut_pdf=const_adlut*exp(-0.5*XMXA[1]);

 const_skeptical=1/((CONSTANT('PI')**(n/2))*(dSS**(1/2)));
 skeptical_pdf=const_skeptical*exp(-0.5*XMXS[1]);

 lp=log(w*adlut_pdf + (1-w)*skeptical_pdf);
endnodata;

/*上記で導出したmixture priorを事前分布として指定*/
prior beta_tm ~ general(lp);

/*他の事前分布を指定*/
/* general(0)は区間[-∞,∞]の一様分布に相当*/
prior beta_m: ~ general(0);
prior R ~ iwish(4, S);

/*解析モデル及び推定値の指定
今回のモデルではlsmの指定は不要だが拡張性も考慮して設定している*/
beginnodata;
```

```
 do j=1 to 4;
  lsm1[j] = beta_m[j] + beta_tm[j];
  lsm2[j] = beta_m[j];
  lsmd[j] = beta_tm[j];
 end;
endnodata;

do j=1 to 4;
 if trt=1 then mu[j] = beta_m[j] + beta_tm[j];
 else mu[j] = beta_m[j];
end;

model y ~ mvn(mu, R);
```

4.4. シミュレーションの結果

4.2 で記載したシミュレーションの結果を以下の表に示す。

・Type I error rate、検出力

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8
17 例/群	null	5.0%	6.1%	21.2%
	50% effect of optimistic	15.3%	17.9%	40.5%
	optimistic	37.4%	36.3%	65.1%
51 例/群	null	4.9%	7.8%	17.4%
	50% effect of optimistic	30.1%	37.8%	58.6%
	optimistic	72.6%	77.1%	90.6%

・バイアス

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8
17 例/群	null	-0.3	7.0	12.1
	50% effect of optimistic	-0.4	0.3	4.7
	optimistic	0.2	-6.3	-2.6
51 例/群	null	0.0	4.5	7.3
	50% effect of optimistic	0.1	1.1	3.9
	optimistic	0.0	-2.4	-0.9

・MSE

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8
17 例/群	null	323.2	113.4	210.2
	50% effect of optimistic	337.3	76.1	81.9
	optimistic	341.6	111.7	55.4
51 例/群	null	113.2	71.2	108.7
	50% effect of optimistic	109.0	66.6	64.1
	optimistic	114.5	58.5	35.4

・ESS

被験者数	シナリオ	improper prior	mixture prior w=0.5	mixture prior w=0.8
17 例/群	null	-	70.4	67.4
	50% effect of optimistic	-	62.6	75.2
	optimistic	-	61.2	86.4
51 例/群	null	-	78.2	64.4
	50% effect of optimistic	-	55.0	76.5
	optimistic	-	75.7	106.2

4.5. シミュレーションの結果の考察

Type I error rate について、成人試験のデータを利用しない improper prior を用いた場合は名義水準(5%)通りの値となったが、成人試験のデータを外挿した事前分布を用いた場合は被験者数に依らず名義水準を超える傾向となった。$w = 0.5$ の mixture prior を用いた場合の Type I error rate は 6.1〜8.8%となり、$w = 0.8$ の mixture prior の場合は 17.4〜21.2%と大きなインフレが見られた。

Type I error rate のインフレは成人試験と小児試験の治療効果に差異が生じたことによるものと考えられ、成人試験の利用の程度が大きい(混合割合 w が大きい)ほどその傾向がより顕著に出ている。また、これは被験者数に依らず同様の傾向であった。

検出力について、全てのシナリオ及び被験者数で improper prior が最も低かった。mixture prior を用いた場合はいずれの場合でも improper prior と比較して検出力が増加した $w = 0.8$ の mixture prior を用いた場合が最も検出力が高く、特に optimistic のシナリオで検出力が大きく増加していた。この事から、事前分布の成人試験の利用の程度が大きい、または成人試験と小児試験の治療効果の差異が小さい場合に検出力が増加することが示唆された。

バイアスについて、improper prior はほぼ 0 であったが、mixture prior では improper prior と比べてバイアスが大きくなる傾向が見られた。mixture prior の場合は事前分布の期待値と小児試験の治療効果の差異が小さい場合にバイアスが小さく、反対に乖離が生じている場合にはバイアスが大きくなる傾向となることが示された。

MSE について、全てのシナリオ及び被験者数で improper prior が最も大きかった。mixture prior の場合は成人試験と小児試験の治療効果の差異が小さい場合に MSE が小さく、反対に乖離が生じている場合には MSE が大きくなる傾向となることが示された。

ESS について、$w = 0.8$ の mixture prior では成人試験と小児試験の治療効果の差異が小さいほど ESS は大きくなった。一方で $w = 0.5$ の mixture prior では同様の傾向が見られず、類似性が高い 50% effect of optimistic のシナリオでの ESS は他のシナリオと同程度か小さい値となった。$w = 0.5$ の mixture prior を用いた場合の事後分布の分散は他の事後分布の分散より大きく、これにより類似していることによる ESS の増大を打ち消しているためと考えられる。

事前分布に成人試験のデータを外挿したことで検出力が大きくなることが示されたが、一方でその利用の程度が大きい場合 type I error rate の顕著な増加が見られ、mixture prior の利用に際しては検出力と type I error rate のバランスを十分に評価、検討することが必要となる。

Mixture prior は外部試験の要約データの情報を事前分布に組み込むベイズ流アプローチであり、個別データを必要としないこともメリットの 1 つであると考えられる。一方で個別データを用いることができる場合は、CMP 法等の異なるベイズ流アプローチも利用可能であるため、本手法の実装を検討する際には他の手法との比較検討も併せて行うことができる。

日本製薬工業協会が 2025 年に発行した「ベイズ流試験デザインにおけるシミュレーションの利用法」[13] ではwを確率変数とみなし、本論文の内容に加え、ベータ分布を事前分布として設定する場合についても検討している。また、R を用いて実装するためのポイントも解説しているので参考されたい。

5. 参考文献

1. 厚生労働省. (2021). 承認申請等におけるレジストリの活用に関する基本的考え方. https://www.mhlw.go.jp/web/t_doc?dataId=00tc5763&dataType=1&pageNo=1
2. 厚生労働省. (2022). 「臨床試験の一般指針」の改正について. 医薬審第 1223 第 5 号. https://www.pmda.go.jp/files/000250244.pdf
3. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. "Rare disease の治療効果の推測法."(2022)
4. 厚生労働省. (2000). 小児集団における医薬品の臨床試験に関するガイダンスについて. 医薬審第 1334 号. https://www.pmda.go.jp/files/000156072.pdf
5. 厚生労働省. (2017). 小児集団における医薬品開発の臨床試験に関するガイダンスの補遺について. 医薬審第 1227 第 5 号. https://www.pmda.go.jp/files/000222107.pdf
6. ICH HARMONISED GUIDELINE. (2024). INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE. https://www.pmda.go.jp/files/000270245.pdf
7. Dunbar Ivy, Damien Bonnet, Rolf MF Berger, Gisela M.B. Meyer, Simin Baygani, and Baohui Li. (2021). Efficacy and safety of tadalafil in a pediatric population with pulmonary arterial hypertension: phase 3 randomized, double-blind placebo-controlled study. Pulmonary Circulation, 11(3) 1–8.
8. ClinicalTrials.gov A Study of Tadalafil in Pediatric Participants With Pulmonary Arterial Hypertension (PAH). https://www.clinicaltrials.gov/study/NCT01824290
9. Statistical Analysis Plan H6D-MC-LVHV (v4). https://cdn.clinicaltrials.gov/large-docs/90/NCT01824290/SAP_001.pdf
10. Ye J, Travis J. (2017). A Bayesian approach to incorporating adult clinical data into pediatric clinical trials. Food and Drug Administration, Office Biostatics (DB V and II). https://www.fda.gov/media/107649/download
11. Food and Drug Administration. (2010). Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials
12. 伊庭 克拓、浅野 豊、松嶋 優貴、毛利 誠. (2023). BGLIMM プロシジャおよび MCMC プロシジャによるベイズ流経時測定データ解析. SAS ユーザー総会 2023. https://sas-user2023.ywstat.jp/download.html?n=10&key=paper.
13. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会(2025). ベイズ流試験デザインにおけるシミュレーションの利用法.

TRAJプロシジャを用いた縦断測定データ解析方法

○石渡量太、橋口正行、志賀剛

（[1]東京慈恵医科大学 医学部臨床薬理講座）

Analysis method using Group-based modeling of longitudinal data

Ryota Ishiwata, Masayuki Hashiguchi, Tsuyoshi Shiga
Clinical Pharmacology and Therapeutics, The Jikei University School of Medicine,

要旨

近年、大規模なデータベースが利用できるようになり、より多くの時点を扱うような経時的な変化に関する縦断的データを扱うことができるようになった。複数の時点で測定された縦断データは、仮説検定において検出力を高めるだけでなく、横断データや 2 時点の観測データでは対処できない仮説を検証できる。これまで多くの分析手法が開発され、マルチレベルモデリング（階層線形モデル（HLM），潜在成長（曲線）モデル，混合効果モデル）は、縦断的データをモデル化する潜在成長曲線モデルの代替手段である。

本発表ではこれらのモデルのうち Nagin らにより開発された集団軌跡モデル（GBTM）(Nagin 1999) に焦点を当て、その概要と TRAJ プロシジャの詳細な使用方法と機能に焦点を当てて解説する。さらに、共変量を入れるステートメントやグラフを描画するためのステートメント、オプションについて解説する。

本稿では，この TRAJ プロシジャを用いて，診療データベースの処方箋レコードデータを用いて GBTM を実装する方法を紹介する。

キーワード：経時測定データ，集団軌跡モデル

はじめに

大規模データベースでは、既に収集されているデータを用いて特定の治療法や介入の効果を後ろ向きに評価する研究が実施される。後ろ向きコホート研究は、過去の医療記録からある疾患の発症と特定の治療法の関係を調べるなど、縦断データを用いて時間の経過とともに要因がどのように変化し、結果にどのような影響を与えるかを分析する。多くの研究では、各時点のデータに平均値を算出して時間的な変化を分析する。時間の経過とともにデータに変化が生じる場合、平均値などでまとめた分析では潜在的な（観測することのできない）個人内の差や個人間の差（変動）を単純化してしまうため、これらの情報を適切に分析できていないことがある。(Nguefack, 2020)

例えば、服薬アドヒアランスを評価する場合、「薬所持率」（MPR）や「対象日数の割合」（PDC）(Pednekar 2018)の 2 つの指標を用いて、評価したい期間の平均値の推移を全症例やサブグループで評価する。服薬アドヒアランスには、個人の行動パターン（生活のリズム、通院の利便性など）や個人の考え（自己判断で服用を中止、経済的な理由など）などさまざまな要因による影響が服薬状況に変化をもたらすと考えられる。(池野 2014)　そのように変化のパターンに違いがあると考えられ、時系列の変化に個人差があるのかを検討したい場合、時系列に伴う変化のパターンの個人差をうまく分析する方法を用いる。潜在クラスモデリング手法、成長混合モデリング（GMM）、集団軌跡モデル（GBTM）、潜在クラス分析（LCA）、潜在遷移分析（LTA）などがある。（大橋 2023）

本稿では、集団軌跡モデル（GBTM）に注目して、SAS の TRAJ プロシジャを用いた経時測定データ解析を行う方法を紹介する.

1.　グループベース軌跡モデリング（GBTM ）

軌跡（trajectory）とは、時間経過に伴って変化する様相や増加または減少といった方向性を表す。疫学調査や後ろ向きコホート研究（縦断研究）では、時間経過に伴う変数の軌跡に加えて、変数（共変量）がその軌跡（変化の様相、形状）にどのように影響を及ぼすかについて検討する。縦断データ（繰り返し測定されたデータ）の様相を分析する方法として、階層マルチレベルモデリングと潜在曲線分析（（階層線形モデル（HLM），潜在成長（曲線）モデル，混合効果モデル））があり、グループベース軌跡モデリング（GBTM）もその一つである。（高橋 2015） 階層マルチモデリングと潜在曲線分析では、母集団の平均軌跡を推定し共変量を用いてこの平均の変動を検討する。GBTM は集団内に潜在的な異なるサブグループが存在すると仮定し、それらのグループごとに異なる変化パターンをモデル化する手法です。この GBTM では、個々の患者の治療経過を時系列データとして扱い、関数を当てはめることで、疾患の進行パターンや治療反応のサブグループを確率に基づいて分類することができる。(Nguena Nguefack, Pagé et al. 2020) 近年、GBTM を用いた研究結果が多く報告されている。(Ishiwata, AlAshqar et al. 2024) (Hamada, Thomas et al. 2025)

GBTM は、最大尤度法（Maximum Likelihood Estimation） に基づいて推定する。この方法では、観測されたデータが最も高い確率で説明されるように、軌跡グループの特徴（平均的な傾向や変化のパタ

ーン）を決定します。個人ごとに、どの軌跡グループに属する可能性が高いかを計算し、最も適したグループに分類する。(Nagin 1999)

経時的に繰り返し測定された値の分布：アウトカムの軌跡を表す分布を$P(Y_i|Time_i)$で表す。ランダムベクトル(Y_i)は個人(i)の繰り返し測定された値（縦断的に時間とともに変化するデータ）、ベクトル$Time_i$は個人のアウトカムが測定された時点(i)を表す。ここで J は、なんからの軌跡を辿る母集団分布で未知の J グループによる有限混合モデルと仮定する。GBTM は軌道の母集団分布が未知の J 有限混合モデル（クラスタ）をもち、それぞれのクラスタが混合する確率分布から生じると仮定している。個人(i)の尤度はグループ数(J)によって条件付けられており、以下のように表す。

$$P(Y_i|Time_i) = \sum_{j=1}^{J} \pi^j \, P\left(Y_i \middle| Time_i, j; \beta^j\right)$$

個人(i)の観測データ(Y_i)の尤度は、グループの数 J に依存する。ここでπ^jはグループメンバーシップ(j)に対する確率$(0 \leq \pi^j \leq 1,\ \Sigma_{j=1}^{J}\pi^j = 1)$であり、$P(.)$は j 番目のクラスのモデルにおける応答変数の条件付き確率密度関数である。

グループ(j)に属する個人（i）の(Y_i)の条件付き分布は、グループ固有の軌道の形状を決める未知のパラメータベクトルβ^j、グループ特有の軌跡の形状（1 次、2 次、3 次関数）を指標化する。(軌道は時間を最大 3 次の多項式によってモデル化する。) グループ(j)について、条件付き独立の仮定のもと、評価期間Ｔポイントでの(Y_i)と(Y_{it})のパラメーターの最適化（逐次合成）によって仮定される。

$$P\left(Y_i \middle| Time_{it}, j; \beta^j\right) = \prod_{t=i}^{T} p\left(y_{it} \middle| Time_{it}, j; \beta^j\right)$$

ここで$p(.)$はグループ(j)に属する個人 (i) のタイムポイント(t) における変数y_{it}の条件付き確率分布である。PROC TRAJ ではこの$p(.)$のモデルにいくつかの確率分布：打ち切り正規分布（CNROM)、ゼロ過剰ポアソン分布（ZIP)、ベータ分布、およびベルヌーイ分布（ロジスティックモデル）を指定できる。Y_iの条件付き分布は、未知のパラメータベクトル β^j（グループ固有の軌道の形状を特定する）に依存する。グループ(j)における、条件付き独立を仮定する。

GBTM では推定される軌跡とその軌跡がグループに所属する確率に加え、個人（i）がそれぞれのグループ（J）に属する確率：事後確率推定量（PPGMj）を算出する。ベイズ則に基づいて、事後確率が最大となるパラメータの推定方法（**Maximum A Posteriori estimation**）により$\hat{\beta}^j$と$\hat{\pi}^j$を解を出す。

$$\hat{P}(j|Y_i) = \frac{\hat{P}(Y_i|j_i)\hat{\pi}_j}{\sum_j^J \hat{P}(Y_i|j)\hat{\pi}_j}$$

グループメンバーシップ事後確率(j)は、時間の経過に伴って変化する個人のデータ(Y_i)が軌跡(j)のア

ウトカムである確率である。

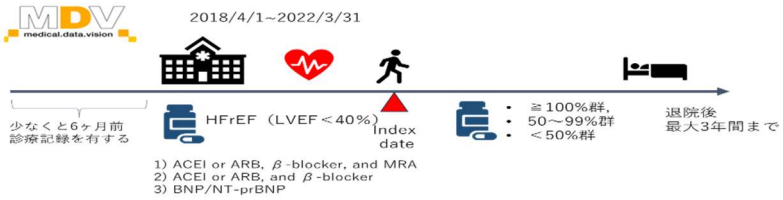
$$\hat{P}(j|Y_i) = \frac{\hat{P}(j|Y_i)}{\sum_j^J \hat{P}(j|Y_i)\hat{\pi}_j}$$

個人(i)がそれぞれのグループ(J)に属する確率の合計は 1 になり、最も高い事後確率（PPGMj）が算出されたグループに割り当てられる。(Maximum Probability Assignment)。内的整合性は、グループの個人(i)の平均事後確率に近似する。そのグループメンバーシップの平均事後確率が＞0.70 ～ 0.80 を超える場合、そのグループは同じ変化パターン（軌跡）を持つ個人のグループとみなす。(Nagin,2025)

2. 本稿で使用する事例データ

本稿では、 MDV 診療データベースを用いた後ろ向きコホート研究のデータを用いる。（石渡,2024）

図 1 試験デザイン（心不全により初回入院した心不全患者を選択）



この事例データから、退院時より少なくとも 1 年間の診療記録を有する患者を対象とし、それらの患者における処方箋（β 遮断薬）の処方パターンを評価する。評価方法として，退院日から monthly（30 日間）期間を区切り、その monthly における処方記録（処方日数分）を 30 日で割り、Monthly PDC を算出した。 Monthly PDC が≧0.8 を服薬アドヒアランス(=1)、<0.80 を服薬 non-アドヒアランス(=0)の二値データと定義した。

表 1 本稿で使用する解析データ (N=19132)

ID	退院後（期間）	処方薬（日数）	PDC	服薬アドヒアランス（PDC≧0.8）
001	1st month	30 日分	30/30 = 1.0	アドヒアランス（1）
	2nd month	15 日分	15/30 = 0.5	ノンアドヒアランス(0)
	3rd Month	0 日分	0/30 = 0.0	ノンアドヒアランス(0)

3. TRAJ プロシジャ

3.1 SAS code

本節では TRAJ プロシジャの基本的な構文を以下に示す。PROC TRAJ は Base SAS, SAS/STAT 分析プロシジャには含まれていないので、以下のリンク （www.andrew.cmu.edu/~bjones）から SAS programs

がダウンロードできる。

```
PROC TRAJ;
      ID; VAR; INDEP;
      MODEL; ORDER; START; WEIGHT;RISK; REFGROUP;TCOV; PLOTTCOV;
      MIN; MAX; RORDER;/* cnorm options */
      IORDER; EXPOS;/* zip options */
      OUTCOME; OMODEL; OCOV;/* outcome model */
      DROPOUT; DCOV; OBSMAR;/* dropout options */
      MULTGROUPS; MULTRISK;/* multi-trajectory */
```

PROC TRAJ 文の主なオプションを以下に示す。

＊Data をサブセットする“where”構文は PROC TRAJ では指定できない。

INPUT ファイル:

DATA　　解析データセット

OUTPUT ファイル:

OUT　　グループ分けとグループに所属する確率

OUTSTAT　　TRAJPLOT macro より算出したパラメータ推定値

OUTPLOT　　軌跡プロットデータ

OUTEST　　パラメータと共分散の行列推定値

ITDETAIL　　モデルの最適化をイテレーティブに最小値を求める

/***/

ID;　　個人識別番号

VAR;　　従属変数　測定時点（例：1 カ月後、2 カ月後..）　例：VAR V1-V12;

INDEP;　　測定時点の変数を指定 例：INDEP T1-T12;;

MODEL;　　従属変数の確率モデルを指定（BETA, CNORM, ZIP, LOGIT）：例：Logit

ORDER;　　多項式：各群の軌跡の関数（0=intercept, 1=linear, 2=quadratic, 3=cubic）

/*確率モデルを指定した場合*/

/* CNORM */

MIN; (CNORM)　　打ち切り最小値 e.g.）1（デフォルト 0）

MAX; (CNORM)　　打ち切りの最大値

RORDER; (CNORM)　　各グループのランダム成長曲線パラメータ（-1=なし，0=切片，　1=1 次関数，2=2 次関数，3=3 次関数）グループに共有のランダム成長曲線パラメータを設定する場合、単一の値を指定する e.g.）RORDER 0 or RODER 0 0 0 0;

/*ゼロ過剰モデル（ZIP）*/

IORDER; (ZIP)　　各グループのゼロ過剰モデルパラメター(0=切片, 1=1 次関数, 2=2 次関数, 3=3 次関数）ポアソン分布を使ったゼロ過剰ポアソンモデルには、どのグループにも-1 を指定する。各グループに共通の過剰モデルを指定する e.g.）IORDER 1 1 1 1 ； or IORDER 2;

EXPOS; (ZIP)　　変数 e.g.）EXPOS E1-E8;

START;　　**パラメータの開始値　e.g.）SAS ログを参照**

RISK;　　独立変数（グループメンバーシップ（群分け）の確率を説明する）を指定 e.g.）教育レベル

REFGROUP;　　独立変数（コントロールレファレンスを指定）　例：デフォルト（0）

TCOV;　　時間依存共変量　e.g.)TCOV C1-C8

PLOTTCOV;　　グループ軌跡を算出する時間依存共変量の値（設定する値は TCOV で指定した変数の数と同じ数にする）e.g.）**PLOTTCOV** 0 0 0 0 1 1 1 1

WEIGHT;　　重み付き尤度関数における重み付け変数（サンドイッチ推定量を用いたパラメータ推定値の共分散行列）

その他の options については以下を参照する。（TRAJ SAS）

3.2　　モデルを選択

GBTM を実施するにあたり、以下の手順によりモデルを選択する。(Troiano , 2023)

1. 医学専門家や参考文献などを基に、適切なグループ分けの数を仮定する。
2. モデルの検討
   i. 検討するグループ数は 1 グループから多くても 7 グループで特定する（各グループは全体の≧5%のサンプル数を有する）
   ii. 軌道の形状（各軌道の一次、二次、および三次関数）を特定する。
3. モデルの適合度をベイズ情報 (BIC)、割当てられた平均事後確率など指標を用いて評価する。
4. グラフ化してその当てはまりを確認して、実質的な解釈などを評価する。

解析結果:TRAJ プロシジャを実行すると，収束診断の結果やプロット，MODEL 文で指定したパラメータおよび パラメータ推定値、事後確率などが出力される。

① SAS code と Output

```
PROC TRAJ
        DATA= PDC OUTPLOT=OP CI95M OUTSTAT=OS
        OUT=POST_OF_all   OUTEST=OE ITDETAIL;
        ID patientid;
        VAR     PDC_beta_1mo  PDC_beta_2mo  PDC_beta_3mo  PDC_beta_4mo
                PDC_beta_5mo  PDC_beta_6mo  PDC_beta_7mo  PDC_beta_8mo
                PDC_beta_9mo  PDC_beta_10mo PDC_beta_11mo PDC_beta_12mo;
                INDEP T1-T12;
        MODEL LOGIT; NGROUPS 4; ORDER  3 3 3 3;RUN;
```

```
%TRAJPLOT(OP,OS,,'',"Adherence",'Beta-blocker with 4 groups')
```

Maximum Likelihood Estimates
Model: Logistic (LOGIT)

Group	Parameter	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
1	Intercept	0.75322	0.07570	9.950	0.0000†
	Linear	-1.21569	0.06757	-17.993	0.0000†
	Quadratic	0.05892	0.01378	4.276	0.0000†
	Cubic	0.00146	0.00075	1.941	0.0522†
2	Intercept	1.64389	0.14565	11.286	0.0000†
	Linear	0.96474	0.08738	11.041	0.0000†
	Quadratic	-0.26547	0.01504	-17.653	0.0000†
	Cubic	0.01318	0.00076	17.436	0.0000†
3	Intercept	1.26819	0.06866	18.469	0.0000†
	Linear	1.11269	0.05631	19.760	0.0000†
	Quadratic	-0.13178	0.01021	-12.908	0.0000†
	Cubic	0.00417	0.00051	8.106	0.0000†
4	Intercept	0.57013	0.10664	5.346	0.0000†
	Linear	-0.96133	0.07513	-12.796	0.0000†
	Quadratic	0.22217	0.01393	15.945	0.0000†
	Cubic	-0.01102	0.00073	-15.056	0.0000†
Group membership					
1	(%)	21.66602	0.30643	70.704	0.0000†
2	(%)	9.70043	0.23667	40.987	0.0000†
3	(%)	60.94793	0.39637	153.764	0.0000†
4	(%)	7.68562	0.26399	29.114	0.0000†

BIC= -68293.42 (N= 229584) BIC= -68269.82 (N= 19132) AIC= -68195.16 ll= -68176.16

表 2 に各グループ(2,3,4,5,6)の PROC TRAJ を実行した解析結果(BIC と各グループの割合(%))を示す。BIC が最も小さい値をもつモデルが優先するため、BIC が改善されなくなるまでグループ分け評価を続ける。各モデルの BIC を用いてそれぞれ比較したいモデルを評価する。その評価方法は、複雑なモデル（より多くの軌道の数を持つモデル）から、よりシンプルなモデル （軌道の数が少ないモデル）の BIC の差（対数ベイズ係数の推定値）で評価する。(Jones et al., 2001)
対数ベイズ係数の推定値を解釈する指標を示す。

BIC の差（2log ≃ 2（BIC））	
0~2	Weak evidence
2~6	Moderate evidence
6~10	Strong evidence
>10	very strong evidence

グループ分けの数は多くても 7 グループ（解釈できる範囲内であること）され、各グループは少なくとも全体の 5%以下にならないことが推奨される。この Guideline に従い、BIC が最も小さい統計量をもつ Group 5 を選択する。医学専門家との協議により Group 5 においてグループ 1 と 2 を区別することは臨床的に解釈しづらいため、本稿では Group 4 を採用した。（図 3）

表 2：BIC 統計量 Group 2,3,4,5,6

Group	BIC	~2BIC	グループ（%）
2	-78436.03		70.6(%) 29.4(%)
3	-71987.50	6448.53	10.7(%) 66.5(%) 22.7(%)
4	-68293.42	3694.08	21.7(%) 9.7(%) 60.9(%) 7.7(%)
5	-66759.23	1534.19	6.5(%) 7.2(%) 59.0(%) 19.4(%) 7.9(%)
6	-67102.28 （収束しない）	-343.05	8.1(%) 7.1 (%) 10.1 (%) 54.8 (%) 14.9 (%) * 4.9(%)

3. 3　　軌道の形状の選び方について

軌道の形状（Liner, Quadric, Cubic）を選ぶにあたり、まずは軌跡の形状をすべて 3 次関数(cubic)に設定したモデルから評価する。e.g.）MODEL LOGIT; NGROUPS 4; ORDER　3 3 3 3;　/*3 次関数*/

パラメータの数を最小限に抑えながら適切なモデルを選ぶ（モデルの倹約性）ために、統計的に有意性が確認できた次項をそれぞれ選択する。(Helgeson, Snyder et al. 2004) ただし、3 次関数と 2 次関数 に統計的な有意性が確認できなかった場合、1 次関数を選択する。 (Louvet, Gaudreau et al. 2009)

選択した 4 グループ分けの結果（SAS code と Output）より、それぞれのグループに設定する多項式を検討する。グループ 1 のパラメーターは 3 次（p-value＝0.052）2 次（p＝0.000）なので、グループ 1 は 2 次を選択した。 同様にグループ 2、3、4 において 3 次（p-value =0.000）を選択する。また、すべての次項を 2 次で選択した場合、BIC は-68510.61 であった。モデル（2,3,3,3）とモデル（2,2,2,2）の BIC の差は、2BIC = -68266.79 - (-68510.61) = 243 であった。

3.4 モデルの適合度

モデルの適合度について、各グループの事後確率の平均値で検討する。症例ごとに算出された事後確率(Posterior probability)より最大の値をとるグループに割り付ける。（Maximum Probability Assignment）。内的整合性は、データからグループ分けされたの平均事後確率に近似するため、そのグループの平均事後確率が＞0.70～0.80 を超える場合、そのグループは同じ変化パターン（軌跡）を持つ個々のグループとする。

例）Group 3: (0.93 + 0.99 + 0.99) / 3 = 0.97 (SAS code : OUT=Post_of_all)

ID	Group 1 Probability	Group 2 Probability	Group 3 Probability	Group 4 Probability	Group
1	0.000000	0.000767	0.932525	0.066708	3
2	0.999998	0.000000	0.000000	0.000002	1
3	0.000000	0.000931	0.997594	0.001475	3
4	0.000000	0.000000	0.058984	0.941015	4
5	0.999991	0.000007	0.000000	0.000002	1
6	0.000000	0.000011	0.999849	0.000141	3

3.5 開始値の設定について

PROC TRAJ は最尤推定の閉形式解（a closed form solution）を持たない非線形モデルを近似する。閉形式解の代替方法として、尤度のパラメータ空間を繰り返し検索して最尤解を求める開始値を設定する。デフォルト（開始値（start value）を設定しない）では、TRAJ procedure で従属変数の範囲または標準偏差に基づいてグループ切片を計算して、その値をデフォルトの開始値として設定する。（グループのサイズ数は均等に設定　例：4 つのグループを仮定した場合、25%）

SAS ログ：

```
Start:
Parameter estimates
-2.523450  0.000000   0.000000  0.000000  -0.174747  0.000000
 0.000000  0.000000   1.647980  0.000000   0.000000  0.000000
-3.000000  0.000000   0.000000  0.000000  25.000000 25.000000
25.000000 25.000000   0.000000  0.000000   0.000000  0.000000
```

モデルの複雑さやデータの特性により GBTM の適合度が悪くなること、またデフォルトの Start value では GBTM が収束しないことがあるので開始値（Start value）の設定を推奨する。その開始値を設定する手順を以下に示す。それぞれのグループの軌道の形状について、次項（0 は切片モデル、1 は線形モデル、2 は二次式、3 は三次式）を特定する。

モデルに組み入れる適切な次項（order）は 1 グループは 2 次関数、2,3,4 グループは 3 次関数とした。SAS ログに設定したモデルのパラメータ推定値が出力されるので、選択した各パラメータの次項（1 次、2 次、3 次）に合わせて開始値、また Intercept を設定する。（以下）

SAS ログ

```
Start:

Parameter estimates

   0.753224    -1.215689     0.058925     1.643889     0.964739    -0.265470
   0.013180     1.268195     1.112693    -0.131777     0.004173     0.570134
  -0.961328     0.222170    -0.011017    21.666019     9.700432    60.947926
   7.685623     0.000000     0.000000     0.000000     0.000000     0.000000
```

SAS code に Start value を設定

```
/*Parameter estimates*/
START    0.753224   -1.215689    0.058925   /*0.001460   */
         1.643889    0.964739   -0.265470    0.013180
         1.268195    1.112693   -0.131777    0.004173
         0.570134   -0.961328    0.222170   -0.011017
         21.666019   9.700432   60.947926    7.685623;
RUN;
```

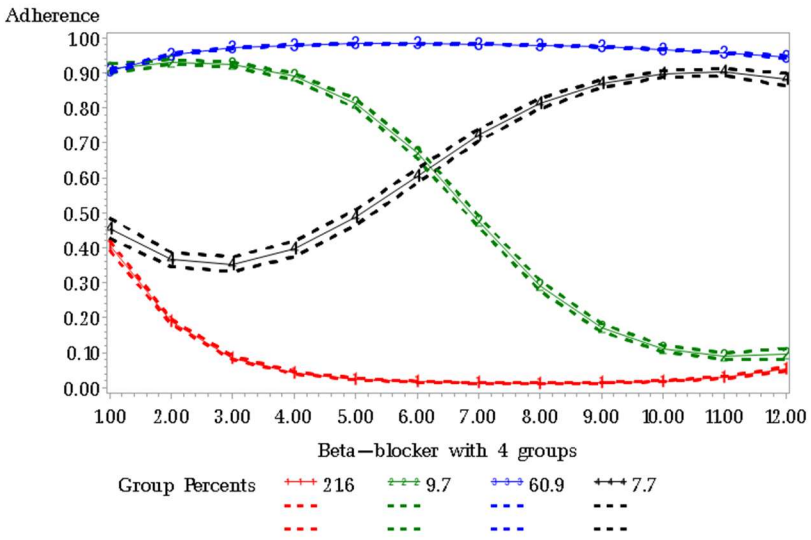
選択した 4 グループ分けモデル解析結果：Group 4　BIC-68266.79, 各グループの事後平均確率：1)0.98, 2)0.95, 3)0.98, 4)0.87 であり全ての Groups は 0.8 以上であった。

LCGM の最近の拡張機能：「CI95M」をプログラムに追加「%TRAJPLOTNEW」により 95%信頼区間を計算可(Jones &Nagin, 2007)

Maximum Likelihood Estimates
Model: Logistic (LOGIT)

Group	Parameter	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
1	Intercept	0.86794	0.04763	18.223	0.0000†
	Linear	-1.33661	0.02711	-49.301	0.0000†
	Quadratic	0.08540	0.00219	39.081	0.0000†
2	Intercept	1.64318	0.14497	11.334	0.0000†
	Linear	0.95979	0.08719	11.008	0.0000†
	Quadratic	-0.26477	0.01501	-17.640	0.0000†
	Cubic	0.01316	0.00075	17.438	0.0000†
3	Intercept	1.26753	0.06877	18.432	0.0000†
	Linear	1.11587	0.05637	19.795	0.0000†
	Quadratic	-0.13239	0.01022	-12.956	0.0000†
	Cubic	0.00420	0.00052	8.155	0.0000†
4	Intercept	0.56972	0.10675	5.337	0.0000†
	Linear	-0.95862	0.07529	-12.732	0.0000†
	Quadratic	0.22189	0.01397	15.880	0.0000†
	Cubic	-0.01100	0.00073	-14.986	0.0000†
Group membership					
1	(%)	21.64858	0.30649	70.635	0.0000†
2	(%)	9.73032	0.23656	41.132	0.0000†
3	(%)	60.92874	0.39695	153.492	0.0000†
4	(%)	7.69237	0.26466	29.066	0.0000†

BIC= -68289.15 (N= 229584) BIC= -68266.79 (N= 19132) AIC= -68196.05 ll= -68178.05

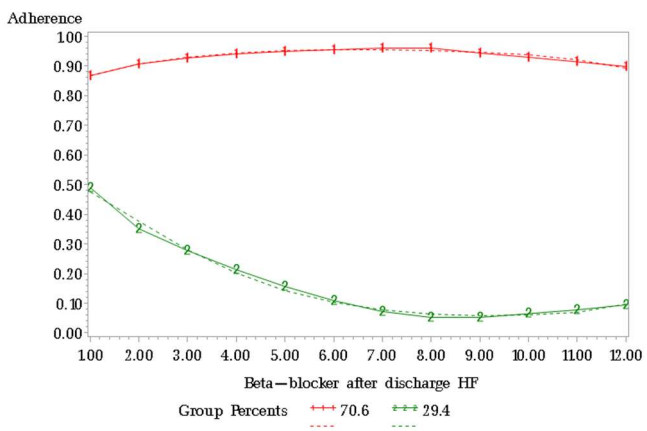



```
%TRAJPLOTNEW(OP,OS,,'',"Adherence",'Beta-blocker with 4 groups')
```

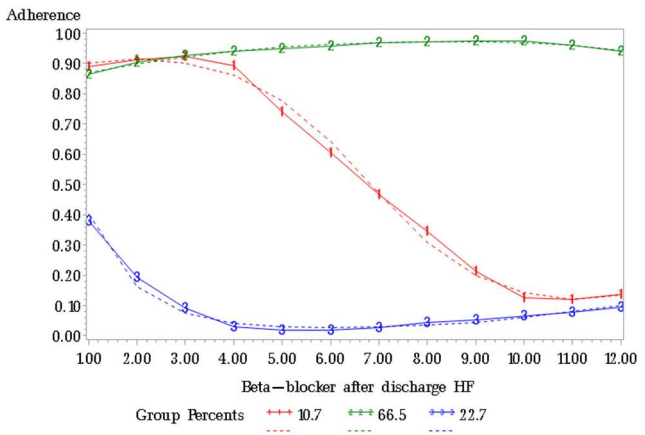
3.6 その他

複数の変数をモデルに組み入れて様々な種類の軌跡を作成できる。この手順は、これまで紹介した手順に従い異なる変数をモデルに組み入れる。ただし、事前にグループ数を定義する必要があり、異なる変数ごとにグループ数が同じであることが前提になる。

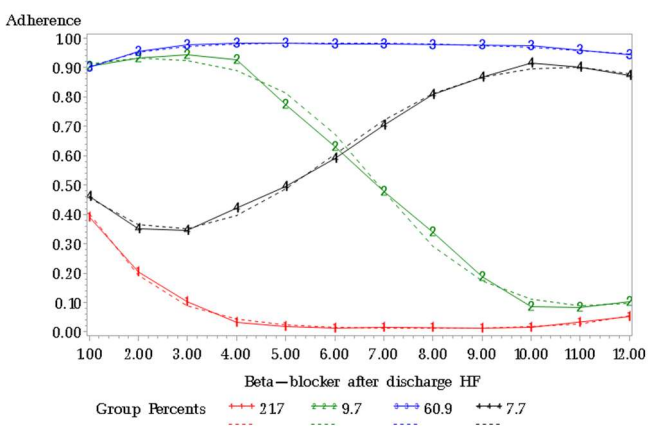
図 3：GBTM plot Group 2, 3, 4, 5 , 6 and BIC



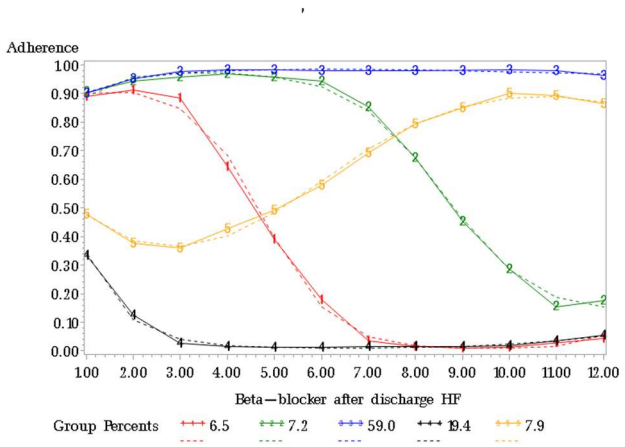
N=2



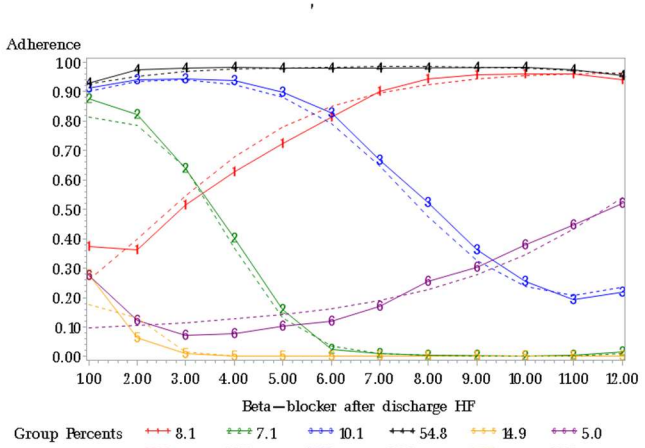
N=3



N=4



N=5



N=6

Group (N=)	BIC
2	-78436.03
3	-71987.50
4	-68293.42
5	-66759.23
6	-67102.28

4. まとめ：

本稿では、PROC TRAJ を用いて、経時的に測定されるデータの解析方法である Group-based trajectory model 方法を紹介した。GBTM は、時間の経過とともに変化する結果を分析するのに役立つ手法で、特にこの手法の利点として、特定のタイプまたは数の軌跡の存在を事前に仮定するのではなく、規定された手順より、実データから仮定した軌跡を得られるかを評価できる。特に、他の解析手法では見落としてしまうような（予測できない）軌跡も、GBTM を用いることで、潜在的に意義のある軌跡を評価することができる。(Nagin、2005)

本稿が SAS で縦断データにおける経時測定データ解析（特に、いくつかの変化のパターンが想定される）を評価する際の参考になると幸いである。

参考文献：

高橋 雄介：時間を含むデータをどう分析するか？2015 http://hdl.handle.net/2115/58521

石渡量太、橋口正行、志賀剛：保険者 DPC データベースを用いた心不全患者における β 遮断薬の適切な漸増投与量と臨床評価：日本臨床薬理学会 2024 https://www.congre.co.jp/jscpt2024/

池野敬、伊藤弘人：精神保健研究 60：49- 54, 服薬アドヒアランス 2014

大橋 洸太郎 (翻訳)：在成長曲線モデル (計量分析 One Point) 共立出版 2023

Helena Troiano Ona Valls Janine Jongbloed Jean-François Giret Chris Edwards (2023) ”GROUP BASED TRAJECTORY MODELLING Examples of analysis.” Comparative Report. July 2023 https://fe.up.pt/complext/images/demo/GBTM_comparative%20report.pdf

Nagin,D.S: Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*(2), 139–157.1999 https://doi.org/10.1037/1082-989X.4.2.139

Hamada, Y., D. Thomas, E. S. Harvey, S. Stevens, M. Fricker, H. Lewthwaite, V. M. McDonald, A. Gillman, M. Hew, V. Kritikos, J. W. Upham and P. G. Gibson (2025). "Distinct trajectories of treatment response to mepolizumab toward remission in patients with severe eosinophilic asthma." Eur Respir J **65**(1).

Helgeson, V. S., P. Snyder and H. Seltman (2004). "Psychological and physical adjustment to breast cancer over 4 years: identifying distinct trajectories of change." Health Psychol **23**(1): 3-15.

Ishiwata, R., A. AlAshqar, M. Miyashita-Ishiwata and M. A. Borahay (2024). "Dispensing patterns of antidepressant and antianxiety medications for psychiatric disorders after benign hysterectomy in reproductive-age women: Results from group-based trajectory modeling." Womens Health (Lond) **20**: 17455057241272218.

Louvet, B., P. Gaudreau, A. Menaut, J. Genty and P. Deneuve (2009). "Revisiting the changing and stable properties of coping utilization using latent class growth analysis: A longitudinal investigation with soccer referees." Psychology of Sport and Exercise **10**(1): 124-135.

Nagin, D. (1999). "Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach." Psychological Methods **4**: 139-157.

Nguena Nguefack, H. L., M. G. Pagé, J. Katz, M. Choinière, A. Vanasse, M. Dorais, O. M. Samb and A. Lacasse (2020). "Trajectory Modelling Techniques Useful to Epidemiological Research: A Comparative Narrative Review of Approaches." Clin Epidemiol **12**: 1205-1222.

Websites

TRAJ SAS: Download Proc Traj and view examples of its application: http://www.andrew.cmu.edu/user/bjones/index.htm

R package for latent class mixed models to implement group-based trajectory models: http://cran.r-project.org/web/packages/lcmm/index.html(link is external and opens in a new window)

Stata: Implementation and example of GBTM: https://ssrc.indiana.edu/doc/wimdocs/2013-03-29_nagin_trajectory_stata-plugin-info.pdf(link is external and opens in a new window)

RMSTREGプロシジャの回帰分析を理解する

○久貝航太郎，魚住龍史
東京科学大学工学院経営工学系

Understanding regression analysis using PROC RMSTREG

Kotaro Kugai and Ryuji Uozumi
Department of Industrial Engineering and Economics, Institute of Science Tokyo

要旨

境界内平均生存時間は，比例ハザード性が成り立たないデータに対する解析においてハザード比の代替指標として近年注目されている．RMSTREG プロシジャには，境界内生存時間の回帰手法として IPCW 回帰と疑似値回帰の 2 つの手法がサポートされている．本稿では，これら 2 つの回帰手法について，恒等リンク・対数リンクをそれぞれ用いた場合のパラメータベクトル推定プロセスを詳細に解説する．

キーワード：生存時間解析　境界内平均生存時間　一般化線形モデル

1　はじめに

本稿では，SAS に実装されている境界内生存時間に関する 2 つの回帰分析を詳細に解説する．

2　境界内平均生存時間

2.1　定義と性質

生存時間を意味する非負の連続確率変数を T とする．生存関数 $S(t)$ は $S(t) = \Pr(T > t)$ で表される．生存関数下の面積は平均生存時間を表し，治療効果の指標として用いられる．

生存関数はノンパラメトリックに推定でき，その推定量は Kaplan-Meier 推定量と呼ばれる．t_i をイベント発生時刻，d_i を t_i の時点でイベントが発生した患者数，n_i を t_i の直前にイベントも打ち切りイベント (リスク) もなかった患者数とすると，カプラン・マイヤー推定量 $\hat{S}_{\mathrm{KM}}(t)$ は次式で計算できる．

$$\begin{aligned}\hat{S}_{\mathrm{KM}}(t) &= \left(1 - \frac{d_1}{n_1}\right) \times \left(1 - \frac{d_2}{n_2}\right) \times \cdots \\ &= \prod_{t_j < t} \left(1 - \frac{d_j}{n_j}\right)\end{aligned}$$

最後の患者を打ち切ると Kaplan-Meier 曲線は閉じず，曲線下面積が無限に発散してしまい，平均生存時間の計算ができなくなる．これを防ぐため，ある境界時間 $t = \tau$ までの曲線化面積を平均生

存時間の近似値として計算する，境界内平均生存時間 (Restrictd Mean Survival Time, 以下 RMST) が提案された．境界内平均生存時間は以下のように定義される．

$$\mu(\tau) = E[\min(T, \tau)]$$

生存時間 T の確率密度関数を $f(t) = \Pr(T = t)$ とすると，

$$\begin{aligned}
\mu(\tau) &= E[\min(T, \tau)] \\
&= \int_0^\infty t \times \Pr(\min(T, \tau) = t) dt \\
&= \int_0^\tau t f(t) dt + \int_\tau^\infty \tau f(t) dt \\
&= [tF(t)]_0^\tau - \int_0^\tau F(t) dt + \tau [F(t)]_\tau^\infty \\
&= \tau F(\tau) - \int_0^\tau (1 - S(t)) dt + \tau (1 - F(\tau)) \\
&= -[t]_0^\tau + \int_0^\tau S(t) dt + \tau \\
&= \int_0^\tau S(t) dt
\end{aligned}$$

と計算でき，生存関数の区間 $[0, \tau]$ の曲線化面積で表せることがわかる．RMST の推定の際には Kaplan-Meier 曲線の区間 $[0, \tau]$ について，長方形の面積を足し合わせるようにして推定する．

$$\begin{aligned}
\hat{\mu}(\tau) &= \int_0^\tau \hat{S}_{\mathrm{KM}}(t) dt \\
&= \sum_{j=1}^{N} \hat{S}_{\mathrm{KM}}(t_{j-1})(t_j - t_{j-1}) + \hat{S}_{\mathrm{KM}}(t_N)(\tau - t_N)
\end{aligned}$$

RMST はハザード比よりも臨床的に解釈しやすく，また群間差を推定するための比例ハザードの仮定を必要としない．そのため RMST は，比例ハザードの仮定が無効なデータに対する解析において，ハザード比の代替統計量となりうるが，事前に臨床的に妥当な境界時間を指定する必要がある (Hasegawa et al., 2020)．

2.2 回帰分析

本節では，RMST に関する回帰手法のうち，SAS に実装されている以下の 2 種類の回帰について，最尤推定量の詳細な計算過程を概説する．

2.2.1 IPCW 回帰

T_i, C_i を被験者 i のイベント発生時間および打ち切り時間とする．このとき，解析側は以下の 3 つの量 $(U, \Delta, \boldsymbol{Z})$ を得る．U_i は被験者 i のイベント発生または打ち切りまでの時間，Δ_i はイベント

発生の有無を表す指示関数，$\boldsymbol{Z}_i$ は q 次元共変量ベクトルである．

$$
\begin{aligned}
U_i &= \min(T_i, C_i) \\
\Delta_i &= I(T_i < C_i) \\
\boldsymbol{Z}_i' &= (Z_{i1}, \cdots, Z_{iq})
\end{aligned}
$$

τ を境界時間とする．

$$
\begin{aligned}
Y_i &= \min(T_i, \tau) \\
\tilde{\Delta}_i &= I(Y_i < C_i) \\
\mu(\boldsymbol{Z}_i) &= E(Y_i \mid \boldsymbol{Z} = \boldsymbol{Z}_i) \\
\boldsymbol{\beta}' &= (\beta_0, \beta_1, \cdots, \beta_q) \\
\boldsymbol{X}_i' &= (1, Z_{i1}, \cdots, Z_{iq})
\end{aligned}
$$

$\eta(\cdot)$ を連続な狭義単調増加リンク関数とする．

恒等リンク　　$\eta(\mu(\boldsymbol{Z}_i)) = \mu(\boldsymbol{Z}_i) = \boldsymbol{\beta}'\boldsymbol{X}_i = \beta_0 + \beta_1 Z_{i1} + \cdots + \beta_q Z_{iq}$

対数リンク　　$\eta(\mu(\boldsymbol{Z}_i)) = log(\mu(\boldsymbol{Z}_i)) = \boldsymbol{\beta}'\boldsymbol{X}_i = \beta_0 + \beta_1 Z_{i1} + \cdots + \beta_q Z_{iq}$

また，これらの定義において，リンク関数の逆関数は以下のようになる．

恒等リンク　　$\eta^{-1}(\mu(\boldsymbol{Z}_i)) = \boldsymbol{\beta}'\boldsymbol{X}_i$

対数リンク　　$\eta^{-1}(\mu(\boldsymbol{Z}_i)) = \exp(\boldsymbol{\beta}'\boldsymbol{X}_i)$

重み付き正規方程式

$$
\begin{aligned}
S_n(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\Delta}_i}{\widehat{G}(Y_i)} \boldsymbol{X}_i \{Y_i - \eta^{-1}(\boldsymbol{\beta}'\boldsymbol{X_i})\} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\Delta}_i}{\widehat{G}(Y_i)} \boldsymbol{X}_i \{Y_i - E(Y_i \mid \boldsymbol{Z} = \boldsymbol{Z}_i)\} \\
&= \boldsymbol{0}
\end{aligned}
$$

重み付き正規方程式を解く．

$w_i = \frac{\tilde{\Delta}_i}{\widehat{G}(Y_i)}$ とおくと，重み付き正規方程式は以下のように表せる．

$$
S_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i \boldsymbol{X}_i \{Y_i - \eta^{-1}(\boldsymbol{\beta}'\boldsymbol{X_i})\} = \boldsymbol{0}
$$

(i) リンク関数が恒等リンクの場合,

$$
\begin{aligned}
S_n(\boldsymbol{\beta}) &= \sum_{i=1}^{n} w_i \begin{pmatrix} 1 \\ Z_{i1} \\ \vdots \\ Z_{iq} \end{pmatrix} \{Y_i - \eta^{-1}(\boldsymbol{\beta}' \boldsymbol{X}_i)\} \\
&= \begin{pmatrix} \sum_i w_i Y_i - \beta_0 \sum_i w_i - \beta_1 \sum_i w_i Z_{i1} - \cdots - \beta_q \sum_i w_i Z_{iq} \\ \sum_i w_i Z_{i1} Y_i - \beta_0 \sum_i w_i Z_{i1} - \beta_1 \sum_i w_i Z_{i1} Z_{i1} - \cdots - \beta_q \sum_i w_i Z_{i1} Z_{iq} \\ \sum_i w_i Z_{iq} Y_i - \beta_0 \sum_i w_i Z_{iq} - \beta_1 \sum_i w_i Z_{iq} Z_{i1} - \cdots - \beta_q \sum_i w_i Z_{iq} Z_{iq} \end{pmatrix} \\
&= \mathbf{0}
\end{aligned}
$$

これを $\boldsymbol{\beta}$ について整理すると

$$
\begin{pmatrix} \sum_i w_i & \sum_i w_i Z_{i1} & \cdots & \sum_i w_i Z_{iq} \\ \sum_i w_i Z_{i1} & \sum_i w_i Z_{i1} Z_{i1} & \cdots & \sum_i w_i Z_{i1} Z_{iq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i w_i Z_{iq} & \sum_i w_i Z_{iq} Z_{i1} & \cdots & \sum_i w_i Z_{iq} Z_{iq} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} = \begin{pmatrix} \sum_i w_i Y_i \\ \sum_i w_i Z_{i1} Y_i \\ \vdots \\ \sum_i w_i Z_{iq} Y_i \end{pmatrix}
$$

逆行列を用いることで $\boldsymbol{\beta}$ が確定する.

$$
\begin{aligned}
\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} &= \begin{pmatrix} \sum_i w_i & \sum_i w_i Z_{i1} & \cdots & \sum_i w_i Z_{iq} \\ \sum_i w_i Z_{i1} & \sum_i w_i Z_{i1} Z_{i1} & \cdots & \sum_i w_i Z_{i1} Z_{iq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i w_i Z_{iq} & \sum_i w_i Z_{iq} Z_{i1} & \cdots & \sum_i w_i Z_{iq} Z_{iq} \end{pmatrix}^{-1} \begin{pmatrix} \sum_i w_i Y_i \\ \sum_i w_i Z_{i1} Y_i \\ \vdots \\ \sum_i w_i Z_{iq} Y_i \end{pmatrix} \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \\ Z_{11} & \cdots & Z_{n1} \\ \vdots & \ddots & \vdots \\ Z_{1q} & \cdots & Z_{nq} \end{pmatrix} \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} \begin{pmatrix} 1 & Z_{11} & \cdots & Z_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Z_{n1} & \cdots & Z_{nq} \end{pmatrix} \right)^{-1} \\
&\quad \begin{pmatrix} 1 & \cdots & 1 \\ Z_{11} & \cdots & Z_{n1} \\ \vdots & \ddots & \vdots \\ Z_{1q} & \cdots & Z_{nq} \end{pmatrix} \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}
\end{aligned}
$$

(ii) リンク関数が対数リンクの場合，
解析的に $\hat{\boldsymbol{\beta}}$ を求められないため，多次元 Newton 法による近似解を求める．

$$
\begin{aligned}
\boldsymbol{S}(\boldsymbol{\beta}) &= \begin{pmatrix} S_0 \\ S_1 \\ \vdots \\ S_q \end{pmatrix} = \begin{pmatrix} \sum_i w_i Y_i - \sum_i w_i \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \\ \sum_i w_i Z_{i1} Y_i - \sum_i w_i Z_{i1} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \\ \vdots \\ \sum_i w_i Z_{iq} Y_i - \sum_i w_i Z_{iq} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \end{pmatrix} \\
&= \begin{pmatrix} 1 & \cdots & 1 \\ Z_{11} & \cdots & Z_{n1} \\ \vdots & \ddots & \vdots \\ Z_{1q} & \cdots & Z_{nq} \end{pmatrix} \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} \left(\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} - \begin{pmatrix} \exp(\boldsymbol{X}_1\boldsymbol{\beta}) \\ \vdots \\ \exp(\boldsymbol{X}_n\boldsymbol{\beta}) \end{pmatrix} \right) \\
\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{\beta}} &= \begin{pmatrix} \frac{\partial S_0}{\partial \beta_0} & \cdots & \frac{\partial S_0}{\partial \beta_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial S_q}{\partial \beta_0} & \cdots & \frac{\partial S_q}{\partial \beta_q} \end{pmatrix} \\
&= \begin{pmatrix} -\sum_i w_i \exp(\boldsymbol{X}_i\boldsymbol{\beta}) & -\sum_i w_i Z_{i1} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) & \cdots & -\sum_i w_i Z_{iq} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \\ -\sum_i w_i Z_{i1} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) & -\sum_i w_i Z_{i1} Z_{i1} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) & \cdots & -\sum_i w_i Z_{i1} Z_{iq} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_i w_i Z_{iq} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) & -\sum_i w_i Z_{iq} Z_{i1} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) & \cdots & -\sum_i w_i Z_{iq} Z_{iq} \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \end{pmatrix} \\
&= -\begin{pmatrix} 1 & \cdots & 1 \\ Z_{11} & \cdots & Z_{n1} \\ \vdots & \ddots & \vdots \\ Z_{1q} & \cdots & Z_{nq} \end{pmatrix} \begin{pmatrix} w_1 \exp(\boldsymbol{X}_1\boldsymbol{\beta}) & & 0 \\ & \ddots & \\ 0 & & w_n \exp(\boldsymbol{X}_n\boldsymbol{\beta}) \end{pmatrix} \begin{pmatrix} 1 & Z_{11} & \cdots & Z_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Z_{n1} & \cdots & Z_{nq} \end{pmatrix} \\
\boldsymbol{J}(\boldsymbol{\beta}^{(k)}) &= \left. \frac{\partial \boldsymbol{S}}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}}
\end{aligned}
$$

k 回目の反復計算による近似解を $\boldsymbol{\beta}^{(k)}$ とし，$k+1$ 回目以降の近似解を以下のように定める．

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - \boldsymbol{J}^{-1}(\boldsymbol{\beta}^{(k)}) \boldsymbol{S}(\boldsymbol{\beta}^{(k)})$$

これを十分に繰り返すことで $\boldsymbol{S}(\boldsymbol{\beta}) = \boldsymbol{0}$ の近似解 $\hat{\boldsymbol{\beta}}$ を得られる．

2.2.2 擬似値回帰

擬似値 (pseudo value) 回帰は，一般化線形モデルを生存時間データに適合させる一般的な手法である (Andersen et al., 2003)．

被験者 i の θ の疑似値は以下のように計算される．

$$\hat{\boldsymbol{\theta}}_i = n\hat{\boldsymbol{\theta}} - (n-1)\hat{\boldsymbol{\theta}}^{-i}$$

ここで，$\hat{\boldsymbol{\theta}}^{-i}$ は被験者 i 以外の標本に基づくジャックナイフ推定量 (Miller 1974) である．

リンク関数 $g(\cdot)$ を用いて，$\boldsymbol{\theta}_i$ を応答変数とする以下のような一般化線形モデルを仮定する．

$$
\begin{aligned}
&\text{恒等リンク} \qquad \eta(\theta_i) = \theta_i = \boldsymbol{\beta}'\boldsymbol{X}_i = \beta_0 + \beta_1 Z_{i1} + \cdots + \beta_q Z_{iq} \\
&\text{対数リンク} \qquad \eta(\theta_i) = log(\theta_i) = \boldsymbol{\beta}'\boldsymbol{X}_i = \beta_0 + \beta_1 Z_{i1} + \cdots + \beta_q Z_{iq}
\end{aligned}
$$

以下の一般化推定方程式を解くことで，回帰パラメータ $\boldsymbol{\beta}$ の推定が可能となる．

$$
\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) \right)^{\top} \boldsymbol{V}_i^{-1} (\hat{\boldsymbol{\theta}}_i - \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta})) = \boldsymbol{0} \tag{1}
$$

V_i は $\hat{\theta}_i$ の作業分散共分散行列であり，$k \times k$ 対角行列 $A_i, W_i, R(\alpha)$ および，拡散パラメータ ϕ を用いて以下のように定義される．

$$
\begin{aligned}
&A = \begin{pmatrix} v(\mu_{i1}) & & 0 \\ & \ddots & \\ 0 & & v(\mu_{ik}) \end{pmatrix}, \qquad W = \begin{pmatrix} w_{i1} & & 0 \\ & \ddots & \\ 0 & & w_{ik} \end{pmatrix}, \qquad R(\alpha) \in \mathbb{R}^{k \times k} \\
&\hat{\phi} = \frac{1}{N-p} \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\frac{\hat{\boldsymbol{\theta}}_i - \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta})}{\sqrt{v(\mu_{ij})/w_{ij}}} \right)^2 \\
&V_i = \hat{\phi} A_i^{\frac{1}{2}} W_i^{-\frac{1}{2}} R(\alpha) W_i^{-\frac{1}{2}} A_i^{\frac{1}{2}}
\end{aligned}
$$

ここで，N はすべての被験者の測定回数の合計値，p は推定パラメータ数である．V_i は非常に煩雑な推定が必要な作業行列だが，RMST の解析においては各被験者に対する測定回数は 1 回であるため，$k = 1$ となる．このとき V_i はスカラーであり $V_i = 1$ としてよく，上記の一般化推定方程式 $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$ は重みのない正規方程式に帰着する．その後の演算は IPCW 回帰の場合と同様で，リンク関数が恒等リンクの場合は最小二乗法，対数リンクの場合は多次元 Newton 法で解が求まる．

疑似値を用いた回帰分析は，共変量と独立した打ち切りの仮定を必要とする (Komukai et al, 2024)．

(i) 恒等リンクの場合

$$
\begin{aligned}
\eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) &= \boldsymbol{X}_i\boldsymbol{\beta} \\
\frac{\partial}{\partial \boldsymbol{\beta}} \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) &= \begin{pmatrix} 1 & Z_{i1} & \cdots & Z_{iq} \end{pmatrix}'
\end{aligned}
$$

これを用いて，推定方程式を整理する．$V_i^{-1} = 1$ とする．

$$
\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \begin{pmatrix} \hat{\boldsymbol{\theta}}_i - \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) \\ Z_{i1}(\hat{\boldsymbol{\theta}}_i - \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta})) \\ \vdots \\ Z_{iq}(\hat{\boldsymbol{\theta}}_i - \eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta})) \end{pmatrix} = \boldsymbol{0} \tag{2}
$$

$\boldsymbol{\beta}$ は，IPCW 回帰と同様に最小二乗法で求まる．

$$\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_q\end{pmatrix} = \left(\begin{pmatrix}1 & \cdots & 1\\ Z_{11} & \cdots & Z_{n1}\\ \vdots & \ddots & \vdots\\ Z_{1q} & \cdots & Z_{nq}\end{pmatrix}\begin{pmatrix}1 & Z_{11} & \cdots & Z_{1q}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & Z_{n1} & \cdots & Z_{nq}\end{pmatrix}\right)^{-1}\begin{pmatrix}1 & \cdots & 1\\ Z_{11} & \cdots & Z_{n1}\\ \vdots & \ddots & \vdots\\ Z_{1q} & \cdots & Z_{nq}\end{pmatrix}\begin{pmatrix}\hat{\theta}_1\\ \vdots\\ \hat{\theta}_n\end{pmatrix}$$

(ii) 対数リンクの場合

$$\begin{aligned}\eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) &= \exp(\boldsymbol{X}_i\boldsymbol{\beta})\\ \frac{\partial}{\partial\boldsymbol{\beta}}\eta^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) &= \Big(\exp(\boldsymbol{X}_i\boldsymbol{\beta}) \quad Z_{i1}\exp(\boldsymbol{X}_i\boldsymbol{\beta}) \quad \cdots \quad Z_{iq}\exp(\boldsymbol{X}_i\boldsymbol{\beta})\Big)'\end{aligned}$$

対数リンクは解析的に解けないため，IPCW 回帰の場合と同様に多次元 Newton 法によって近似解を求める．こちらも，$V_i^{-1}=1$ とする．

$$\begin{aligned}\frac{\partial\boldsymbol{U}}{\partial\boldsymbol{\beta}} &= \begin{pmatrix}\frac{\partial U_0}{\partial\beta_0} & \cdots & \frac{\partial U_0}{\partial\beta_q}\\ \vdots & \ddots & \vdots\\ \frac{\partial U_q}{\partial\beta_0} & \cdots & \frac{\partial U_q}{\partial\beta_q}\end{pmatrix}\\ &= -\begin{pmatrix}1 & \cdots & 1\\ Z_{11} & \cdots & Z_{n1}\\ \vdots & \ddots & \vdots\\ Z_{1q} & \cdots & Z_{nq}\end{pmatrix}\begin{pmatrix}\exp(\boldsymbol{X}_1\boldsymbol{\beta}) & & 0\\ & \ddots & \\ 0 & & \exp(\boldsymbol{X}_n\boldsymbol{\beta})\end{pmatrix}\begin{pmatrix}1 & Z_{11} & \cdots & Z_{1q}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & Z_{n1} & \cdots & Z_{nq}\end{pmatrix}\\ \boldsymbol{J}(\boldsymbol{\beta}^{(k)}) &= \left.\frac{\partial\boldsymbol{U}}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}}\end{aligned}$$

k 回目の反復計算による近似解を $\boldsymbol{\beta}^{(k)}$ とし，$k+1$ 回目以降の近似解を以下のように定める．

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - \boldsymbol{J}^{-1}(\boldsymbol{\beta}^{(k)})\boldsymbol{U}(\boldsymbol{\beta}^{(k)})$$

これを十分に繰り返すことで $\boldsymbol{U}(\boldsymbol{\beta})=\boldsymbol{0}$ の近似解 $\hat{\boldsymbol{\beta}}$ を得られる．

3 おわりに

本稿は，境界内生存時間および RMSTREG に実装されている回帰手法についての詳しい計算過程を解説することを目的に作成された．また，RMSTREG を用いずに IPCW 回帰及び疑似値回帰の両リンク関数の推定を再現した SAS プログラムを付録に示している．本稿がこれから RMST について学習する学生や研究者の一助となれば幸いである．

参考文献

[1] Andersen PK, Klein JP, and Rosthøj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*. 2003;**90**:15-27.

[2] Andersen PK, Hansen MG, and Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis*. 2004;**10**:335-350.

[3] Andersen PK and Perme MP. Pseudo-observations in survival analysis. *Statistical Methods in Medical Research*. 2010;**19**(1):71-99.

[4] Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*. 1972;**34**(2):187-220.

[5] Hasegawa T, Misawa S, Nakagawa S, Tanaka S, Tanase T, Ugai H, Wakana A, Yodo Y, Tsuchiya S, Suganami H; JPMA Task Force Members. Restricted mean survival time as a summary measure of time-to-event outcome. *Pharmaceutical Statistics*. 2020;**19**:436–453.

[6] Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958;**53**(282):457-481.

[7] Komukai S, Yokota I, and Sakamaki K. Survival Analysis with Pseudo Observations: A Review. *Japanese Journal of Biometrics*. 2024;**45**(2):135–154.

[8] Miller RG. The jackknife - a review. *Biometrika*. 1974;**61**(1):1-15.

[9] Nelder JA and Wedderburn RWM. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*. 1972;**135**(3):370-385.

[10] Tian L, Zhao L, and Wei LJ. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*. 2014;**15**(2):222-233.

[11] 中川雄貴. RMST の概要と SAS での実装について. *SAS* ユーザー総会論文集 *2021* 年. 2021:112-121.

[12] 花田圭佑, 小島将裕, 守屋順之. 共変量を調整した RMST の SAS での実装と性能評価. *SAS* ユーザー総会論文集 *2022* 年. 2022:5.

[13] 長谷川貴大，田中慎一，棚瀬貴紀，若菜明，淀康秀，飯森孝行，鵜飼裕之，仲川慎太郎，三澤早織．生存時間解析の評価指標に関する最近の展開 - RMST(restrictcd mean survival time) を理解する -. *SAS* ユーザー総会論文集 *2018* 年. 2018:355

付録：SAS プログラム

プログラム 1：生存時間データを生成するプログラム

```
%let n = 500; /*sample size*/
%let tau = 12; /*specific time point tau*/
```

```
data sas2025;
/*call streaminit(2025);*/
do id = 1 to &n;
        if id <= (&n / 2) then TRT = 1; else TRT = 0;
        Factor = rand("normal",0,1);
        lambda = 0.1341 - 0.0764 * TRT + 0.01 * Factor;
        T = rand("exponential") / lambda; C = rand("exponential") / 0.05;
        OS = min(T,C);
        if OS = T then Event = 1; else Event = 0;
        output;
end;
run;
```

プログラム 2：IPCW 回帰 (両リンク関数共通部分)

```
/*calculate Y and delta_tilde*/
data reg0; set sas2025;
        if Event = 0 & OS < &tau then do;
                        Y = 99999; delta = 0;
                end; else do;
                        Y = min(OS, &tau); delta = 1;
                end;
run;
/*calculate G(Y)*/
ods output ProductLimitEstimates = KME1;
proc lifetst data = sas2025; time OS * Event(1); run;
data KME2; set KME1; if not missing(Survival);
keep OS Survival Failed; run;
proc sort data=reg0 out=sorted; by OS; run;
data KME3; merge KME2 sorted; by OS; run;
data KME4; set KME3; retain S_ret 1;
if OS < &tau then do;
        if Survival = . then do; Survival = S_ret; end;
        else do; S_ret = Survival; end;
end;
else do; Survival = S_ret; end;
if delta = 1 then w = delta / Survival; else w = 0; run;
data KME5;set KME4; if not missing(id); keep OS Y Event TRT Factor w; run;
```

プログラム 3：IPCW 回帰 (恒等リンク)

```
proc iml; /*weighted least square estimation*/
use KME5;
```

```
    read all var {w} into vec_w; /*nx1*/
    read all var {Y} into vec_y; /*nx1*/
    read all var {TRT Factor} into X; /*nxq (q=2)*/
    W = diag(vec_w);  /*nxn*/
    A = J(&n,1,1);   /*nx1*/
    Z = A || X; /*nx(q+1)*/
close KME5;
    beta = inv(t(Z) * W * Z) * t(Z) * W * vec_y;
    print beta; /*calculate beta*/
quit;
run;
```

プログラム 4：IPCW 回帰 (対数リンク)

```
data beta; run;
data beta_init;
input repeated beta_0 beta_1 beta_2;
cards;
0 0 0 0
;
run;
data beta; set beta_init; run;

%macro loopproc(rep = );
%do i = 1 %to &rep;
proc iml; /*weighted least square estimation*/
    use KME5;
        read all var {w} into vec_w;  /* nx1 */
                read all var {Y} into vec_y;   /* nx1 */
                read all var {TRT Factor} into X;  /* nxq (q=2) */
                A = J(&n,1,1);        /* nx1 */
                Z = A || X;        /* nx(q+1)*/
    close KME5;
    use beta;
                read all var {beta_0 beta_1 beta_2} into m_beta; /* &ix(q+1) */
                vec_beta = m_beta[&i,]; /* 1x(q+1) */
    close beta;
        vec_bX = Z * t(vec_beta); /* nx1 */
        vec_expbX = exp(vec_bX); /* nx1 */
        W = diag(vec_w);   /* nxn */
        expbX = diag(vec_expbX); /* nxn */
        beta_new = t(vec_beta) - inv(- t(Z) * W * expbX * Z) * t(Z) * W * (vec_y -
```

```
            vec_expbX); /*calculate beta*/
        create beta_k var{"repeated" "beta_0" "beta_1" "beta_2"};
        tmp = &i || t(beta_new);
        append from tmp;
        close beta_k;
quit;
run;
data beta; set beta beta_k; run;
%end;
%mend;

%loopproc(rep = 20) /*repeat 20 times*/
```

プログラム 5：疑似値回帰 (両リンク共通部分)

```
/*create dataset for calculating theta_i*/
data reg0; set sas2025 nobs = n;
do i = 1 to n;
    if id ^= i then output;
end; run;
proc sort data=reg0 out=reg1; by i; run;

ods output ProductLimitEstimates = KME1;
proc lifetst data = sas2025;
time OS * Event(0);
run;

data KME2; set KME1; if not missing(Survival);
keep OS Survival Failed; run;
data KME3; set KME2;
retain t_ret S_ret ANS_ret 0;
if Failed = 0 then do;
    t_ret = OS; S_ret = Survival; ANS_ret = 0; theta = 0;
end; else do;
    delta_t = (min(OS,12) - t_ret); delta_ans = delta_t * S_ret;
    theta = ANS_ret + delta_ans;
    t_ret = min(OS,12); S_ret = Survival; ANS_ret = theta;
end;
run;
data KME4; set KME3 end = last; if last; keep theta; run;
ods output ProductLimitEstimates = KME5;
proc lifetst data = reg1; time OS * Event(0); by i; run;
```

```
data KME6; set KME5; if not missing(Survival);
keep i OS Survival Failed; run;
data KME7; set KME6;
retain t_ret S_ret ANS_ret check_ret 0;
if Failed = 0 then do;
    t_ret = OS; S_ret = Survival; check_ret = 0; ANS_ret = 0; theta = 0;
end; else do;
    delta_t = (min(OS,12) - t_ret); delta_ans = delta_t * S_ret;
    theta_i = ANS_ret + delta_ans;
    t_ret = min(OS,12); S_ret = Survival; ANS_ret = theta_i;
        if t_ret = 12 then
        do;
                if check_ret ^= 1 then
                do; check = 1; check_ret = 1; end;
        end;
end;
run;
data KME8; set KME7;if check; keep i theta_i; run;
data KME9; if _N_ = 1 then set KME4; set KME8; run;
data JN; set KME9 nobs =n; jne = n * theta - (n-1) * theta_i; run;
data reg2; merge sas2025 JN; keep id OS Event TRT Factor jne; run;
```

プログラム 6：疑似値回帰 (恒等リンク)

```
proc iml; /*least square estimation*/
    use reg2;
                read all var {jne} into vec_theta;   /*nx1*/
                read all var {TRT Factor} into X;  /*nxq(q=2)*/
                A = J(&n,1,1);   /*nx1*/
                Z = A || X;   /*nx(q+1)*/
        close reg2;
        beta = inv(t(Z) * Z) * t(Z) * vec_theta;
        print beta; /*calculate beta*/
quit;
run;
```

プログラム 7：疑似値回帰 (対数リンク)

```
data beta; run;
data beta_init;
input repeated beta_0 beta_1 beta_2;
cards;
0 0 0 0
```

```
;
run;
data beta; set beta_init; run;

%macro loopproc(rep = );
%do i = 1 %to &rep;
proc iml;
    use data_reg;
                read all var {jne} into vec_theta;   /*nx1*/
                read all var {TRT Factor} into X;  /*nxq(q=2)*/
                A = J(n,1,1);   /*nx1*/
                Z = A || X;   /*nx(q;1)*/
        close data_reg;

        use beta;
                read all var {beta_0 beta_1 beta_2} into m_beta; /* &ix(q+1) */
                vec_beta = m_beta[&i,]; /* 1x(q+1) */
        close beta;
        vec_bX = Z * t(vec_beta); /* nx1 */
        vec_expbX = exp(vec_bX); /* nx1 */
        expbX = diag(vec_expbX); /* nxn */
        beta_new = t(vec_beta) - inv(- t(Z) * expbX * Z) * t(Z) * (vec_theta -
            vec_expbX); /*calculate beta*/
        create beta_k var{"repeated" "beta_0" "beta_1" "beta_2"};
        tmp = &i || t(beta_new);
        append from tmp;
        close beta_k;
quit;
run;
data beta; set beta beta_k; run;
%end;
%mend;

/* repeat 20 times */
%loopproc(rep = 20)
```

境界内平均生存時間による生存時間解析の実践

○魚住 龍史
東京科学大学工学院経営工学系

Practical application of survival analysis with restricted mean survival time

Ryuji Uozumi
Department of Industrial Engineering and Economics, Institute of Science Tokyo

要旨

生存時間データの定量的な治療効果の指標として，Cox 回帰から推定されるハザード比がよく用いられている．しかし，近年の医学分野では，Cox 回帰の前提条件である比例ハザード性が成り立たない生存時間データを扱う場面が増えており，ハザード比の代替指標として境界内平均生存時間の利用が注目されている．本稿では，SAS による境界内平均生存時間の分析機能についてチュートリアル形式で解説しながら，境界内平均生存時間による生存時間解析の実践として，実際の臨床試験にどのような位置付けで応用されているか，議論する．

キーワード：生存時間解析　境界内平均生存時間　区間推定　仮説検定　LIFETEST　LSMEANS　RMST　RMSTREG

1　はじめに

医学分野では，被験者のイベント発生までの時間に対して統計解析がよく実施される．このような分析手法の総称は生存時間解析と呼ばれ，SAS/STAT では，LIFETEST プロシジャや PHREG プロシジャが該当する．著者は，SAS/STAT の新たな機能拡張のたびに，生存時間解析に関連した SAS/STAT の活用報告を行ってきた (大橋ら, 2016; 魚住ら, 2017, 2018, 2019; 魚住・浜田, 2018; 魚住, 2022)．がん患者を対象とした臨床試験では，Kaplan-Meier 法による生存時間曲線の作成，ログランク検定による群間比較，Cox 回帰からのハザード比による治療効果の推定がよく用いられ，SAS/STAT では LIFETEST プロシジャと PHREG プロシジャで計算可能な方法である．Cox 回帰はセミパラメトリックな統計モデルであり，ハザード比が時点に依らず一定だという比例ハザード性の仮定のもと，ハザード比が推定される．

しかし，近年のがん免疫療法の臨床試験では，治療効果が観察されるまでに一定の時間を要する特徴から，比例ハザード性が成り立たず，Cox 回帰から推定されるハザード比を要約指標とすると，臨床的な解釈が難しいと指摘されている．比例ハザード性が成り立たない場合，ノンパラメトリック検定であれば，重み付きログランク検定による対応が挙げられる．重み付きログランク検定で用いられる重みとして，Fleming-

Harrington 型の重みが有用といえるものの，重みの事前規定が難しく，治療効果の推定には代替手法が必要といえる．そこで，ハザード比に代わる生存時間の要約指標として，境界内平均生存時間 (restricted mean survival time, RMST) が用いられることがある．境界内平均生存時間に関する議論は，Uno et al. (2014) 以降活発に行われており，SAS ユーザー総会でも 2018 年に企画セッションとして議論されている (長谷川ら, 2018)．同時期に，SAS/STAT 15.1 からは LIFETEST プロシジャで境界内平均生存時間の計算が可能となった (中川, 2021)．さらに，SAS/STAT 15.1 では RMSTREG プロシジャも新たに追加され，境界内平均生存時間による回帰分析が可能となった (花田ら, 2022)．以上より，SAS による境界内平均生存時間に関連した機能は整備されており，書籍でも取りあげられている現状である (大橋ら, 2022)．日本製薬工業協会医薬品評価委員会データサイエンス部会 タスクフォース 4 生存時間解析チーム (2018) による成果物や国際ジャーナルでの出版物 (Hasegawa et al., 2020) でも詳しく解説されている．

このような現状を考慮すると，臨床試験に従事する実務統計家にとっては，Cox 回帰によるハザード比の代替指標として，境界内平均生存時間が実際の臨床試験にどのような位置付けで応用されているか気になるところではないだろうか．例えば，境界内平均生存時間が生存時間データの主解析として用いられている事例 (Guimarães et al., 2020) もあれば，Cox 回帰によるハザード比を治療効果の指標として報告することが計画されていたのにもかかわらず，最終的には当初の計画から変更され，境界内平均生存時間が報告されている事例 (Gupta et al., 2025) もある．

また，境界内平均生存時間による統計解析結果として，もしも当該報告書に必要な項目が出力できない場合，別のソフトウェアを用いることになるか，あるいは新たなコーディングが必要となるだろう．そこで，実際の臨床試験に境界内平均生存時間による統計解析を計画する場合，LIFETEST プロシジャや RMSTREG プロシジャを用いれば，出力結果が不足なく出力されているか気になるのではないだろうか．

本稿では，SAS による境界内平均生存時間の分析機能についてチュートリアル形式で解説し，生存時間データを評価項目としたいくつかの臨床試験の事例をもとに，境界内平均生存時間による生存時間解析がどのように実践されているか議論する．本稿の構成は次の通りである．2 節では，境界内平均生存時間の概要と数理とともに解説する．3 節では，簡便なデータセットを用いて，SAS/STAT のプロシジャによる境界内平均生存時間のプログラム及び出力結果について解説する．4 節では，境界内平均生存時間を用いた臨床試験の事例を導入する．最後に，5 節では，4 節の事例をもとに，SAS 社への要望を示す．本稿のまとめを示す．

2 境界内平均生存時間の概要

2.1 境界内平均生存時間の推定

生存関数 $S(t)$ とすると，境界内平均生存時間は境界時間 τ までの生存関数の曲線下面積であり，

$$\mu(\tau) = \mathrm{RMST}(\tau) = \int_0^\tau S(t)\,dt$$

と定義される．境界内平均生存時間は，Kaplan-Meier 曲線にもとづいて推定されることが多く，境界時間 τ までの範囲内でイベントが発生した時点を $t_1, t_2, \cdots, t_D$，各時点に対応するイベント数を $d_1, d_2, \cdots, d_D$，リスク集合の大きさを $n_1, n_2, \cdots, n_D$ とすると，

$$\hat{\mu}(\tau) = \int_0^{\tau} \hat{S}(t)\,dt = \sum_{j=1}^{D} (t_j - t_{j-1})\hat{S}(t_{j-1}) + (\tau - t_D)\hat{S}(t_D)$$

と推定できる．ただし，$t_0 = 1$ である．

境界内平均生存時間の推定値が得られれば，区間推定や仮説検定も興味があるだろう．そこで，Greenwood の公式を用いれば，$\hat{\mu}(\tau)$ の分散

$$\hat{V}[\hat{\mu}(\tau)] = \sum_{j=1}^{D} \left[\sum_{i=j}^{D} (t_i - t_{i-1})\hat{S}(t_{i-1}) + (\tau - t_D)\hat{S}(t_D) \right]^2 \frac{d_j}{n_j(n_j - d_j)}$$

と推定される．バイアス補正項を含んだ

$$\hat{V}_{BC}[\hat{\mu}(\tau)] = \frac{m}{m-1}\hat{V}[\hat{\mu}(\tau)]$$

として，分散を推定することもある．ただし，m は総イベント数を表す $(m = \sum_j d_j)$.

2.2　境界内平均生存時間による群間比較

群 j における境界内平均生存時間の推定値を $\hat{\mu}_j(\tau)$，分散の推定値を $\hat{V}_j[\hat{\mu}(\tau)]$ とすると，群 1 と群 0 の群間比較を行う場合，帰無仮説：$\mu_1(\tau) = \mu_0(\tau)$ のもと，カイ二乗検定統計量

$$\chi^2 = \frac{\{\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)\}^2}{\hat{V}_1[\hat{\mu}(\tau)] + \hat{V}_0[\hat{\mu}(\tau)]} \tag{1}$$

が自由度 1 のカイ二乗分布にしたがう．

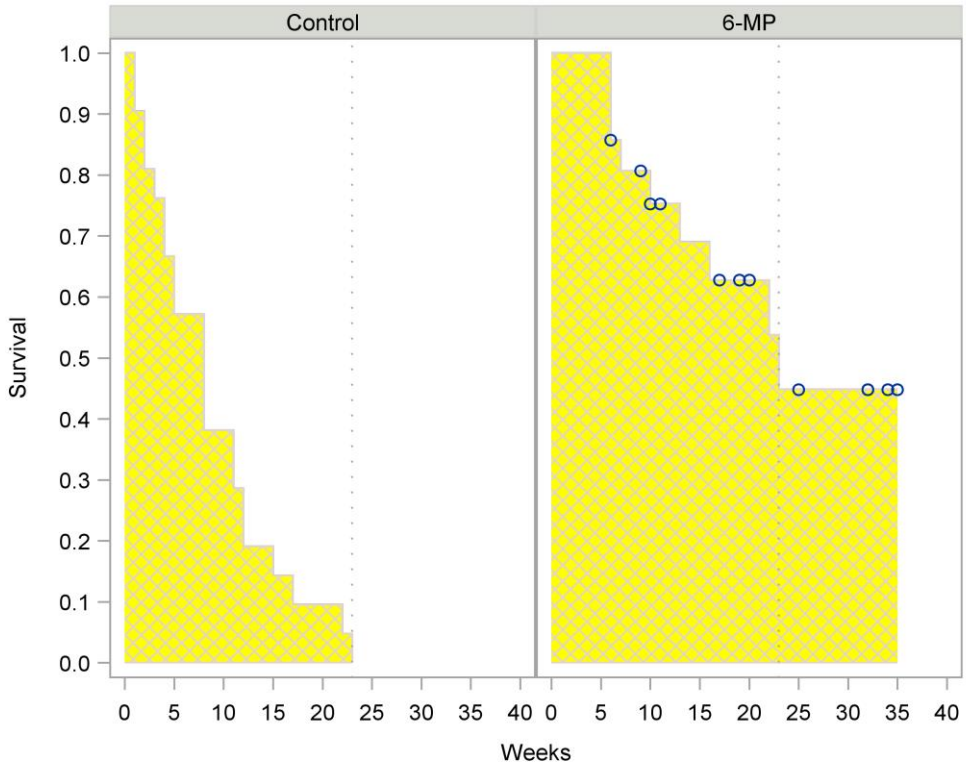


図 1：Gehan の白血病データセット (大橋ら, 2016, 2022)

さらに，境界内平均生存時間による群間差 $\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)$ の 95%信頼区間は

$$\{\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)\} \pm z_{\alpha/2}\sqrt{\hat{V}_1[\hat{\mu}(\tau)] + \hat{V}_0[\hat{\mu}(\tau)]} \tag{2}$$

と構成される．ただし，z_κ は標準正規分布の上側 κ パーセント点である．

3　SAS による境界内平均生存時間

図 1 は，『生存時間解析 第 2 版：SAS による生物統計』や『生存時間解析 応用編：SAS による生物統計』で用いられている Gehan の白血病データセット(大橋ら, 2016, 2022) の Kaplan-Meier 曲線及び境界内平均生存時間に相当する範囲を塗りつぶした状態で示している．最終イベント観測時間は 23 週であり，最終打ち切り観測時間は 35 週である．本節では，Gehan の白血病データセットを用いて，SAS/STAT のプロシジャによる境界内平均生存時間の分析方法を示す．

3.1　LIFETEST プロシジャ

PROC LIFETEST 文の RMST オプションを指定すれば，Kaplan-Meier 法による生存関数の推定結果などの LIFETEST プロシジャのデフォルト出力に加えて，境界内平均生存時間による分析結果が出力される．

SAS プログラム

```
proc lifetest data=Gehan rmst(cl);
  time Time*Status(0);  strata Drug;
run;
```

出力結果

RMST 分析の情報

Tau	23

RMST 推定値

層	Drug	推定値	標準誤差	95%信頼区間	
				下限	上限
1	0	8.667	1.377	5.967	11.366
2	1	17.909	1.553	14.865	20.953

RMST 同等性の検定

要因	カイ 2 乗値	自由度	P 値
Strata	19.822	1	< 0.001

図 2：LIFETEST プロシジャの境界内平均生存時間に関する出力結果

図 2 は，Gehan の白血病データセットに対して，LIFETEST プロシジャによる境界内平均生存時間の推定を行った結果を示している．図 2 の出力結果のうち，RMST 同等性の検定として出力されているカイ 2 乗値は式 (1) から計算される．RMST オプションとして，RMST(CL) と記述することで，境界内平均生存時間

の信頼区間も出力されているが，各群の信頼区間のみであり，群間差の信頼区間が出力されていないことがわかる．

境界時間 τ を変動させたときの境界内平均生存時間の推移を確認したい場合，図 3 のように PROC LIFETEST 文のオプションとして，PLOTS = RMST を記述すれば，図 3 のような出力結果が得られる (LIFETEST プロシジャの出力結果を SGPLOT プロシジャで出力し直している)．PLOTS = RMST(CL) と記述すれば，各群の境界内平均生存時間の信頼区間も追加して出力することが可能である．なお，本稿では SAS/STAT 15.3 を使用しており，SAS/STAT 15.3 では「境界内平均生存時間」という日本語でラベルされる．

SAS プログラム

```
proc lifetest data=Gehan plots=rmst;
  time Time*Status(0);  strata Drug;
run;
```

出力結果

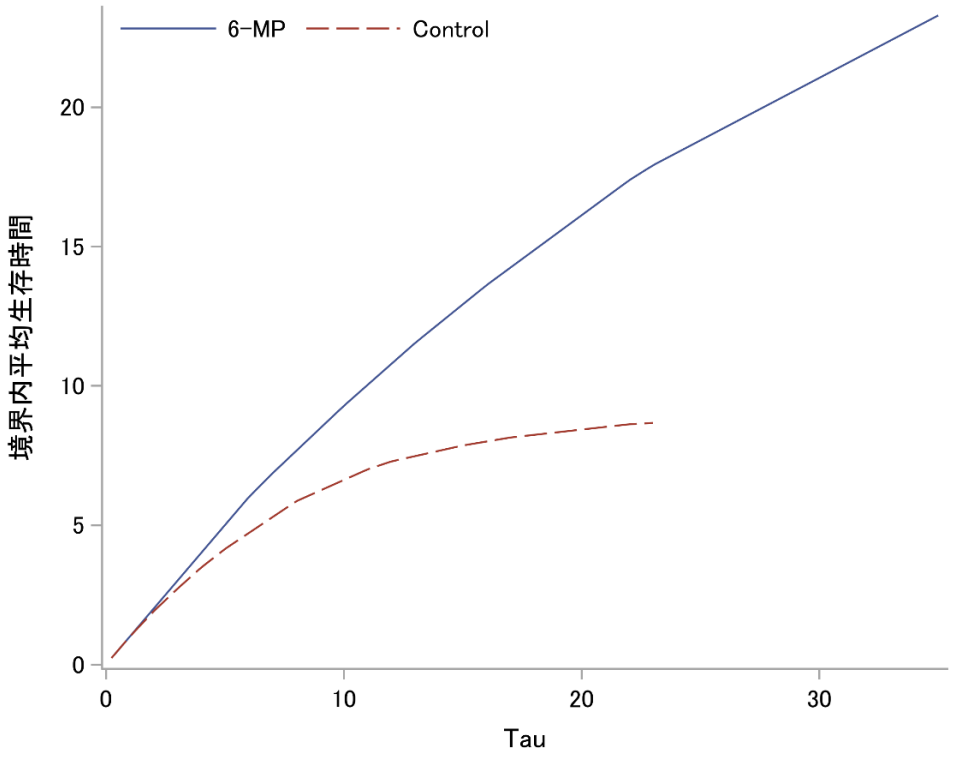


図 3：境界時間に対する境界内平均生存時間の推移図

3.2 RMSTREG プロシジャ

LIFETEST プロシジャでは，各群の境界内平均生存時間の推定，境界内平均生存時間による群間比較として，カイ二乗検定統計量を算出し，p 値を出力した．RMSTREG プロシジャは，境界内平均生存時間に対する回帰分析に相当し，共変量調整を行ったうえでの境界内平均生存時間の推定が可能である (花田ら, 2022; 久貝・魚住, 2025)．PHREG プロシジャを用いた Cox 回帰による共変量調整と同じように，モデルによる推定後のプロセスとして，SAS の多くのプロシジャでは LSMEANS 文や LSMESTIMATE 文を利用すれば (魚住, 2014; 大橋ら, 2016)，モデルによる推定値を用いて，各群の境界内平均生存時間の推定可能である．

SAS プログラム

```
proc rmstreg data=Gehan tau=23;
  class Drug / ref=first;
  model Time*Status(0) = Drug / link=linear method=pv;
  lsmeans Drug / cl diff;
run;
```

出力結果

最小 2 乗平均

Drug	推定値	標準誤差	Z 値	P 値	95%信頼区間	
					下限	上限
1	17.749	1.514	11.72	< 0.001	14.782	20.717
0	8.382	1.376	6.09	< 0.001	5.684	11.080

最小 2 乗平均の差

Drug	_Drug	推定値	標準誤差	Z 値	P 値	95%信頼区間	
						下限	上限
1	0	9.368	2.046	4.58	< 0.001	5.357	13.378

図 4：RMSTREG プロシジャによる出力結果の一部

図 4 は，RMSTREG プロシジャによる SAS プログラム，出力結果の一部として LSMEANS による出力結果部分を示している．PROC RMSTREG 文のオプションとして，境界時間 ＝23 週を指定している．これは LIFETEST プロシジャによる境界時間と合わせるためであり，境界時間を指定しないと，最長の打ち切り発生時間である 35 週が境界時間と指定されてしまう．MODEL 文では恒等リンク関数を指定し，LSMEANS 文で出力される最小二乗平均の差が境界内平均生存時間の群間差に相当するようにしている．もしも境界内平均生存時間の群間比を求めたい場合，対数リンク関数 (LINK = LOG) を指定すればよい．さらに，推定方法として，デフォルトの擬似値回帰法 (pseudo-value regression) (Andersen et al., 2003) を用いている．擬似値回帰法以外にも，IPCW (inverse probability censoring weighting) 法 (Tian et al., 2014) が指定可能である．

図 2 では，群 1 の境界内平均生存時間 17.9 週，群 0 の境界内平均生存時間 8.7 週より，境界内平均生存時間の群間差は 9.2 週となる．図 4 の擬似値回帰にもとづく境界内平均生存時間の回帰分析からは，群 1 の境界内平均生存時間 17.7 週，群 0 の境界内平均生存時間 8.4 週より，境界内平均生存時間の群間差は 9.4 週となり，LIFETEST プロシジャによる結果からは若干数値が異なることがわかる．また，LIFETEST プロシジャでは，境界内平均生存時間の群間差の推定値と信頼区間は出力されなかった．一方，RMSTREG プロシジャでは，LSMEANS 文や LSMESTIMATE 文を活用すれば，図 4 のような境界内平均生存時間の群間差の推定値 9.4 週に加えて，95%信頼区間も出力可能である．

4　境界内平均生存時間を用いた臨床試験事例

4.1　補足的解析：比例ハザード性が仮定できない場合の対応

HER2 陽性の転移性乳癌患者を対象に，ヨーロッパ，北米，南米，アジア，オーストラリアの 28 ヶ国で実施された第 III 相臨床試験 (Saura et al., 2020) は通称 NALA 試験では，無増悪生存時間と全生存期間が主要評価項目として設定され，ラパチニブ群に対するネラチニブ群の非劣性を検証することを目的とした群間比較が行われている．プロトコルでは，ログランク検定による群間比較と Cox 回帰によるハザード比の推定が主解析である統計解析計画が記述されており，比例ハザード性の仮定ができない場合は事前に設定した補足的解析として，境界内平均生存時間による群間比較が記述されていた．結果報告において，Kaplan-Meier 曲線の形状から比例ハザード性の崩れが顕著であり，仮説検定でも確認できる数値であったため，境界内平均生存時間による群間差の推定値，信頼区間，p 値が報告されていた．また，各群の境界内平均生存時間の推定値とその信頼区間も示されていた．

4.2　補足的解析から主解析への変更：比例ハザード性が仮定できない場合の対応

小児 B 細胞急性リンパ芽球性白血病患者を対象に，オーストラリア，カナダ，ニュージーランド，米国で実施された第 III 相臨床試験 (Gupta et al., 2025) は通称 AALL1731 試験と呼ばれ，無病生存期間が主要評価項目として設定され，化学療法単独群に対する化学療法ブリナツモマブ併用療法群の優越性を検証することを目的とした群間比較が行われている．プロトコルでは，ログランク検定による群間比較が主解析である統計解析計画が記述されており，境界内平均生存時間による群間比較が記述されていなかった．しかし，論文による結果報告では，境界内平均生存時間による群間差 72 日 (95%信頼区間：36 to 108 日) が報告され，ログランク検定による p 値 ＜0.001 と並列で示されていた．付録を確認すると，Schoenfeld 残差などを用いた比例ハザード性の評価が行われ，比例ハザード性が仮定できないと判断された．そこで，統計解析結果としてハザード比は一切用いられず，境界内平均生存時間による群間比較結果が報告された．境界内平均生存時間による群間比較では，各群の境界内平均生存時間が等しいという帰無仮説のもと，正規近似による検定統計量が構成され，SAS/STAT 15.1 の RMSTREG プロシジャを用いたことが記述されている．RMSTREG プロシジャでは，IPCW 法と擬似値回帰法が指定可能であるが，本事例では擬似値回帰法を用いたと明示されている．境界時間は最終イベント観測時間が用いられていた．さらに，フォレストプロットによるサブグループ解析の結果でも，境界内平均生存時間による群間差を横軸としたグラフ報告がされている．

4.3　補足的解析：イベント数不足の対応

HER2 陽性の高齢の早期乳癌患者を対象に，本邦で実施されたランダム化臨床試験 (Sawaki et al., 2020) は通称 N-SAS BC07/RESPECT 研究と呼ばれ，無病生存時間が主要評価項目として設定され，化学療法トラスツズマブ併用療法群に対するトラスツズマブ単独群の非劣性を検証することを目的とした群間比較が行われている．デザイン段階にハザード比 1.69 の治療効果が見積もられ，臨床医を対象とした事前アンケートから非劣性マージン 1.69 と規定されていた．Cox 回帰が主解析として統計計画されたが，盲検下モニタリングにより，イベント数が想定よりもはるかに少なく，ハザード比にもとづく非劣性の評価では検出力が不十分であると懸念された．そこで，境界内平均生存時間による補足的解析が追加された．結果報告では，ハザード比による結果と並列して，境界内平均生存時間による群間差とその信頼区間，p 値が示されていた．無病生存時間について，ハザード比 1.36 (95%信頼区間：0.72 to 2.58，p 値 ＝0.51) となり，非劣性は検証されな

かった．ハザード比の結果を補足する形で，境界内平均生存時間の群間差の推定値 −0.39 年 (95%信頼区間：−1.71 to 0.93，p 値 = 0.56) と報告された．

4.4 主解析

生体僧帽弁置換術後の心房細動患者を対象に，ブラジルで実施されたランダム化臨床試験 (Guimarães et al., 2020) は通称 RIVER 試験と呼ばれ，複合エンドポイントによるイベント発生までの時間が主要評価項目として設定され，ワルファリン群に対するリバーロキサバン群の非劣性を検証することを目的とした群間比較が行われている．デザイン段階から境界内平均生存時間による群間比較が主解析として計画されており，各群のハザードからハザード比 0.79 の治療効果が見積もられ，関連研究から非劣性マージン 8 日 (365 日の約 2%相当) と規定されていた．結果報告では，境界内平均生存時間による群間差 7.4 日 (95%信頼区間：−1.4 to 16.3 日) が報告され，非劣性の p 値 < 0.001，優越性の p 値 = 0.10 が並列で示されていた．主要評価項目に対するサブグループ解析では，境界内平均生存時間による群間差を横軸としたフォレストプロットが報告されている．なお，副次評価項目に対しては Cox 回帰によるハザード比が報告されていた．

表 1：臨床試験事例の要約

事例	試験名の通称	計画時の主解析	境界内平均生存時間の事前規定	結果の表示
4.1	NALA	Cox 回帰	比例ハザード性が仮定できない場合	Cox 回帰と RMST*
4.2	AALL1731	Cox 回帰	なし	RMST* のみ
4.3	N-SAS BC07 / RESPECT	Cox 回帰	イベント数不足の場合	Cox 回帰と RMST*
4.4	RIVER	RMST*	RMST で解析計画	主要評価項目：RMST* 副次評価項目：Cox 回帰と RMST

*RMST: 境界内平均生存時間

5 臨床試験事例にもとづく SAS/STAT プロシジャに対する要望

表 1 では，4 節で取りあげた臨床試験事例をまとめている．4.1 節では，境界内平均生存時間による群間比較の p 値のみが出力されており，LIFETEST プロシジャで出力可能な p 値である．すべての事例では，境界内平均生存時間による生存時間解析の結果報告として，境界内平均生存時間による群間差の信頼区間が報告されている．3.2 節で述べたように，RMSTREG プロシジャを用いれば，境界内平均生存時間による群間差の信頼区間は構成可能である．LSMEANS 文や LSMESTIMATE 文を活用すれば，各群の境界内平均生存時間の推定結果も得ることができ，4.2 節の事例で報告されているような数値も得ることが可能である．

しかし，LIFETEST プロシジャでは，各群における境界内平均生存時間の信頼区間のみ構成可能で，群間差の推定値及び信頼区間はいずれも出力できない現状である．現在の LIFETEST プロシジャでは，各群の境界内平均生存時間の推定値や標準誤差までは出力される．そこで，例えば図 5 のようなデータハンドリングにより，境界内平均生存時間による群間差の推定値及び信頼区間は容易に出力できる．データハンドリングで

容易に対応であるとはいえ，カイ二乗検定統計量が計算できるのであれば，信頼区間についても同じようにLIFETEST プロシジャの出力に加えてもらいたい内容である．

```
ods listing close;
ods output rmst=rmst00;
proc lifetest data=Gehan rmst(cl); time Time*Status(0); strata Drug;
run;
ods listing;
*============================================================*;
* LIFETESTプロシジャの出力データセットRMST00からデータハンドリング ;
*============================================================*;
proc transpose data=rmst00 out=se00 prefix=se; id trt; var StdErr;
proc transpose data=rmst00 out=est00 prefix=est;
  id trt; var Estimate;
data cl00; merge est00 se00; keep est1 est0 se1 se0;
data cl00; set cl00;
  alpha=0.05; za=probit(1-alpha/2); se=sqrt(se0**2+se1**2);
  est=est1-est0; lower=est-za*se; upper=est+za*se;
run;
```

図 5 : LIFETEST プロシジャを用いた境界内平均生存時間の群間差の信頼区間構成プログラム

境界内平均生存時間による群間差の信頼区間に関しては，上記のプログラムの場合，Wald 型の信頼区間といえる．しかし，FREQ プロシジャによる 2 項割合の差の信頼区間のように (魚住ら, 2015)，境界内平均生存時間の信頼区間についても多くの方法が考案されている (Tang, 2021)．Tang (2021) によると，式 (2) は (プールしない) Greenwood の分散にもとづく信頼区間と示され，別の種類の信頼区間として，プールした Greenwood の分散にもとづく信頼区間も示されている．プールした Greenwood の分散にもとづく信頼区間では，各群の生存時間分布は等しいが，打ち切り分布は異なる可能性があると仮定している．

また，RMSTREG プロシジャにおけるリンク関数として，対数リンク関数も指定可能なことから，境界内平均生存時間の群間比の推定値と信頼区間について，RMSTREG プロシジャから出力可能である．したがって，LIFETEST プロシジャの出力においても，境界内平均生存時間の群間比の推定値と信頼区間を加えてもらうことで，生存時間解析に従事するユーザーにとって．より利便性が増すだろう．

6 まとめ

本稿では，境界内平均生存時間の分析機能について，SAS/STAT の LIFETEST プロシジャと RMSTREG プロシジャで利用可能な方法をチュートリアル形式で解説し，生存時間データを評価項目とした 4 つの臨床試験事例をもとに，境界内平均生存時間による生存時間解析がどのように実践されているか議論した．結果，境界内平均生存時間による解析が主解析として用いられている事例も散見された．しかし，多くの場合，Cox 回帰による生存時間解析が主解析として計画され，比例ハザード性が仮定できない状況への対応として，境界内平均生存時間による解析が事前規定されている事例が多いといえる．さらに，臨床試験事例の結果報告と SAS/STAT のプロシジャによる出力の現状を比較すると，LIFETEST プロシジャによる出力結果と

して，境界内平均生存時間による群間差の推定値と信頼区間が出力されるべきである．また，RMSTREG プロシジャで指定可能なリンク関数を考慮すると，対数リンク関数を指定できることからも，境界内平均生存時間による群間比の推定値と信頼区間についても出力されるべきである．SAS Institute Japan の皆様には，本稿での議論内容を是非 North Carolina の米国本社に共有いただきたい．臨床試験にデザイン及び統計解析に従事する多くの実務家にとって，本稿の内容がお役に立てれば幸いである．

参考文献

[1] Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis*. **10**:335–350, 2004.
[2] Guimarães HP, Lopes RD, de Barros E Silva PGM, et al. Rivaroxaban in Patients with Atrial Fibrillation and a Bioprosthetic Mitral Valve. *New England Journal of Medicine*. **383**:2117–2126, 2020.
[3] Gupta S, Rau RE, Kairalla JA, et al. Blinatumomab in Standard-Risk B-Cell Acute Lymphoblastic Leukemia in Children. *New England Journal of Medicine*. **392**:875–891, 2025.
[4] Hasegawa T, Misawa S, Nakagawa S, et al. Restricted mean survival time as a summary measure of time-to-event outcome. *Pharmaceutical Statistics*. **19**:436–453, 2020.
[5] Saura C, Oliveira M, Feng YH, et al. Neratinib Plus Capecitabine Versus Lapatinib Plus Capecitabine in HER2-Positive Metastatic Breast Cancer Previously Treated With ≥ 2 HER2-Directed Regimens: Phase III NALA Trial. *Journal of Clinical Oncology*. **38**:3138–3149, 2020.
[6] Sawaki M, Taira N, Uemura Y, et al. Randomized Controlled Trial of Trastuzumab With or Without Chemotherapy for HER2-Positive Early Breast Cancer in Older Patients. *Journal of Clinical Oncology*. **38**:3743–3752, 2020.
[7] Tang Y. Some new confidence intervals for Kaplan-Meier based estimators from one and two sample survival data. *Statistics in Medicine*. **40**:4961–4976, 2021.
[8] Tian L, Zhao L, Wei LJ. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*. **15**:222–233, 2014.
[9] Uno H, Claggett B, Tian L, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. *Journal of Clinical Oncology*. **32**:2380–2385, 2014.
[10] 魚住龍史. LS-Means 再考 –GLM と PLM によるモデル推定後のプロセス–. SAS ユーザー総会 論文集 2014, 449–463.
[11] 魚住龍史・飯塚政人・浜田知久馬. 非劣性試験における割合の差の信頼区間と例数設計. SAS ユーザー総会 論文集 2015, 422–431.
[12] 魚住龍史・矢田真城・浜田知久馬. PHREG プロシジャにおける共変量調整解析に関連したオプション機能. SAS ユーザー総会 論文集 2017, 237–248.
[13] 魚住龍史・浜田知久馬. SAS による競合リスクを伴う生存時間解析の理解. SAS ユーザー総会 論文集 2018, 75–90.
[14] 魚住龍史・矢田真城・篠崎智大・川口淳・浜田知久馬. SAS による時間依存性 ROC 曲線と C 統計量. SAS ユーザー総会 論文集 2018, 6–24.
[15] 魚住龍史・中川雄貴・矢田真城. SAS による区間打ち切りデータを伴う生存関数の推定. SAS ユーザー総会 論文集 2019, 3–17.
[16] 魚住龍史. Cox 回帰とその拡張. SAS ユーザー総会 論文集 2022.
[17] 大橋靖雄・浜田知久馬・魚住龍史. 生存時間解析 応用編 –SAS による生物統計. 東京大学出版会, 2016.
[18] 大橋靖雄・浜田知久馬・魚住龍史. 生存時間解析 第 2 版 –SAS による生物統計. 東京大学出版会, 2022.
[19] 久貝航太郎・魚住龍史. RMSTREG プロシジャの回帰分析を理解する. SAS ユーザー総会 論文集 2025.
[20] 中川雄貴. RMST の概要と SAS での実装について. SAS ユーザー総会 論文集 2021.
[21] 日本製薬工業協会 医薬品評価委員会データサイエンス部会 タスクフォース 4 生存時間解析チーム. 生存時間型応答の評価指標（第 2 版）–RMST (restricted mean survival time)を理解する–. 日本製薬工業協会, 2019. Available at https://www.jpma.or.jp/information/evaluation/results/allotment/rmst.html.
[22] 花田圭佑・小島将裕・守屋順之. 共変量を調整した RMST の SAS での実装と性能評価. SAS ユーザー総会 論文集 2022.
[23] 長谷川貴大・田中慎一・棚瀬貴紀ら. 生存時間解析の評価指標に関する最近の展開 –RMST(restrictcd mean survival time) を理解する–. SAS ユーザー総会 論文集 2018, 355.

連絡先

E-mail : *uozumi.r.ab@m.titech.ac.jp*

IMLライセンスを利用せずに
Proc FCMPで基本的な行列計算を行う方法について

森岡　裕

（イーピーエス株式会社）

Performing Basic Matrix Calculations with Proc FCMP without Using an IML License"

Yutaka Morioka

(EPS Corporation)

SAS® IMLは行列計算に強力な機能を備えた言語であるが，IMLライセンスがない環境でも，SASのProc FCMPを利用することで基本的な行列演算を実行することが可能である．本発表では，Proc FCMPに用意されている行列操作ルーチンを活用した行列計算の実装例を示した．

まず，配列を行列とみなし，積・逆行列・転置・単位行列生成・行列式計算・コレスキー分解といった基本的な操作を行う方法を紹介した．さらに，応用として連立方程式の解法，重回帰分析における偏回帰係数の導出，対称行列の対角化の例を示し，統計解析における利用可能性を考察した．

Proc FCMPはユーザー定義関数やサブルーチンを作成できる特性を持ち, IMLにはない拡張性を提供する一方，可変引数や複雑な戻り値を伴う処理の実装は難易度が高い．また，固有値分解や特異値分解などの高度な行列操作はProc Princomp等の既存プロシジャを併用する方が効率的であることも示した．

総じて，Proc FCMPはIMLに比べて機能は限定的ながら，IMLが利用できない環境における行列計算手段として有効であり，シミュレーションや臨床試験データ解析の場面でも十分活用可能であると考えられる．さらに，FCMPを用いたユーザー定義関数によって不足する機能を補う余地があり，行列処理の選択肢を広げる意義があると結論づけた．

SAS Packages Frameworkでのパッケージとして開発された「love_matrix」についても紹介する．このパッケージは本発表で紹介したFCMPの機能をマクロとしてパッケージ化したものである．

SASシステムでリアルワールドデータを解析する際の注意点

○岡本　史子、平野　勝也

（株式会社インテリム　データサイエンスユニット　メディカルデータアナリシスグループ
リアルワールドデータ解析チーム）

承認申請, 再審査申請へのリアルワールドデータの利活用を推進する取り組みが種々行われる中、ICH-M14「医薬品の安全性評価においてリアルワールドデータを活用する薬剤疫学調査の計画、デザイン、解析に関する一般原則」[1]が2024年5月にリリースされた。Step 4の合意・採択に向けて準備が進められており、今後、ますますリアルワールドデータを利活用する機会、SASシステムで解析する機会が増加すると考えられる。レセプトデータ、レジストリデータ、電子カルテデータ等のリアルワールドデータは、臨床試験データに比して大規模になる傾向があるため、大規模データを始めて扱う人に向けて、SASシステムで大規模データを解析する際の注意点として、①プロジェクト（共有）フォルダを利用する場合の注意点、②SASシステム起動時設定の注意点、③SASプログラミング時の注意点（ifよりwhere、数値型より文字型、compressオプションの利用）を紹介する。

本報告では、2012年のSASユーザー総会の企画セッション「PMDAの医薬品副作用データベースを用いた副作用リスクの評価」[2]で、シグナル検出、副作用リスクポジショニング、潜在的リスク分析に利用されたJADERデータ（更新：2025年4月）を、利用可能な大規模データ（医薬品情報, 副作用テーブルは、それぞれ1,540,317, 338,359レコード[更新：2012年5月]から4,745,288, 1,587,986レコード[更新：2025年4月]に増加している）として利用しながら注意点を紹介する。

最後に、ICH-M14[1]では「国内又は海外のデータベースから得られた自発報告を用いた医薬品安全性監視における調査」は適用範囲外とされていること、PMDAでの安全対策措置においてデータマイニング手法によりJADERデータから検出されるシグナルが利用されていること[3]、JADERデータ利用時の留意事項を含めた「研究発表でJADERデータを利用する際のチェックリスト」が提案されていること[4]を参考にされたい。

参考文献

[1] ICH M14：医薬品の安全性評価においてリアルワールドデータを活用する薬剤疫学調査の計画、デザイン、解析に関する一般原則（Step 2）

[2] 高橋行雄・半田　淳．2012．PMDAの「医薬品副作用データベース」を用いた副作用リスクの評価－自社医薬品の副作用リスクのポジショニングへの活用－（1）企画の趣旨およびデータベースの概要．SASユーザー総会2012．

[3] 松井和浩・飯村康夫. 2014. データマイニング手法の安全対策業務への導入と活用. ファルマシア. 50(6).

[4] 酒井隆全．2020．JADERを用いた研究発表の際に留意すべきチェックリストの提案．薬剤疫学．25(2).

SAS Packages - the way to share

○Jablonski, Bartosz[1]

([1]Takeda Pharmaceutical / yabwon / Warsaw University of Technology)

SAS Packages offer a solution for sharing complex SAS code by bundling macros, formats, functions, and data into a single, structured file. This reduces issues when moving code from local environments to servers or cloud platforms like SAS Workbench. Using the SAS Packages Framework and SAS Packages Archive, developers can easily build, organize, and distribute code, minimizing deployment errors and simplifying collaboration. The presentation shows how to make code-sharing a piece of cake.

SASPAC - SAS packageの作り方と事例紹介 -

○仲家 諒[1]、鳴澤 輝子[1]

([1]武田薬品工業株式会社)

近年、製薬業界におけるSASプログラミングの現場ではオープンソース文化の拡大という大きな後押しもあり、再利用性や保守性を高めるためにコードやリソースを統一的な形式でパッケージ化し、チームやコミュニティで共有するニーズが高まっている。Bartosz Jablonski氏の開発したSAS Package Framework(SPF)は、そのニーズに応える形でSASにおけるパッケージ管理・配布・利用を効率的に行える仕組みを提供している。

本発表では、まずSAS packagesとは何か、その目的や利点を簡潔に説明した上で、パッケージ開発の基盤となるフォルダ・ファイル構成を解説し、SPFが提供する%generatePackage()を用いた開発プロセスを紹介する。

さらに具体的な事例として、

- パッケージ作成を効率化するパッケージ「SASPACer」
- リアルワールドデータの効率的なハンドリングを支援するパッケージ「RWDExpress」

の2つを取り上げ、それぞれの設計方針や特徴などを共有する。

本発表を通じて、参加者がSAS Packagesに関する活動やパッケージ開発の魅力を理解し、自らのSAS packageを作成・公開する一歩を踏み出すきっかけとなることを目指す。

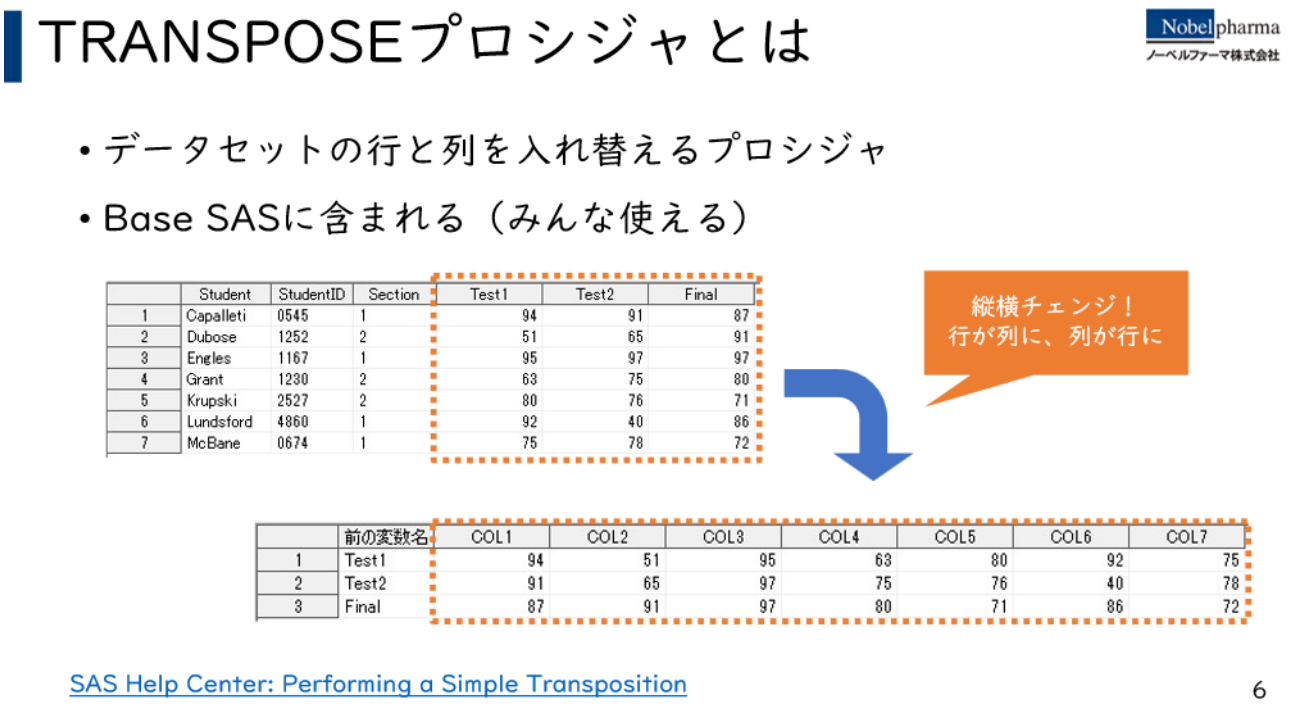
小さく始めるPROC TRANSPOSE

～ 転置無用？ ～

○森田　祐介、太田　裕二、内藤　寿輝、浜田　泉

（ノーベルファーマ株式会社　データサイエンス部）

SASにおけるデータ加工の基本プロシジャのひとつであるTRANSPOSEプロシジャは、データセットの行と列を入れ替えることで、解析の前処理やレポート作成において柔軟なデータ加工を可能にする有用なプロシジャである。一方で、TRANSPOSEプロシジャはその構文がやや独特であり、初学者にとっては「思ったような出力にならない」「エラーの原因がわかりにくい」といったつまずきがあり、習得に苦手意識を持たれがちである。本発表では、こうした課題を踏まえ、SAS初学者を主な対象としてTRANSPOSEプロシジャの使い方をStep by Step形式で解説する。具体例を交えながら、基本構文とともに、出力の形を指定する上で重要となるステートメント（BY / ID / VARなど）の役割や組み合わせのコツについて紹介する。



	Student	StudentID	Section	Test1	Test2	Final
1	Capalleti	0545	1	94	91	87
2	Dubose	1252	2	51	65	91
3	Engles	1167	1	95	97	97
4	Grant	1230	2	63	75	80
5	Krupski	2527	2	80	76	71
6	Lundsford	4860	1	92	40	86
7	McBane	0674	1	75	78	72

	前の変数名	COL1	COL2	COL3	COL4	COL5	COL6	COL7
1	Test1	94	51	95	63	80	92	75
2	Test2	91	65	97	75	76	40	78
3	Final	87	91	97	80	71	86	72

クリックで見せる！Geminiを活用したハイライトグラフ作成法

○橋本菜乃[1]、○小野洋平[1]

([1]聖路加国際大学公衆衛生学研究科)

本文　新型コロナウイルス感染症の拡大により、外出自粛や通院控えが推奨されたが、慢性疾患患者の通院行動や処方内容の変化に関する国内研究は限られている。特に精神科外来処方について、都道府県ごとの時系列変化を俯瞰することは地域医療の課題を把握する上で有用である。

目的　厚生労働省が公開するNDBオープンデータを用いて、精神科外来処方薬の経年変化を都道府県別に可視化し、パンデミックの影響と地域差を明らかにすることを目的とした。

方法　NDBオープンデータから、対象期間における精神科外来処方データを抽出し、都道府県別の処方量をSASにて集計・加工した。次に、Geminiの支援下でインタラクティブなグラフを描画し、任意の都道府県をクリックでハイライト表示できるマクロを実装した。院外処方の内服向精神薬の処方量を第1~10回(平成26年度~令和5年度)まで都道府県別に経時的な変化を可視化した。COVID-19による影響の大きさについては、2018~2019平均に対する2020~2021平均の変化率が最も小さかった「その他の循環器薬」をベースラインとして変化を比較した。また、前年と比較した変化率についてもグラフを作成し、その変化の仕方をタイプ別にクラスター分類を行った。また都道府県ごとの自殺者数、精神障害者保健福祉手帳交付台帳登載数などとの関連を検討し、それぞれの都道府県特有の精神保健領域おける課題について探索的な考察を行った。

結果　対前年比の変化について、コロナ禍の影響が現れたと考えられる2020~2022年に処方量が顕著に増えた都道府県はなかった。外来で処方された精神科の内服薬の人口一人あたりの処方量でみると全国で最も多かったのは秋田県であった。秋田県の自殺率は2015年は全国1位(25.7)であったが2023年には5位(22.6)に3.1ポイント減少していた。精神科の内服処方薬の過去10年間における変化トレンドが全国と最も似ていたのが神奈川県であった。2022年の精神福祉障害者福祉手帳交付台帳登録数の全人口に対する割合では、秋田県は0.88%で全国平均の1.1%と比較して低く、これは全国の都道府県で8番目に少ない交付割合であった。

考察　秋田県は外来における人口一人あたりの向精神薬(内服)の処方量が10年間全国最多であったが、自殺率や精神保健福祉手帳の登録者の割合の高さとは一致せず、精神科医療資源の量的充足よりも質や多様性に課題がある可能性が示唆された。本研究は地域特有の医療提供体制やパンデミック対応の違いを把握する手がかりを提供する。

結語　今回の探索的な検討によりこれまで顕在化できていなかった地域課題についての新たな示唆を得ることができた。この結果をもとに、今後QGISを用いて秋田県を含めたいくつかの都道府県における精神科医療資源へのアクセシビリティの比較検討を行い、秋田県の現状が医療資源の量的不足によるものなのかさらに検討を行う予定である。本発表では、SASでのインタラクティブなグラフ表示手法を提案した。これまでSASのOutput Delivery System (ODS) のgraphicsで多項目グラフを扱う際には、出力されるグラフに組み込まれているhtmlにより色や線種による識別に限界があり注目したい領域の抽出が困難だった。本手法により、ユーザーの関心に応じた動的なグラフ表示が可能となり、視覚的訴求力と分析の柔軟性が飛躍的に高まった。これにより、今後大量の時系列グラフを対象とする可視化において、データビジュアライズが視覚的な理解のしやすさだけでなくユーザー主導の探索的分析を支援する有効な手段となりうることが示された。この手法がSASユーザーにとって、データのさらなる活用の一助となることが期待される。

SAS Viyaを中核とする統合解析環境のさらなる進化

○坂井 聡、○松下 和希、北西 由武

（塩野義製薬株式会社 DX推進本部 データサイエンス部）

塩野義製薬データサイエンス部は、RWD (Real World Data) の解析をはじめ、新規事業の創出や社内DX支援など、多岐にわたる業務を担っている。2023年の下期に構築が完了した”SAS Viyaを中核とする統合解析環境”は、「業務効率化/スピードアップ」という側面でそれらの業務に寄与しており、特にRWD活用業務においては、エビデンス創出の加速によるSHIONOGI製品の価値向上、つまり“育薬”への貢献を実現した。

人々の健康に寄り添うヘルスケア産業での活動において、多様なアウトプット創出のためには解析成果物の質の担保と高い倫理観が必要である。そのため、統合解析環境の次のステージでは、「質の向上」と「解析業務多様化への対応」をキーとして、さらなる進化に挑戦している。具体的には、より解析成果物の質が求められる臨床試験/臨床研究や、機械学習などのモダンな技術の活用が必要となる多岐な業務への適用を見据えて、”CSV (Computerized System Validation) の実施”や、”解析環境のグローバル化”、”Viya APIの活用検討”に取り組んでいる。

本発表では、統合解析環境のさらなる利活用を探求し、「“育薬”にとどまらない“創薬”への挑戦」と「多様な活用シーンに柔軟に対応可能なグローバルな解析環境の構想」について紹介する。

PharmaSUG 2025 U.S. レポート

―海外学会で得た最新の知見と気づき―

○家田　維哉

（日本たばこ産業株式会社）

PharmaSUG 2025 U.S.に参加し，初めての海外出張として国際学会の雰囲気や現地での交流を経験する貴重な機会を得た。本発表では，出張の概要や滞在スケジュールを紹介しながら，学会を通じて得た知見や印象に残った出来事について報告する。

PharmaSUG とは，製薬業界における統計解析やプログラミングに関する実務者が集う国際的なカンファレンスであり，業務効率化や新技術の活用に関する講演が数多く行われている。今回は，SAS や R を用いたプログラミング自動化，生成 AI の実務応用，オープンソースツールの導入事例など，現場で即活用できる技術的トピックが豊富に取り上げられていた。

学会で得た最新の知見として，SAS プログラムのヘッダー更新を自動化するマクロに関する発表と，SAS ライクな R パッケージである SASSY に関する発表についての概要を紹介し，それぞれについての技術的背景と実務への応用可能性についての所感を述べる。

講演だけでなく，参加者同士の活発な議論を通じて，多様な視点や課題意識に触れることができ，大きな刺激を受けた。現地で発表を行った日本人参加者の声も交えながら，文化的・言語的な違いを含むコミュニケーション面での気づき，海外学会参加・発表の意義，今後の業務への活用やキャリアについて考察する。

海外学会への参加は，最新の技術や知見を直接収集し，業務に活かす貴重な機会である。専門外の分野にも触れることで視野が広がり，今後の業務やキャリアを考えるきっかけとなる。また，国内外の参加者との交流を通じて人脈が広がり，継続的な情報交換や連携の可能性も生まれる。さらに，海外学会での発表は，自身の技術を国際的に発信し，多様な視点からのフィードバックを得るチャンスである。発表を通じて議論が深まり，技術の応用可能性が広がり，さらなる発展につながる可能性がある。

本発表を通じて，参加者の方々が今後，海外学会への参加や発表に挑戦する一歩を踏み出すきっかけとなれば幸いである。

AI-SAS for RWE：STEP1実装とSTEP2開発の展望

○米田 卓司[1]，藤田 智紀[1]，北西 由武[1]

([1]塩野義製薬株式会社)

本文

リアルワールドエビデンス（RWE）創出には、標準化・効率化・透明性・容易性の課題が存在する．塩野義製薬では, 臨床試験解析で培ったAI-SAS（人工知能解析プログラマシステム）のノウハウを活用したAI-SAS for RWEの開発によりこれらの課題解決に取り組んでいる．STEP1では，簡易な解析条件を記載したコンセプトシートから，各種ドキュメントを準自動生成する．これらのドキュメントを基に，SASプログラムを準自動生成され，解析用データセット及び図表が出力される．これら一連の解析フローはGit上に自動反映され，いつ誰がどのような結果を出力したかが自動記録される．このSTEP1のフローにより従来数ヶ月を要していた解析作業が，RWD解析の熟練者でなくても数時間で完了した．現在進行中のSTEP2では，生成AI技術を活用することで，コンセプトシートやプロトコルの自動生成など，さらなる自動化と効率化を目指している．本発表では構築が完了したSTEP1の詳細と開発中のSTEP2の展望について紹介する．

SASによるRWD研究のデータ分析の落とし穴

～SAS/ACCESS Interface to Amazon Redshiftの機能を通して～

○土屋 裕章[1]、高見 光広[2]、北村 嵩英[2]

([1]Johnson & Johnson Innovative Medicine、[2]イーピーエス株式会社)

近年の臨床研究におけるRWD (Real World Data)の利活用は増加傾向である。RWD利活用の課題として、データを適切に扱える人材不足が挙げられる。これまで治験等の統計解析業務に従事してきたSASプログラマーはTB級のデータやRDB (Relational database)の扱いに慣れておらずRWD研究推進の阻害要因の1つである。さらに、統計解析手法や研究デザインはその重要性が広く認識されており、学ぶ機会も多い。一方で、データハンドリングに関する学習機会は限られている。

SAS/accessはRDBに格納されているデータに対してODBC接続機能を用いてSASのインターフェースからデータを扱うことが可能である。従来のSASプログラマーがRWDを扱う際の第一選択の方法になるが、現場のSASプログラマーはその機能や利用方法に対する知識が不十分である。

本発表では発表者がこれまで経験したSAS/ACCESS Interface to Amazon Redshiftを用いたRWD研究での落とし穴やその解決策について解説する。SASプログラムの書き方による実行速度の違いやエラー解釈/回避方法を含んだ内容を発表する。

観察研究を外部対照とする際の症例数と検出力の計算

～それでもPower procedureを使いますか？～

○馬場 崇充[1,2]，堀江 悠生[1]，藤田 智紀[2]，坂口 弘樹[1]

([1]塩野義製薬株式会社 解析センター，[2]塩野義製薬株式会社 データサイエンス部)

観察研究を臨床試験の外部対照として間接比較を行う場合，臨床試験と観察研究間で被験者が無作為に割り付けられているわけではないため，共変量の分布が偏り，単純な比較ができない．そこで，傾向スコアの分布を両試験間でそろえたうえで比較を行うため，傾向スコアを用いた逆確率重み付け法がしばしば用いられる．そのような場合でも，研究計画時には想定する結果に基づく必要症例数を計算し，十分な被験者データの収集が可能であるかの検討が重要である．Power procedureは無作為化比較試験の必要症例数計算で多用されているが，データが群間で独立であることを仮定しており，このような間接比較でその仮定は成立しない．そこで，傾向スコアを用いた逆確率重み付け法による間接比較に適した症例数計算法が提案されている(Liu et al. 2025)．本発表の目的は，当該提案手法の内容を概説するとともに，SASで実装し，シミュレーションにより，Power procedureとは大きく異なる結果が算出される場合があることを示すことである．

提案手法では，連続値応答と2値応答を扱い，その平均値 (2値応答では割合に相当する) の差により群間比較を行う設定を考えている．逆確率重み付け法による群間差の推定量は漸近正規性を有することから，その分布収束先の正規分布を用いて症例数計算が可能となる．効果サイズを想定値で与えるとすると，漸近分散をいかに推定するかに焦点が当たる．提案手法では，共変量が単変量の正規分布に従うことを仮定し，逆確率重み付け法を用いて推定した群間差の漸近分散を明示的に求めている．その漸近分散の算出のためには，無作為化比較試験で症例数設計を行う際に必要なパラメータに加えて，傾向スコアの群間での重複度，結果変数と共変量の相関係数が必要となる．

当日は，Liu et al. 2025の理論を概説するとともに，症例数計算で必要なパラメータをどのように決定するかについて，当該論文を引用しながら議論したい．また，必要症例数を算出するプログラムをSASで実装し，Case 1:臨床試験と外部対照のデータを前向きに集める場合，Case 2:単群の臨床試験を終えており (例数固定)，外部対照のデータを後ろ向き/前向きに集める場合の2ケースを考え，シミュレーションによりどのような場合にPower procedureとは大きく異なる結果が算出される場合があるかを明らかにする．結果的に，逆確率重み付け法を考慮した症例数計算が必要であることを示す．

引用文献

1. Bo Liu & Fan Li (2025). Sample size and power calculations for causal inference of observational studies. *arXiv*, 2501.11181.

外部データの取り込み・加工に関する拡張機能

○井上 貴博

（ノバルティス ファーマ株式会社）

私が従事している製薬業界においては、臨床試験をはじめとする解析業務において、SAS データセットとして受領したデータを用いて SAS で解析を行うケースが多く存在する。一方で、DB 研究に代表されるように、二次利用を目的としたデータを取り扱う機会も増加傾向にある。また、すべての作業を SAS のみで完結できるわけではなく、SQL でデータを取得し、SAS で加工し、さらに BI ツールで可視化するなど、複数のアプリケーションを連携させる必要性も高まっている。

本発表では、外部データの取り込み・加工に関して注意すべき点を「データの選択」「大規模データの取り込み」「コンバージョンせずにデータを取り込む」の 3 つのセクションに分け、それぞれに対応する SAS の機能を紹介する。

「データの選択」では、外部ファイルのコンテンツ情報を取得するために使用する External file functions（fopen、finfo 等）の利用方法について説明する。「大規模データの取り込み」では、長大文字列（32KB 超）を取得するために使用する VARCHAR Data Type の特徴について述べる。「コンバージョンせずにデータを取り込む」では、VALIDVARNAME=ANY オプションを使用してデータを取り込む方法と、その際の注意点について解説する。

MMRMにおけるCohen’s dの考察

○高橋 祐策、吉田 祐樹、北西 由武

（塩野義製薬株式会社　DX推進本部　データサイエンス部）

群間差の検討においてP値に依存しすぎることの危険性が指摘されており、治療効果の解釈にはそれだけでなく効果量を含めた総合的な観点での評価が推奨されている。群間を平均値で比較する際は、Cohen（1988）で定義されたCohen's d（群間差を標準偏差で割った指標）が広く用いられている。

近年、臨床研究では反復測定データの解析にMMRM（Mixed Models for Repeated Measures）が多く用いられており、この状況でのCohen’s dの算出方法についての議論がある。Schnitzerら（2024）は、変形性関節症に対する3つの臨床試験データを用いて、Cohen’s dの算出に用いる標準偏差の定義（ベースライン時のSD、評価時点のSD、全時点のSDの中央値）によって値が変動することを示した。このような違いは、治療効果の解釈に影響を与えるだけでなく効果量を基にしたサンプルサイズ設計にも影響を及ぼし、過小なサンプルサイズによる検出力不足の可能性が生じる。Schnitzerらはこうした問題を踏まえ、効果量を報告する際には平均値やSDなどの構成要素を明示することを推奨している。加えて、将来的にはSDの定義方法を標準化することで、異なる研究間での治療効果の大きさの比較をより正確に行えるようになると述べている。

そこで、本研究ではMMRMにおけるCohen’s dの統一的な定義の確立を目指し、その前段階として既存手法の特徴と性質を把握することを本発表での目的とする。本発表では、Schnitzerらの方法に加えて他の算出方法も含めた複数の手法を対象に数値シミュレーションを行い、算出値の違いや治療効果の解釈への影響を比較・考察する。この結果を当日の発表にて報告する。

参考文献

1. Cohen J. (1988) Statistical power analysis for the behavioral sciences. Second edition. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
2. Schnitzer, T.J. *et al.* (2024) Effect size varies based on calculation method and may affect interpretation of treatment effect: an illustration using randomised clinical trials in osteoarthritis. *Adv Rheumatol* **64**, 31.

ログの確認と対処法

○後藤　駿輔[1]、井上　貴博[1]

([1]ノバルティス ファーマ株式会)

本文

SASはプログラム実行時に、エラー・警告・NOTEを含むログを生成する。ログの確認は、プログラムが意図通りに動作したかを判断する上で不可欠である。

本発表は、2017年にノバルティス ファーマ株式会社の井上貴博氏が報告した論文「初/中級者が陥りやすいプログラミングエラー」から、「ログの確認と対処法」を抜粋した発表である。該当論文は「ログの確認と対処法」「浮動小数点の丸め誤差」「Program Data Vector (PDV)」を含む三部構成であるが、「ログの確認と対処法」に関する発表はこれまで未実施であり、本講演が初の公開となる。なお、当該発表はSAS OnDemand for Academics環境で作成されており、他環境では結果が異なる可能性がある。

本発表では、SASログの構造と、確認時に注意すべきポイントを体系的に整理する。さらに、エラーメッセージの種類を分類し、各種エラーの特徴と対処法を具体例とともに解説する。加えて、エラー発生の予防や、エラー発生箇所の特定に有効なステートメントやオプションについても紹介する。

ADS specを基にしたSASコード生成の検討

○藤田 智紀[1]、任 点[1]、米田 卓司[1]、吉田 祐樹[1]、北西 由武[1]

([1]塩野義製薬株式会社)

本文

疫学研究においては，解析を実施する前段階として，解析に用いるデータセットの変数定義を記載した「解析用データセット仕様書（ADS spec）」を作成し，それに基づいてSASプログラムをコーディングすることで，各解析用変数および解析用データセットを構築するのが一般的である．ADS specに基づく変数作成のSASコーディング作業には，データソースのテーブル構造や各変数の特性・関係性を理解する必要があり，加えてSAS言語への習熟も求められるため，経験者と未経験者の間で作業時間に大きな差が生じうる．

この差を縮める手段として，近年急速に発展・普及している生成AI，特にLarge Language Model（LLM）によるコード生成の活用が期待される．ChatGPT（OpenAI社），Claude（Anthropic社），Gemini（Google社）など，汎用的なLLMベースの生成AIサービスは多数展開されており，日常業務にも広く浸透しつつある．自然言語で実装内容を伝えることで，目的に沿ったコードを提示してくれるLLMのコード生成能力は，プログラミングを行う者にとって魅力的なツールであり，業務効率化への貢献が期待される．

しかしながら，SASコードの生成に関する事例や知見はまだ十分に蓄積されておらず，公開されているLLM基盤モデルがどの程度の品質でSASコードを生成できるかは不透明である．特に，ADS specの記載内容から解析用データセットを作成するためのSASコードをLLMで自動生成できるのであれば，プログラミング作業時間の大幅な削減が見込まれる．

本検討では，レセプトデータベース研究におけるADS specの一例を対象に，テーブル構造の理解を補助するためにデータベース仕様書も併せて提供し，解析用データセット作成用のSASコードがLLMによって生成可能かを検証した．その結果を踏まえ，理想的なSASコードを得るために必要な情報の種類とその与え方，そして本工程におけるLLM活用の可能性と限界について考察し，得られた知見を紹介する．

有害事象データの可視化とその有用性

笹原 大敬

（イーピーエス株式会社）

臨床試験における安全性解析は、新規治療法が被験者に及ぼす危害のリスクを評価する上で必要不可欠である。有害事象（Adverse Events）は治療の継続や、被験者へリスクを与えることから安全性解析において重要な収集対象となるデータである。従来の有害事象における安全性解析は一覧表、集計表を用いて評価を行うが、有害事象の発現の全体像や発現パターン、経時的な変化などを視覚的かつ直感的に把握することは難しい。

本発表では、これまでに報告されている可視化手法の紹介および実例の提示を通して、有害事象の発現状況の把握の一助となることを目的としている。

紹介する可視化手法は、「ヒートマップ」、「タイムライン」、「ラインプロット」の3種類である。いずれもSGPLOTプロシジャを用い、グラフの作成を行うことができる。ヒートマップではFREQプロシジャにより有害事象ごとの発現頻度を集計し、その結果を視覚的に表現した。タイムラインでは治療開始・終了日および有害事象発現日・終了日を基に、TEAE（Treatment-Emergent Adverse Event：治療下有害事象） の経時的発現状況を表現した。ラインプロットでは有害事象の発現率と持続時間を表現するため PBRF（Probability of Being in Response Function）を群別・時点別に算出し、群間の経時的な差異を可視化した。

これらの可視化手法を用いることで、従来の集計表では把握が難しかった有害事象の発現傾向や経時的な発現状況を視覚的に捉えることが可能となり、安全性解析における補助的なアプローチとしての有用性が過去の報告と同様に確認された。

[参考文献]

- Potuganti, P., & Nalluri, S. (2021). Graphical Representation of Clinical Data using Heat Map Graphs. PharmaSUG 2021, Paper DV-078. Eliassen Group, Biometrics and Data Solutions.
- Matange, S. (2012, March 8). Adverse event timeline – revisited. SAS Blogs: Graphically Speaking. https://blogs.sas.com/content/graphicallyspeaking/2012/03/08/adverse-event-timeline-revisited/.
- Harris, K. (2017). Adverse Event Data over Time. PharmaSUG 2017, Paper SP06. SAS Specialists Limited, Hertfordshire, United Kingdom.

SASにおけるPROC SQL ～DATAステップとの違いとサブクエリ活用まで～

○武藤　隆成

（エイツーヘルスケア株式会社）

本文

2016年の特に古隅弘樹氏の論文でSQLプロシジャの利活用について、まとめられているが、SAS独自の処理方式であるDATAステップとの違いには触れられていない。本講演では、SAS初学者を対象に、PROC SQLとDATAステップの基本構文、処理の考え方、および実行順序の違いについて図を交えて丁寧に紹介する。さらに、SQLの中でも少し発展的なテーマである「サブクエリ」の基本的な種類（スカラー、IN句、EXISTS）と使いどころを、Russ Laveryらが発表した「An Animated Guide: Explaining SQL to Beginners -A SQL process flowchart, and rules, makes SQL easy」の内容を参考に実例付きで紹介する。両者を比較する中で、「どちらを使うべきか」ではなく「どちらが便利か」を判断できるような視点を養うことを目的としている。SASを使ったデータ加工や集計を効率化するための第一歩として、実務にも役立つ構成を目指している。

ダミーデータを用いた疫学研究の手法の確認

○矢田 徹

（株式会社セブントゥワン）

既存治療に追加する治療の効果・安全性等を非介入観察研究で調査する場合、追加治療集団と既存治療集団の比較に際し、追跡期間の起点日を適切に設定しないとイモータルバイアスを生じる恐れがある。すなわち、既存治療開始日と追加治療開始日をそれぞれの集団の起点日とした場合、追加治療集団では、既存治療開始から追加治療開始までの期間に追加治療を行わないと判断させるイベント（例えば、死亡イベント）は発生しえない。それは追加治療集団に分類される患者が、既存治療集団に分類される患者に比べて、既存治療に対する安全性が高い、あるいは既存治療で効果を示しにくいという傾向があることを示唆しており、これがバイアスとなる。

Prevalent New User Designはイモータルバイアスを回避するために提案された手法である。本発表では、追加治療集団の既存治療開始から追加治療開始までの期間に相当する期間を既存治療集団でも考慮して、イモータルバイアスを調整した結果と調整しなかった結果をダミーデータを用いて作成し、比較する。

ダミーデータは、対象疾患の発症から時系列で重症度や診断・治療状況が変換するモデルで作成し、追加治療が対象疾患の重症度の改善や効果が示し、同時に安全性上のリスクも増加させるという想定でシミュレーションを行って作成する。

このダミーデータに対して、Prevalent New User Designを適用するにあたり、PROC PSMATCHを用いてSASで解析を行い、その際の留意点等についても検討する。

SASにおけるMCP利用の検討

○中松 建

（個人）

近年生成AIが注目を集める中で、2024年末頃から主要テック企業が相次いでAIエージェント関連のモデルや規格、ビジョンを発表したことにより、2025年はAIエージェント元年とも呼ばれている。その中でも利用が進んでいるものとして、MCP(Model Context Protocol)が挙げられる。本発表ではMCPに関してサーバー側に加え、クライアント側でのSASの利用について、検討を行ったものについて報告を行う。

サーバー側については、MCPの公式やその他からもpythonのライブラリが提供されている。これをsaspyと組み合わせることで、VSCodeのチャットウィンドウなどMCPクライアントから、チャット形式でのやり取りによりSASの実行や結果の利用が可能なことを確認した。またクライアント側については、MCPの仕様にあわせてproc httpを利用することで、SASから通信の形式にhttpを使用しているMCPサーバーのツールを実行可能なことを確認した。

上記の実際の例の紹介とあわせて、サーバーの例およびクライアント・ホスト用のSASマクロについてgithubで公開を予定している。

EPS-R BaseによるR活用を支えるサービスの紹介

○魚田 孝志[1]、森岡 裕[1]

([1] イーピーエス株式会社)

これまで臨床試験等におけるデータマネジメントや統計解析業務は、SASをはじめとするプロプライエタリソフトウェアを中心に行われてきた。一方、海外でのオープンソースソフトウェア活用事例の拡大を背景に、最近ではRを業務に利用する流れが加速している。FDAでは既にRによる統計解析結果を用いた承認事例が存在し、PMDAも受け入れに前向きな姿勢を示している。近い将来、日本国内においてもRの利用が急速に広がることが予想される。そのため、Rを業務に取り入れるための環境整備やガバナンス構築が、多くの企業にとって課題となりつつある。

当社では、EPS-R BaseとしてR導入を支援するサービスを始める予定である。本サービスは、医薬品や医療機器の統計解析業務等において注目されているRを安全かつ効率的に業務へ取り込むための支援を目的としている。特に、規制要件に対応した環境を整えるためには、Computerized System Validation（CSV）の実施が不可欠であり、我々はその活動を包括的にサポートする。

当社には、これまで多数のソフトウェアに対してCSVを行ってきた実績豊富な専門チームが在籍しており、さらにRの利用経験を有するメンバーが加わることで、柔軟な支援が可能となる。具体的なサポート内容として、文書の作成支援、利用目的に応じた適切なパッケージ選定、依存関係の複雑さを考慮したパッケージ管理やリスクマネジメントへの対応などを挙げている。また、パッケージごとの性質や利用範囲に応じてテストの粒度を設計し、過剰な負担を避けつつも規制要件を満たすバランスの取れたアプローチを提案する。

R導入においては「どの環境を用意すべきか」「どのパッケージを採用すべきか」「どのレベルの検証を行うべきか」といった課題が必ず発生する。自由度が高い反面、ガイドラインに沿った運用を実現するには高度な知識と経験が不可欠であり、導入企業にとって大きな負担となる。当社は、そのような不安や課題に対して、導入から運用までの一貫したサポートを行う。本セミナーでは、そのEPS-R Baseで予定しているサービス内容を紹介する。

GTL（Graph Template Language）未経験者のためのグラフ作成の基礎とSGPLOTとの比較を通したGTLの利点に関する紹介

○安 正鎬[1]、藤川 桂[1]、宮辻 淳[1]

（[1]株式会社インテリム データサイエンスユニット）

GTL（Graph Template Language）でグラフを作成する際には、Templateプロシジャを用いてプログラマーがグラフの細部要素を一つずつ制御することが可能である。これにより、GTLではニーズに合わせたグラフを作成することができ、SGPLOTで表現が困難なグラフも作成可能である。また、GTLでは、ODS Graphics Designerからコードを参照できる大きな利点がある。しかし、GTLで作成するほとんどのタイプのグラフは、SGPLOTに比べ記述しなければいけないコードの量が多く、GTL未経験のSASユーザーにとっては比較的難易度の高い言語であると言われている。本発表では、GTL未経験のSASユーザーを対象に、GTLの基礎知識と共に、GTLとSGPLOTを使用したいくつかのグラフ作成例の比較を通して、GTLでグラフを作成することの利点について紹介する。

SAS・R・Pythonにおける条件式の比較

○和田 尚磨、谷 里志

（株式会社タクミインフォメーションテクノロジー）

臨床試験における統計解析ではSASが長年標準ツールとして利用されてきたが、近年はRやPythonの活用も進んでいる。オープンソースソフトウェア（OSS）の普及や生成AIの登場により、SASエンジニアも複数言語の習得と効率的な条件ロジックの理解が求められている。本発表では、条件式の内部処理や記述方法に着目し、SAS・R・Pythonの相違点を比較検討した。

各言語における条件式評価の仕組みを整理し、SASのIF文／WHERE句の使い分け、Rのベクトル化の特性、Python のtruthy/falsy判定（Python）を例示した。また、臨床試験データの加工を想定したサンプルプログラムを提示し、条件ロジックを簡潔に記述する方法（IFC関数、WHICHC関数、ifelse関数、NumPyのnp.whereなど）を検証した。

SASは論理値を数値型で扱う一方、Rはlogical型、Pythonはbool型を採用しており、非ゼロ・非欠損の扱いに差異が見られた。フィルタリングや条件分岐においては、SASではWHEREが大規模データに有効であること、Rではベクトル演算による効率性、Pythonでは簡略化した条件記述が有用であることを確認した。さらに、関数利用により条件ロジックの可読性と保守性が向上し、異なる言語間でも同等の処理が実現できることを示した。

SAS・R・Pythonの条件式評価にはそれぞれ特徴があり、用途に応じた使い分けが重要である。今後、生成AIとのハイブリッド活用が進む中で、条件ロジックの整理と標準化はコードレビューや再利用性の向上につながる。さらに、本研究で得られた知見は教育・研修や多言語対応の実務にも応用でき、医薬品開発における品質保証の観点からも意義を持つ。将来的にはAIを補助的に活用し、効率性と信頼性を両立した解析環境の構築を目指したい。

SAS・R・Pythonにおける変数処理の比較

○谷 里志、和田 尚磨
（株式会社タクミインフォメーションテクノロジー）

臨床試験解析やデータマネジメント業務では、長年にわたりSASが標準ツールとして活用されてきました。しかし近年、RやPythonなどのオープンソースソフトウェア（OSS）の利用が急速に拡大し、医薬品開発やデータ分析の現場ではマルチ言語環境での作業が一般化しつつあります。こうした背景を踏まえ、言語ごとの仕様の違いを正しく把握し、適切に実装することは、効率化と品質確保の両面で不可欠です。

本発表では、SAS・R・Pythonにおける「変数型変換」「長い文字列の処理」「日付・時間変数の扱い」の重要な三領域を比較します。具体的には、SAS特有の固定長文字列、保存形式と表示形式の分離、RやPythonにおける可変長文字列と日付・時間オブジェクトの管理の特徴を整理し、コード移植の際に生じやすい非互換・不整合や処理エラーを防ぐための実践的なポイントを提示します。

さらに、生成AIの業務活用に関連して、コード生成と検証プロセスの効率化、エラー低減、コーディングの標準化ルールの策定といった取り組み事例についても触れます。本発表を通じて、マルチ言語環境における変数処理を安定かつ効率的に運用するための知見を共有し、現場での生産性と再現性の向上に資することを目指します。

SASの“あるある”をスマートに解決
実務に効く3つのマクロをパッケージ化

○鈴木 慎吾[1]、西井 雅之[1]、毎田 豊[1]、土生 敏明[1]、小原 裕美子[1]

([1]SAS Institute Japan)

SAS Institute Japanには、お客様の業務課題に対する解決策を検討・実現する部門としてコンサルティングサービス本部・カスタマーアドバイザリー本部、SASユーザー様のご利用促進を支援する部門としてカスタマーサクセス本部がございます。

これら部門の有志メンバーが、これまでSASプログラムを用いた実案件において頻繁に遭遇してきた“SAS固有の煩わしさ”とその解決法について議論し、今後の案件推進の加速、成果物の均質化に役立つと思われるマクロのパッケージ化を進めております。

本発表では、その成果として作成した以下の3つのパッケージをご紹介いたします。

- ✓ フォルダにあるデータセットの文字コードを一括判別し自動的にUTF-8変換
- ✓ フォルダにあるExcel/CSVを一括で読み込みデータセット化
- ✓ 指定ディレクトリ内（サブディレクトリも含む）の .sas ファイルを一括インクルード

PharmaForestでのパッケージ紹介

森岡　裕 Yutaka Morioka

（イーピーエス株式会社）EPS Corporation.

PharmaForest (https://github.com/PharmaForest) は，製薬業界におけるSAS業務を支援するSAS Packageを集約・共有するGitHubリポジトリであり，PHUSE Japan Open Source Technology Working Groupのサブチームメンバーにより運営されている．個人が抱えているSASノウハウを共有知へと変え，効率化や標準化，品質向上，コミュニティの活性化を目指すことを目的としている．本発表では，PharmaForestで現在公開されている20個以上のパッケージについて，それぞれの概要とユースケースを紹介する．

#	**Package**	#	**Package**
1	OncoPlotter (PharmaForest original package)	11	devil (PharmaForest original package)
2	sashash	12	xattr_kit
3	sas_dataset_json	13	cloak
4	SASPACer	14	vis_review_kit
5	misc (PharmaForest original package)	15	odstable_layoutshowcase
6	SAS FAKER	16	maxims4sas (PharmaForest original package)
7	SASLogChecker	17	laglead
8	sas_compare	18	ctcae5_grader
9	rtfCreator	19	love_matrix
10	shutter_chance	20	SASPACerShiny (R Package)
		21	change_tracer_dog



次世代に向けたSASの取り組みについて

○土生 敏明[1]、○吉原 也真人[1]、○西井 雅之[1]、成松 怜央[1]

（[1]SAS Institute Japan）

本セッションでは企画セッションという形で, SASが推進する先端技術と事例, 新ソリューションの紹介, ユーザーに対する新しい取り組みといった3つのテーマにてお話します．各内容の概略は以下の通りです.

1. SASが推進する先端技術と事例
   - ✓ SASの量子AI

     SASの量子AIに対する取り組みの最新動向・大規模最適化問題へのアプローチを紹介
   - ✓ SASによる合成データ

     SAS Data Makerを用いた合成データ生成技術ならびに事例
   - ✓ SASとLLMの組み合わせ

     SASと大規模言語モデル（LLM）を組み合わせたドキュメント業務効率化事例
2. ソリューション説明
   - ✓ SAS Viya Copilot for Clinical Data Discovery

     対話型にて臨床データの抽出・集計等を実施するソリューション
   - ✓ SAS Clinical Acceleration

     臨床データをSAS Viya上でレギュラトリーに準拠して取り扱うソリューション
3. SAS コミュニティサイトの御紹介
   - ✓ SAS Analytics Explorers

     SASユーザーが学習コンテンツやチャレンジを通じてスキルを磨き、バッジやポイントを獲得できるゲーミフィケーション型の学習コミュニティの御紹介
   - ✓ SAS Support Communities

     SAS製品に関する質問や情報交換をユーザー同士で行える、公式のオンラインフォーラムの御紹介

PharmaForestについて

山野辺　浩己

（マルホ株式会社）

PharmaForest: An Overview of a SAS Package Repository for Drug Development

Hiroki Yamanobe

(Maruho Co., Ltd.)

医薬品開発分野において広く利用されている統計解析ソフトウェアSASには、RにおけるCRANやPythonにおけるpipのような公式のパッケージ管理機能が存在せず、コード資産の共有や再利用に課題があった。Bartosz Jablonski氏が開発したSAS Packages Framework（SPF）は、この課題に対応する仕組みを提供し、フォーマット、関数、マクロ等をパッケージ化して管理・利用可能とした。これにより、SAS言語におけるコードの効率的な配布と再利用が実現され、開発効率や再現性が大きく向上した。

このSPFを基盤として、医薬向けSASプログラムのレポジトリ「PharmaForest」が創設された。PharmaForestは、PHUSE Japan OST（Open Source Technology）ワーキンググループのサブチームが運営する共同リポジトリであり、製薬業界に特化したSASパッケージを集約している。従来、個人や組織内に蓄積されがちであったSAS開発のノウハウをオープンに共有することで、知識の集積を促進し、生産性や品質の向上、標準化、さらにはユーザーコミュニティの活性化を目指している。

PharmaForestに収載されるパッケージは、統計解析業務で汎用的に利用される関数群や、CDISC標準に基づくデータ処理テンプレート、さらには日常業務の効率化を支援する補助的ツールまで幅広く、利用者は容易に導入・活用できる。これにより、重複作業の削減や再現性の確保が進み、医薬品開発におけるデータ解析の質と効率を同時に高めることが期待される。また、オープンリポジトリとして持続的に改善・拡張される仕組みを備えており、ユーザー参加型の発展が見込まれる点も大きな特徴である。

PharmaForestは、SAS利用の効率化を支援するだけでなく、製薬業界における共同開発やオープンサイエンス推進の観点からも重要な役割を担う。本発表では、PharmaForestの概要、設立経緯、レポジトリの構造や運用方針に加え、公開済みパッケージや導入事例、さらに今後の開発予定を含む活動状況と展望について紹介する。

Estimand初心者集まれ！

Estimandの復習と解析業務への影響点の紹介

○菊池 宏和、中西 豊支

（シミック株式会社）

本発表は、Estimand初心者を対象として、ICH E9(R1)で定義されたEstimandの基礎および解析業務への具体的な影響点について、分かりやすく解説する。
まず、ICH E9(R1)で示されたEstimandの5要素、中間事象（Intercurrent Event）とその対応ストラテジー、ならびに感度分析の定義について、基礎から体系的に説明する。

次に、Estimandの定義が統計解析業務にもたらす変化について、具体的な事例を用いて紹介する。特に仮想ストラテジーを適応した状況下において、MMRMによる解析を行う際のデータハンドリングの留意点や、SASを用いた解析事例を示し、実践的な対応方法について論じる。
さらに、感度分析の一例としてTipping Point Analysisを取り上げ、SASを用いた解析事例を通じて理解を深めることを目指す。

本発表により、Estimandの基礎から実務に応用可能な解析手法まで、体系的知識の習得につながるものとする。

JMPによるCDISC標準サンプルデータ(SDTM)の可視化手法の紹介

○中野 薫、坂本 佳代子、上野 恵子

（株式会社タクミインフォメーションテクノロジー）

JMPは、プログラミングの知識がなくてもマウス操作のみでグラフ作成や統計解析が可能なため、日常的にコードを扱わない研究者にとっても扱いやすい可視化ツールである。

本発表では、北海道大学病院 医療・ヘルスサイエンス研究開発機構 データサイエンスセンターが公開するCDISC標準サンプルデータを用いて、SDTMのLBドメイン（臨床検査）およびAEドメイン（有害事象）を可視化した事例を紹介する。

1. データ準備と加工

LBドメインにDMドメインを結合、同様にAEドメインにDMドメインを結合。

LB(+DM)ドメインに可視化を補助する列を追加：

・Page Title列：LBCAT、LBTEST、LBSTRESUを結合してグラフのタイトル要素とした。

・Changes from Baseline列：LBBLFL="Y"を基準に変化量を算出。

・Changes from Baseline (%)列：前述の変化量の割合を算出。

AE(+DM)ドメインに可視化を補助する列を追加：

・BODSYS_DECOD列：AEBODSYS、AEDECODを結合してグラフの要素とした。

2. グラフの作成

各ドメインについて、以下のような多面的なグラフを作成した：

LB(+DM)ドメイン : 散布図に平均線を重ねたグラフ／症例別推移図／時点ごとの散布図＋平均値オーバーレイ／要約統計量の表

AE(+DM)ドメイン：重症度ごとの発症数を示す棒グラフ／重症度ごとの有害事象の度数分布表

3. インタラクティブ機能の追加

可視化の操作性を高めるため、以下の機能を追加した：

・列スイッチャーを用いて列の切り替えを実現（例 : 測定値の切替 : LBSTRESN ／ Changes from Baseline）

・ローカルデータフィルターを設定し、項目（例 : LBTEST、LBCAT、SEX）による動的な絞り込みを実現。

・JMPプロジェクト機能により、複数グラフを一画面に統合。タブの切替により直感的な閲覧を実現。

本発表では、グラフの作成を手作業で行ったが、データ結合や基本的な可視化処理をJMPスクリプトで自動化し、アドイン化することで、JMPに不慣れなユーザーでも一連の処理を容易に再現できる。

Enhancing Statistical Analysis Workflows: LLM-Driven Approach with SAS HTTP Procedure

○高浪 洋平[1]、仲家 諒[1]

([1]武田薬品工業株式会社)

近年、大規模言語モデル（LLM）は、ビジネスプロセスや日常業務を効率化するツールやアプリケーションを開発する上で重要な役割を果たしている。医薬品開発の統計解析・プログラミング業務においても、統計解析担当者や統計プログラマーが、解析計画書、データセットや解析帳票の仕様書、プログラムの作成、統計解析の実施やレポート作成など、多くの関連する成果物の作成においてLLMの利活用を進めている。LLMユーザーは、Python等のプログラミングコードを通じてAPI経由でLLMに容易にアクセスできるが、SASではHTTPプロシジャを使用してLLMにプロンプトを送信し、回答を受け取ることができる。本発表では、HTTPプロシジャによるLLMへのアクセス方法を紹介するとともに、統計解析業務の効率化を考察する。

- HTTPプロシジャ

HTTP（Hyper Text Transfer Protocol）は、WebサーバーとWebクライアント（例：Webブラウザ）がデータを送受信するために使用されるプロトコルであり、主に、Webページを構成するHTMLファイルや画像、動画などのデータをやり取りする際に利用される。SASのHTTPプロシジャは、HTTPリクエストを作成して、Webサーバーや Webサービスからデータを取得したり送信したりすることができる。HTTPプロシジャの主な構文・オプションを以下に示す。

```
PROC HTTP URL="URL-to-target" METHOD="http-method" IN=fileref OUT=fileref CT = "content-type" ;
```

- HTTPプロシジャによるLLMへのアクセスと統計解析業務効率化の考察

本発表では、Google Gemini 2.5 Pro（以下Gemini）を用いて、SASプログラムを自動生成して実行する枠組みを考察する。まず、LLMに与えるPromptをExcelファイルからJSONファイルに変換し、次に、SASのHTTPプロシジャでGeminiへアクセスしてその応答をJSONファイル形式で取得する。続いて、LibnameのJSONエンジンを用いて応答ファイルをSASデータセットへ変換し､その中に含まれる実行可能なSASプログラムを抽出する。さらに、call execute文を用いて、抽出したSASプログラムを動的に実行し、最終的に解析結果をRTF形式で出力する。本手法により、LLMを介したSASプログラムの自動生成から実行・成果物出力までを一連のプロセスとして以下の処理フローを単一のSASプログラムで実現できることを示す。



SASによるMMRM（反復測定のための混合効果モデル）の実装と臨床試験への応用（加齢黄斑変性（AMD）に伴う地図状萎縮（GA）進行抑制効果の解析を例として）

○ZU YANHUI[1]、ZHAO YUNXI[2]

（[1]Clinchoice株式会社、[2] Clinchoice株式会社）

MMRMは「Mixed Model Repeated Measures」の略で、線形混合効果モデルの一種で、不完全な経時測定データを解析するために利用される統計モデルである。特に、同じ患者さんから複数回測定されたデータ（例えば、投薬前と投薬後、1週間後、2週間後…といった時系列データ）を解析する際によく用いられる統計手法である。

臨床試験では、反復測定データに基づく治療効果の評価が不可欠であるが、欠測データの存在や時間効果の調整が課題となる。本発表では、加齢黄斑変性 (Age-related Macular Degeneration, 略称: AMD)における地図状萎縮 (Geographic Atrophy, 略称: GA)進行の評価を例に、MMRMの眼科領域への実装方法を解説する。薬剤AのGA病変進展抑制効果を検証するため、24ヶ月間の多施設共同二重盲検ランダム化比較試験を実施した。

統計モデル:

$$\%Change=\beta0+\beta1\ base+\beta2\ trt01pn+\beta3\ avisity+\beta4(trt01pn\times avisity)+\beta5(base\times avisity)+\beta6\ gapcat+\beta7\ mltcat+\beta8\ fovcat+\epsilon$$

に基づき、SAS PROC MIXEDを用いた解析において、①両眼データの取り扱い、②眼科特有の共変量調整、③経時的変化のモデル化など、臨床試験における実践的な応用ポイントを紹介する。主要評価項目は、ベースラインからのGA病変面積の変化率 (%Change)であり、負の値はGA面積の縮小 (改善)率、正の値は拡大 (悪化)率を示す。

本研究では、MMRMモデルが眼科臨床試験のデータ解析に有効であることを示した。特に、経時的なデータ解析や両眼データの取り扱いにおいて、その優位性が確認された。今後の臨床研究において、本手法を活用することで、より信頼性の高い治療効果評価が可能となると考える。

Pharmacometrics解析のための
データセット作成プロセスの改善
SASとRのCollaborationの提案

○碇 新里[1]、Dimple Patel[2]、永田吉範[1]、Douglas Marsteller[2]

（[1]協和キリン株式会社 開発本部 医科学薬理部、[2]Kyowa Kirin, Inc.）

目的: PMx解析用データセットの準備は複雑で時間がかかり、高度なデータ専門知識を必要とします。特に複数の臨床試験からデータを統合する母集団解析では、各試験のSDTMの変数定義や構造に若干の違いがあり、1つの解析用データセットへと統合するためには深い理解が必要です。また、モデル解析のプロセスでは頻繁にモデル構造を変更するため、解析用データセットはモデル構造に合わせた修正が必要になることもあります。この解析用データセットの作成にかかる時間を短縮し、品質を確保することは臨床試験中のタイムリーな解析と効率的なフィードバックを可能にします。

方法: 時間的、品質的課題の改善のために、SDTMプログラマーがSDTMドメインと並行して統合データセットを作成するプロセスを考案しました。この統合データセットは、関連するSDTMのドメインをLBドメインに類似したシンプルな縦型構造の一つのデータセットに再構築したものです。この統合データセットの仕様書はSDTMの仕様書に類似した様式を用いました。統合データセットそのものにはPMx特有のデータセット構造を含めませんでした。代わりに、PMx解析に必要な情報は統合データセットの変数の定義を最適化することで織り込みました。統合データセットからモデル構造に適した解析用データセットへの変換はRパッケージであるIQRtoolsによってsemi-automaticに行うことができます。

結果: 作成した統合データセット仕様書を用いて、SDTMプログラマーはSDTMの作成後、速やかに統合データセットを作成することができました。統合データセットの変数の定義は異なるphaseや適応症が異なるstudyに対応できるよう考案しました。常に含める基本的な指標に加え、疾患や薬剤プロファイルに依存するオプションパラメータも容易に追加可能です。

結論: このデータセット作成プロセスの改善により、SDTMプログラマーがSDTMと並行して統合データセットの作成を進行できます。このプロセスで作成された統合データセットはシンプルな縦型構造であり、試験デザインに依存しないため、各phaseの試験データを縦に積み上げることで統合解析に利用することができます。

AIエージェントとSAS Intelligent Decisioningの融合

～　CRMと与信を超えて広がる新しいパラダイム～

○　小野 潔
（コムチュア株式会社 デジタルイノベーション本部）

抄録：

AI エージェントは、大規模言語モデル（LLM）を基盤とし、自然言語を通じて情報処理や意思決定を支援する新しい枠組みである。本研究では、オープンソース基盤 Dify と SAS Intelligent Decisioning（ID）の役割を整理し、両者を統合したエージェント駆動型 CRM モデルを提案する。具体的には、Dify が顧客プロファイル生成や施策案提示を担い、ID がルールベースの制御と説明可能性を保証する二層構造を設計した。この仕組みにより、生成 AI の柔軟性と ID の堅牢性を補完的に活用し、透明性とガバナンスを確保した形で実務に適用することが可能となる。さらに本研究では、ChatGPT エージェントを加えた三層モデルを将来的展望として位置づけ、金融・マーケティングを超えて医療・物流・公共政策など多様な分野に拡張可能な、新しい意思決定基盤としての可能性を示す。

現場で役立つSAS活用術：リアルワールドデータ解析の実践

小川武則[1]、金沢 信一[2]

([1]メディカル・データ・ビジョン株式会社、[2]エイツーヘルスケア株式会社)

近年、医薬品の有効性や安全性を評価する手段として、リアルワールドデータ（RWD）を用いたデータベース研究が注目されている。本セミナーでは、メディカル・データ・ビジョン株式会社（MDV）が保有する5000万人超のDPC（診療群分類包括評価）データおよび健康保険データの概要を紹介し、SASを用いたRWD解析の基本的なアプローチから実務的なテクニックまでを解説する。特に、膨大なデータを扱うDB研究において重要となる前処理工程については、エイツーヘルスケア株式会社（A2）が開発したMDVデータ専用加工ツール「BluePat（ブルーパット）」を用いた処理方法を取り上げる。また、MDVが提供するフィージビリティツール「MDV analyzer for Clinical Insight」の機能と、データベース研究への応用可能性についても紹介する。

SASプログラミングコンテスト

リバーシーチャンピオンシップ



SAS ユーザー総会 2025

SASユーザー総会 2025

Executive summary

概要:
昨年のSASユーザー総会のアンケートに「プログラミングコンテストを開催してほしい」という声があり、リバーシーの思考部分を題材としてコンテストを企画。
出場希望者は、配布するプログラム開発用対戦プログラムを利用し、盤上データから最適手を選択するプログラムを作成し提出する。

期待される成果:
出場者のプログラムはウェブサイトで公開し、SAS学習者の参考資料とする。
作成したプログラム(処理・思考部分等)を題材とした論文・プレゼンテーション発表。

評価基準:
SASユーザー総会当日、会場にて対戦し、成績優秀者を決定。

表彰と特典:
成績優秀者は表彰式にて表彰。
発表者と異なり参加費無料等の特典は付与しない予定(別途、論文・プレゼンテーションの発表を期待します)

出場資格:
SASユーザー総会2025 発表者または参加者(コンテスト実施時は会場での出席をお願いします)

スケジュール: **プログラム・自己紹介スライド 締切: 9月1日** **SASユーザー総会 当日: ライブ対戦**	**最小催行人数:** **3名**	**担当:** **井上 貴博(takahiro.inoue@novartis.com)** **出場希望者は井上までご連絡ください。**
プログラム条件: **詳細は次項参照**	**開発環境:** **SAS OnDemand for academic**	

プログラム・参加条件

- コンテスト実施時は会場にご出席ください。
- マクロで作成ください。マクロ名は世話人より指定します(姓5文字+名2文字+連番の予定)
  例: ITOYO1, MORIOYU1, INOUETA1 (連番は重複がなければ1)
- 出場者のマクロをSASユーザー総会の公式HP(ウェブサイト)で公開することに同意ください。
  SAS学習者に参考になるようにコメント追記等、分かりやすい構文を心掛けてください。
- 盤面データとして、SASデータセット「_info」を読み込み、
  選択するセルを、グローバルマクロ変数「_xy」にx,yで入れてください。
- 対局開始時に黒または白の担当が決まります。どちらでも動作するプログラムを作成ください。
  右記の場合、17手目[No]で指し手 [Turn]は黒(-1)で、黒の指し手担当となります。
  (対局中、白黒の担当が変わることはありません)
  変数[Y, X1-X8]は、前手(16手目)までの盤面で、黒(-1), 白(1)の状態を示しており、
  17手目で指し手(黒)が置けるセルを選択可能セルとして99が入力されています。
- 選択可能セル(99)から一つ選び、グローバルマクロ変数「_xy」にx,yの形式で入力ください。
  例(X=7, Y=3): 7,3
- マクロ内で、_(アンダースコア)から始まるデータセット名、マクロ変数名は作成しないでください。
  文字変数を使用する場合、入力値は1バイト文字のみをご利用ください。なお、コメントには日本語等を利用可能です(分かりやすい構文を心掛けてください)。
- 対戦プログラムで「_info」の他にもデータセット等が作成されますが、使用できません。
  コンテスト当日はライブ対戦用のプラットフォームは別のプログラムを使用するため、「_info」を除いた他のデータは別の名前と構造になります。
- 動作不良は負けとなります。対局は対戦者同士のマクロと交互に実行されます。データセット名の重複に注意してください。細かい命名規則等は定めませんので、原則、1手ごとに独立で動く(数手前のデータセットを利用しない)マクロを推奨します。
- ライブ対戦の関係上、1手何秒以内という制限を設ける予定です(合計で何分以内ではありません)。
  SAS OnDemandで実行した際に、1手が3秒以内(4秒未満)を目安としてください。
  変数[No]で、序盤、中盤、終盤で処理内容を変える等の工夫をお願いします。

Length No Y Turn X1 - X8 8. ;
No: ○手目
Turn: 指し手の色
　(黒 -1、白 1)
Y: 上から○マス(行)目
　(Y=1は1行目、Y=2は2行目...)
X1: 左から1マス(列)目
　(黒 -1、白 1、選択可能セル 99)
X2～X7: 左から2～7マス(列)目
X8: 左から8マス(列)目

提供データ(_info.sas7bdat)

No	Turn	Y	X1	X2	X3	X4	X5	X6	X7	X8
17	-1	1	0	0	0	0	0	0	0	0
17	-1	2	0	0	0	0	99	99	0	0
17	-1	3	0	0	-1	-1	1	1	99	0
17	-1	4	0	0	-1	-1	1	99	99	99
17	-1	5	0	-1	-1	1	1	1	1	1
17	-1	6	0	99	1	-1	-1	-1	99	0
17	-1	7	0	0	99	99	-1	1	99	0
17	-1	8	0	0	0	0	0	99	99	0

<イメージ>

Turn	Y	X1	X2	X3	X4	X5	X6	X7	X8
●	1								
●	2					99	99		
●	3			●	●	○	○	99	
●	4			●	●	○	99	99	99
●	5		●	●	○	○	○	○	○
●	6		99	○	●	●	●	99	
●	7			99	99	●	○	99	
●	8						99	99	

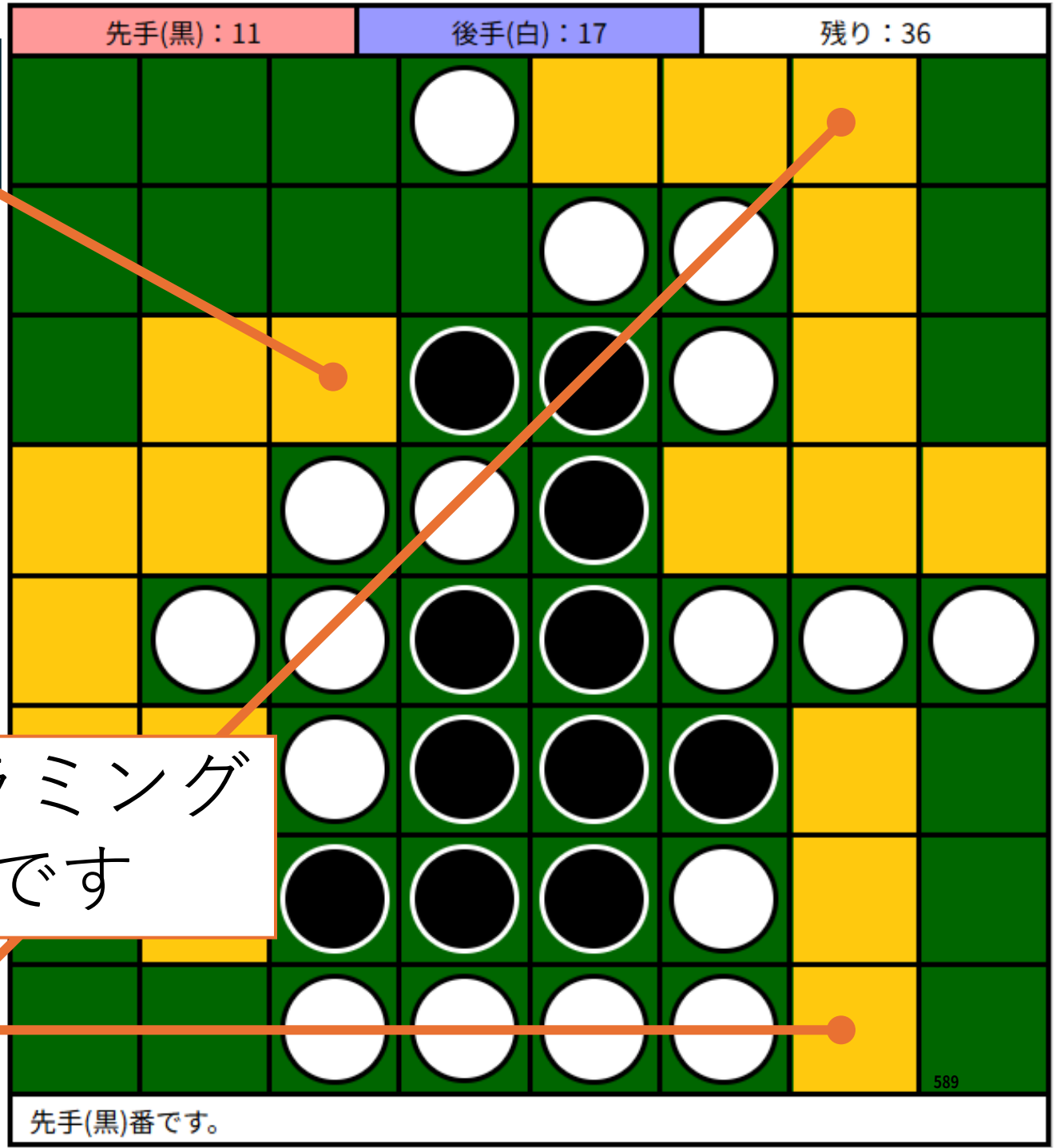
一番多く反転するけど
次で戻される

先手(黒)です
どこに置く？

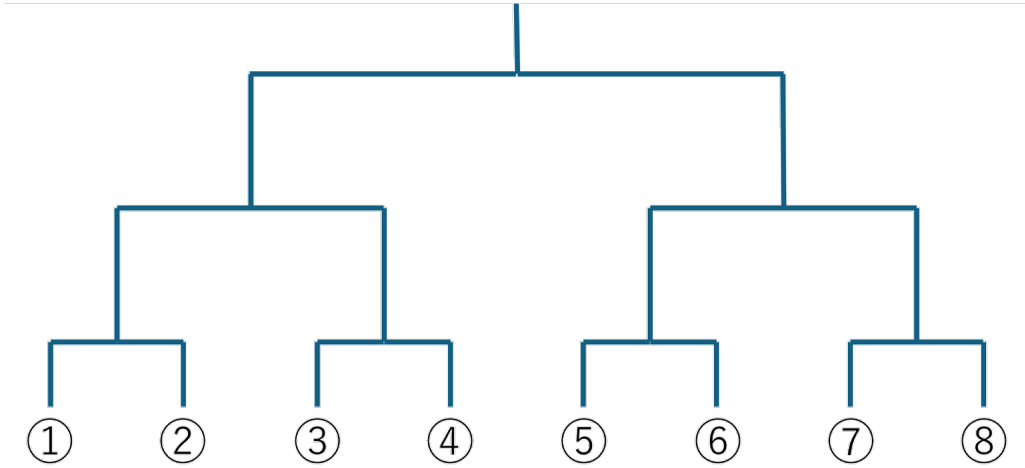
黄色が置ける
ところ

こういった思考をプログラミング
強い人がチャンピオンです

ここに置くと角が
取られやすい？



対戦形式



- 対戦者2名のプログラムが
  先手(黒)、後手(白)に分かれ、
  リーバシーで対戦。
- 勝ち残り式トーナメント
- 基本ルールは一般的なリーバシーのルールに従う。
  制限時間内に応答がない場合、負けとなる。
- 先後入替で2戦予定(時間がない場合は1戦のみ実施)
  1勝1敗の場合、獲得セル数の合計が多いほうが勝ち。
  (終局時の空白セルは勝者の獲得セルとしてカウント)
  (同数の場合は世話人の独断で勝敗を決定)

対戦結果

チャンピオン(優勝):
飯島泰平
[ネイチャーインサイト株式会社]
準優勝:
飯島康浩

(黒)24-40(白)
(白)9-55(黒)

(黒)31-33(白)
(白)50-14(黒)

(黒)14-50(白)
(白)16-48(黒)

(黒)21-43(白)
(白)46-18(黒)

(黒)38-26(白)
(白)19-45(黒)

(黒)22-42(白)
(白)27-37(黒)

(黒)41-23(白)
(白)32-32(黒)

①	②	③	④	⑤	⑥	⑦	⑧
IIJIMYA1 飯島康浩	KAWAKTA1 川上貴弘	IZUMITO1 伊住寿雄	SEKINSH1 関根伸吾	SEKINSA1 関根暁史	OTAYU1 太田裕二	IIJIMTA1 飯島泰平	HIRAIKE1 平井健太

※組み合わせは事前に抽選で決定

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

① IIJIMYA1

お名前（ご所属）	飯島 康浩
プログラムの特徴	角や辺を意識した定石を詰め込んだロジックです。また、序盤は中央を避け、終盤に向けて盤面を広く使う設計です。ランダム要素も少し残して、運も味方にしたいです。
コメント	何とか最後まで戦える展開になることを願っています。

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

② KAWAKTA1

お名前（ご所属）	川上　貴弘（DXC Technology Japan株式会社）
プログラムの特徴	自分の打った手が次の相手にどのように影響するかをチェックして、できるだけ不利にならないような手を選ぶようにした。
コメント	趣味でSASプログラマーをやっている者だ。

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

③ IZUMITO1

お名前（ご所属）	伊住 寿雄（ソーバル株式会社）
プログラムの特徴	1手先読みを行い、相手が石を反転させる数の最大値が最小＆自分が反転させる石の最大の場所を選択する。（minimax法というそうです。） 実行速度を上げるために少ないデータステップで処理する。 その為にはハッシュオブジェクトを多用する。
コメント	ハッシュオブジェクトを多用することにより、極力ディスクの読み書きを減らす＆メモリ上での処理を増やしたため、処理する項目は多いものの意外と実行時間は早いものができたと思います。

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

④ SEKINSH1

お名前（ご所属）	関根伸吾(ネイチャーインサイト株式会社)
プログラムの特徴	探索を50000手(約3秒)を超えるまで実施している。 次の手の候補をData Stepで一気に計算することにより、速度の向上につながっている。 スコアとして自分、相手、空きの角、辺の数、全体の数、各マスのスコアを組み合わせ、スコアが最も高い手をMin-Max法を用いて実装した。
コメント	まずは、面白いコンテストを開催していただきありがとうございました。 SASユーザー総会は登壇してみたい気持ちがありましたが、論文を書けるほどではないので、このような取り組みがありよかったです！ プログラムについては、さまざまな特徴量を作って学習させたが、実際には当初のシンプルな重みが一番良かったので、そちらを採用しました。1勝でもしたいです。

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

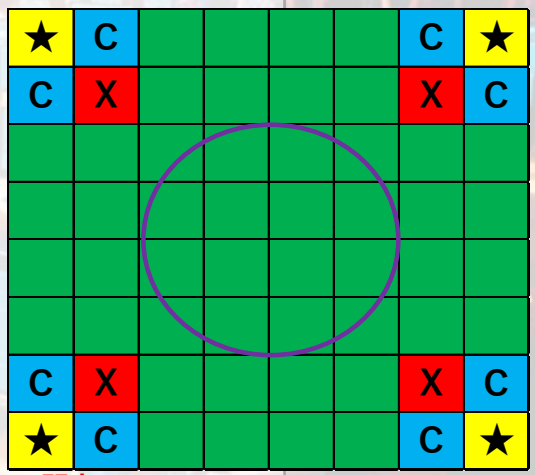
⑤ SEKINSA1

お名前（ご所属）	関根 暁史（藤本製薬株式会社）
プログラムの特徴	回帰分析や主成分分析のアルゴリズムを取り入れた、オセロからかけ離れた何とも不思議なプログラムです。（プロシジャ由来のWARNINGが出ることはあります。）
コメント	四隅だけは取らせません！

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

⑥ OTAYU1

お名前（ご所属）	太田　裕二（ノーベルファーマ株式会社）
プログラムの特徴	• AIに随時盤面を共有し、最善手を聞き出せたら強いかも（私の技術的にハードルが高い） • すべての実現可能盤面に対して最善手をAIから教えてもらって事前にリストを作成 実現可能盤面は一説には10^{28}通りもあるとのこと、制限時間もあるし。。。 ↓ 図書館で初心者向けの本を調査、初心者向けテクニックとして、 下に挙げた5つ（実装できたのは4つ）をピックアップし、 ゴリゴリのデータハンドリングのみで最新技術に勝負を挑む！！ ① 取れる角★はとる！→複数ある場合は、ランダムに選択 ② 角を相手にとられるリスクが高くなる角の斜め内側（Xと呼ばれるらしい）と 角の上下左右（Cと呼ばれるらしい）以外からできるだけ中心に近いところをとる →複数ある場合は、ランダムに選択 ③ 序盤はできるだけ裏返すセルを少なくする →AIに提案してもらったものの十分に理解できなかったため実装しなかった。。。 ④ 角の上下左右を取らざるを得ない場合は、 隣接する角が自分のセルになっているところを優先→複数ある場合は、ランダムに選択 ⑤ 角の斜め内側を取らざるを得ない場合は、 隣接する角が自分のセルになっているところを優先→複数ある場合は、ランダムに選択
コメント	ランダムに選択する状況が多いため、運を味方につける必要あり！ →乱数のシードには20030606を使用（弊社の創立記念日2003年6月6日より）



SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

⑦ IIJIMTA1

お名前（ご所属）	飯島 泰平（ネイチャーインサイト株式会社）
プログラムの特徴	自分の手番の候補手を打った後の盤面を全通り探索し、様々な観点での点数の合計値である優先スコアを算出し、優先スコアが最大の候補手を選択する。 1手先しか読まないので、比較的計算が早い。 例外として、序盤2手目（白）は縦取り、3手目（黒）は縦取りの場合には斜め取り（虎定石）に固定。 【ご参考】優先スコアの観点：次の相手の選択肢が少ない。中盤まで自分の石数が少ない。真ん中の4石に自分の石が多い。自分の確定石（ひっくり返されない石）が多い。相手に角を取られにくい。その他、辺の攻防など。
コメント	オセロ経験者（約20年前の少年時代）＆SAS使いなので参加するしかないと思ってやってまいりました。自分の分身を作るつもりで手動で改良していきまして、そこそこマシな手が打てるプログラムになったかと思います。しかし1手先しか読んでいないので、たまにやらかしてしまうかもしれません。大悪手が出ないことを祈りつつ、いい勝負になればいいなと思います。 対局よろしくお願いいたします！

SASユーザー総会 2025
SASプログラミングコンテスト
リバーシーチャンピオンシップ

⑧ HIRAIKE1

お名前（ご所属）	平井健太（株式会社タクミインフォメーションテクノロジー　ライフサイエンスグループ）
プログラムの特徴	複雑な処理は避けて、単純な処理の積み重ねを意識しました。 盤面を4隅、4隅周り、中央の3区画で考えてそれぞれで処理を分けました。 中盤までと、終盤でそれぞれ別のロジックで優先度をつけて、より有利（と思われる）なマスほど重みが高くなるようにしました。
コメント	サンプルで提供いただいたランダム置きに対して勝率8割程度を目指しましたが、それは達成できたかと思います。 ＋1手先を読むと実行時間がかかってしまったので、実行速度とプログラムの複雑さのバランスをとるのに苦心しました。 予期しないエラーが出ないことを祈っています。

主催：SAS ユーザー会 世話人会

代表世話人 伊藤 陽一 北海道大学病院
世話人(50 音順)
井上 貴博 ノバルティス ファーマ株式会社
上村 鋼平 東京大学
岸本 淳司 九州大学 ARO 次世代医療センター
菅波 秀規 興和株式会社
土生 敏明 SAS Institute Japan 株式会社
林 行和 エイツーヘルスケア株式会社
森岡 裕 イーピーエス株式会社

SAS ユーザー総会 2025 についての問い合わせ先：info@sas-user2025.ywstat.jp

協賛(50 音順)

イーピーエス株式会社
エイツーヘルスケア株式会社
株式会社エス・シー・エー
SAS Institute Japan 株式会社
塩野義製薬株式会社
シミック株式会社
スタットコム株式会社
武田薬品工業株式会社
株式会社タクミインフォメーションテクノロジー
メディカル・データ・ビジョン株式会社
株式会社レビティジャパン