

DVD-ROM
添付



無料統計
ソフト

EZRでやさしく学ぶ統計学

EBMの実践から臨床研究まで

[改訂4版]

神田善伸 ● 著

自治医科大学附属病院・
附属さいたま医療センター
血液科教授

中外医学社

改訂 4 版の序

前版から約6年を経て、第4版の発刊に至りました。本書は無料統計ソフト EZR の公式マニュアルの位置づけとなっています。EZR は幅広くご好評をいただき、日本国内の学会だけでなく Asia Pacific Blood and Marrow Transplantation (APBMT)、International Congress of BMT (ICBMT) などの国際学会でも EZR についての招聘講演の機会をいただくようになりました。EZR の開発と使用方法について紹介した論文 (2013 年 Bone Marrow Transplantation 誌掲載) は、2026 年 5 月時点で 15,000 編以上の論文に引用されています (Web of Science 調査)。引用論文の内訳は腫瘍領域を筆頭に外科領域、血液学領域、消化器領域、循環器領域等と多岐にわたっており、NATURE 誌、JOURNAL OF CLINICAL ONCOLOGY 誌、CIRCULATION 誌 (いずれもインパクトファクター 40 前後) などの一流誌の論文でも引用されています。多くの研究者から信頼をいただいていることと受け止めています。

第3版発刊時の EZR のバージョンは 1.52 ででしたが、現在はバージョン 1.70 までアップデートしています。この間にグラフ作成機能を大幅に強化し、スイマープロット、(サブグループ解析の) フォレストプロット、サンキーダイアグラム、CONSORT ダイアグラム等を描画できるようにしました。また、二値変数を従属変数とする多変量解析において、ポアソン分布にあてはめることによってリスク比での表示を可能にする修正ポアソン回帰などの解析機能を追加しています。本書第4版は、これらの解説に加えて、かねてから要望の多かった予後予測モデルの構築方法について、データ分割法やブートストラップ法によるモデルの検証、スコア化とノモグラムの作成、Lasso 回帰、欠損値の多重代入法の詳細を紹介しています。

今後もバージョンアップなどの案内は自治医科大学血液学部門のホームページ (<https://www.jichi.ac.jp/usr/hema/EZR/statmed.html>) で行います。質問、要望などは中外医学社ホームページのお問い合わせフォーム (<https://www.chugaiigaku.jp/inquiry/inquiry01.php>) にお寄せください。全てのお問い合わせに回答することはできませんが、可能な範囲でお答えいたします。頻度の高い質問についてはホームページに FAQ として掲載していますのでご覧ください。

EZR 動作環境

- ・ Windows XP、Vista、7、8、8.1、10、11 (Arm 版を含む)
- ・ Mac OS 10.5 (Leopard) ~ 26 (Tahoe)
- ・ Linux Ubuntu 11.10 ~ 24.04
- ・ XGA (横 1024 × 縦 768 ドット) 以上の画面サイズが必要 (縦のドット数が 768

目次

第1章 統計の基礎知識 1

1. なぜ統計解析が必要なのか？	1
2. 変数の種類とその要約	2
①変数の種類	2
②変数の要約、信頼区間	2
3. 群間の比較、 P 値とは？	4
4. エフェクトサイズ	6
5. 多重比較の問題	8
6. 回帰、単変量解析と多変量解析	9
7. バイアス	10
8. 多変量解析と変数選択	13
9. EBM と臨床研究	19
10. 観察研究と介入研究	21
11. ケースコントロール研究とコホート研究	22
12. 介入研究の種類	24
13. 主要評価項目	27
14. Intention-to-treat (ITT) 解析	28
15. 必要症例数の設定	28
16. 臨床試験のプロトコール	30
17. 臨床研究の倫理	32
18. 臨床試験の報告に必要な要素	33
19. メタアナリシス	34
20. 臨床決断分析	35

第2章 EZR のインストール 39

1. Windows Vista、7、8、8.1、10、11 のユーザーアカウント 制御に関する問題点	39
--	----

2. EZR の構成要素とインストール方法の概要	39
3. Windows 専用 EZR インストールプログラム EZRsetup.exe を 用いたインストール	40
4. データやスクリプトを保存するフォルダーの作成	46
5. EZR の設定の変更	48
6. EZR を終了する	51
7. EZR のアンインストール	53
8. インターネットからの手作業での EZR のインストール (Windows の場合)	53
9. R のアンインストールの方法	62
10. R のアップデートの方法	62
11. Mac OS への EZR のインストール方法	63
12. Linux への EZR のインストール方法	66
13. トラブルシューティング	68

第3章 R のスクリプトに触れてみる 69

第4章 R の基礎知識 79

1. データの型とデータを格納する形式	79
2. ベクトル	80
3. 行列	83
4. リスト	86
5. データフレーム	86
6. データフレームの読み込み、作成、保存	87
①ベクトルからのデータフレームの作成	87
②既存のファイルを読み込んでデータフレームを作成する方法	88
③データエディタの使用	91
④複数のデータフレームの結合	93
⑤データフレームの保存、読み込み	93
7. データフレームの中の要素へのアクセス	94
8. 関数の定義	96
9. 引数の省略と初期値	98

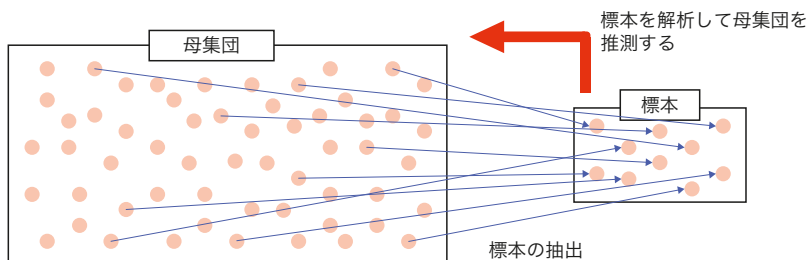
1 なぜ統計解析が必要なのか？

人間は自分自身の経験に基づいて、感覚的にものごとを判断しがちである。例えばある疾患に対する標準治療薬の有効率が50%であったとする。そこに新薬が登場し、ある医師がその新薬を5人の患者に使ったところ、4人が有効と判定されたとしたら、多くの医師はこれまでの標準治療薬よりも新薬のほうが有効性が高そうだと感じることだろう。しかし、たまたま有効性が出やすい5人に治療が行われたにすぎないかもしれない。同じ疾患を有する患者であったとしても、疾患の細かな分類や進行度、患者の年齢、性別、臓器の状態などによって有効率は左右される。さらに背景の条件が全く同じであったとしても有効率にばらつきは生じる（偶然に高い有効率になることもある）。治療に対して思い入れが強ければ強いほど、治療結果に大きく一喜一憂し、客観的な評価が困難となる。印象に残る結果は感覚的な判断を偏らせてしまう。

統計解析の目的の1つは、前提としてこのような様々なばらつきが存在する状況の中で、限られた**標本 (sample)** から**母集団 (population)** を**推測**し、より一般的な結論を導き出そうとすることである。**母集団**の定義は状況によって異なるが、例えばある疾患に対する新薬の有効性を評価するのであれば、その疾患を有する全ての患者が母集団となる。統計解析をしていると、目の前にあるデータだけを対象としているような錯覚にとらわれることがあるが、実際に行っていることは、その標本を用いて本当の**母集団の全体像を推定**しようとしているのである（選挙の出口調査による全体の投票数や議席数の予測をイメージすればよい）。



統計解析の目的の1つは、母集団から抽出した標本（サンプル）を用いて解析することによって、母集団を推測することである



第9章 生存期間の解析

1 生存期間の要約

医学研究では**評価項目**（従属変数、目的変数）が連続変数になることは意外に少なく、二値変数（有効、無効、生死など）あるいは**生存期間の解析**（死亡までの期間、再発までの期間など）であることが多い。例えば高血圧の治療や脂質異常症の治療などでは、短期間の評価項目としては血圧、コレステロール値などの連続変数になるが、やはり長期的には心血管系イベントの発生率、発生までの期間が**より重要な評価項目（hard endpoint）**となる。

生存期間（あるいはあるイベントの発生までの期間）を要約する場合は**生存曲線**を描くのがわかりやすい。生存曲線とは、ある明確に定義された**イベント（出来事、エンドポイント）**の発生を時間に対してプロットした曲線であり、そのイベントは必ずしも死亡である必要はなく、1度だけ生じるものであればよい（複数回生じるイベントの解析には別の解析方法が必要）。単にイベント発生までの期間を連続変数として解析する方法も考えられるが、この場合は解析時点でまだイベントが発生していないサンプル（すなわち解析時点で観察が**打ち切り**となっているサンプル）を正しく扱うことができない。生存期間の解析の特徴は**打ち切り（censor）**サンプルを扱うことができる点である。例えば生死を扱う場合、解析時点で生存中の患者や、ある時期まで生存したことは知られているが、その後について有用な情報がない患者などを打ち切りとして扱う。ただし、その患者が観察から脱落した理由が重要であり、その理由が生存に影響するものであれば解析結果に偏りを生じる可能性があるため注意が必要である（例えば末期がん患者のホスピスへの転院による観察終了）。単に生死を評価する場合は**全生存率（overall survival, OS）**、疾患がない状態での生存を評価する場合は**無病生存率（disease-free survival, DFS）**、腫瘍の増悪がない状態での生存を評価する場合は**無増悪生存率（progression-free survival, PFS）**などの用語が用いられる。全生存率は全患者が対象となっていることが多いが、無病生存率は完全寛解状態となった患者だけを対象として寛解到達時点を観察開始時期にしている場合が多く、**対象症例**や**観察開始時期**が異なることが多いので注意しなくてはならない。また、**無イベント生存率（event-free survival, EFS）**では**イベントの定義**が何であるかも確認する必要がある。

生存曲線の作成には**生命保険数理法**や**Kaplan-Meier法**などが用いられる。前者では月や年などの一定間隔ごとに生存率が計算される。後者では新たにイ

第12章 メタアナリシス

1 メタアナリシスの研究計画

システマティックレビュー (systematic review) は臨床上の疑問について、明確に定義された手法に従って収集された文献情報に基づいて、批判的吟味を行いながら検討する手法であり、必ずしも統計学的な解析は行われず (34頁を参照)。一方、**メタアナリシス**はシステマティックレビューの1つ (統計学的統合を行うシステマティックレビュー) として位置づけられ、やはり統計学的解析に入る前に統合する研究材料をしっかりと選別することが重要である (Garbage in, garbage out)。

メタアナリシスの実施時にも、他の臨床研究と同様に、最初に Research question を明確にする。そして、文献検索にもちいるデータベース、検索キーワードを設定し、また、どのような研究を選択するかの適格・除外条件 (学会抄録を含めるか? 英語以外の論文を含めるか? 無作為化比較試験以外を含めるか? など) についても事前に定めておく必要がある。さらに個々の論文の背景や考察などに紹介されている研究を適宜手作業で追加する。同一の研究が2報以上に分けて発表されていることがあるので重複に注意する。次に、選定された研究の質 (バイアスのリスク等) を評価する。無作為化や割り付けの方法、盲検化の有無や方法、脱落症例の記述などに基づいてスコアリングがなされる。そして、各研究から、主観的な解釈を避けるようにデータを抽出する。文献検索、文献選定、文献評価、データ抽出は2名の研究者が独立して行うことが望ましい。メタアナリシスの研究結果は **PRISMA (Preferred reporting items for systematic review and meta-analysis)** に沿って発表することが求められているが、研究開始前に PRISMA での発表を意識しながら研究計画をたてることによって、適切な研究が実施できるようになる。

2 固定効果モデルとランダム効果モデル

複数の研究結果を統合する場合に最も簡単な方法は研究結果を全てプールして1つの研究にまとめてしまうことであるが、この方法は正しくなく、**シンプソンパラドックス**などの問題が生じる。シンプソンパラドックスの一例としてしばしば紹介される Schwarz らのデータを紹介する。ある大学が2つの教育プログラムへの入学試験を行い、その合格率に男女差があるかどうかを検討した。プログラム A では男性は600人中480人 (80%) が合格、女性は200人中180

1 サマリー表の出力

① 解析サンプルの背景表の作成

群別する変数(0~1つ選択)

Age
Age70
Anthracyclines
Blast80
BM_blast
Cre15
CRP
CRP1
Days
Hb

群別化する変数を指定する。

複数の選択は Ctrl キーを押しながらクリックする。

正規分布する連続変数は平均値と標準偏差で、正規分布しない連続変数はオプションで指定した方法で要約される。

↓ 複数の選択はCtrlキーを押しながらクリック。

カテゴリ変数(名義変数、順序変数)

Karyotype
LDH
LDH250
Plt
Plt2
PS
PS2
Sex
Survival
WBC

連続変数(正規分布)

Age
Age70
Anthracyclines
Blast80
BM_blast
Cre15

連続変数(非正規分布)

BM_blast
Cre15
CRP
CRP1
Days
Hb.

自動選択にすると2群に分ける場合は Fisher の正確検定、3群以上の場合はカイ 2 乗検定が選択される。

カテゴリ変数の検定方法

独立性のカイ2乗検定(連続補正有り) 最小値と最大値 No

フィッシャーの正確検定 四分位数範囲(Q1-Q3) Yes

自動選択

出力先 表示言語

クリップボード 英語

CSVファイル 日本語

クリップボードへの出力はWindowsのみで可能

↓ 一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用