

# 論文化のための統計手法選択について

榎本 哲郎<sup>†</sup>第70回国立病院総合医学会  
(平成28年11月11日 於 沖縄)

IRYO Vol. 72 No. 8 / 9 (355-358) 2018

## 要旨

私は精神科医であり、統計の専門家ではない。20年ほど前に原著論文を書いたときには、統計学の専門家や先輩医師の指示どおりにデータを解析した。自分で悩んで選択した統計検定手法ではなかったために身につかなかった。平成22年2月に「医療」の編集委員となり、査読の際に統計検定手法に注目して投稿論文を読む必要が生じた。また、平成24年に精神科の学会誌に論文を投稿した際、査読者から統計検定手法の選択について間違いを指摘され、手探りで統計の勉強をした。私の統計に関する知識は、書籍とインターネットを利用した独学である。これまでに私が学んだ統計の書籍と統計ソフトを紹介し、次いで正しい統計検定手法選択について述べ、最後に普段論文の査読をしていて気になっている点を報告した。車のエンジンが動く仕組みを知らなくても、ルールを守れば安全に車の運転ができる。それと同様に、統計検定での数式を知らなくても、適切な統計検定手法を選択すれば、学術論文の完成という目的地に到着できるのである。

キーワード 統計検定手法, 査読者, 学術論文

## はじめに

私は精神科医であり、統計の専門家ではない。20年ほど前に原著論文を書いたときには、統計学の専門家や先輩医師の指示どおりにデータを解析した。自分で悩んで選択した統計検定手法ではなかったために、統計に関する知識は身につかなかった。その後は、統計とは縁のない生活をおくっていた。ところが、平成22年2月に「医療」の編集委員となり、

査読の際に統計検定手法に注目して投稿論文を読む必要が生じた。また、平成24年に精神科の学会誌に論文<sup>1)</sup>を投稿した際、査読者から統計検定手法の選択について間違いを指摘され、手探りで統計の勉強をした。私の統計に関する知識は、本とインターネットを利用した独学である。こうして統計への理解を少し深めることができ、ある程度使いこなせるようになった。

今回のシンポジウムで「論文化のための統計手法

国立国際医療研究センター国府台病院 精神科 <sup>†</sup>医師

著者連絡先：榎本哲郎 国立国際医療研究センター国府台病院 精神科 〒272-8516 千葉県市川市国府台1-7-1

e-mail: domani-e@hospk.ncgm.go.jp

(平成29年7月18日受付, 平成30年6月8日受理)

How to Choose the Statistical Analysis Appropriate to Your Medical Research

Enomoto Tetsuro, Kohnodai Hospital, National Center for Global Health and Medicine

(Received Jul. 18, 2017, Accepted Jun. 8, 2018)

Key Words: statistical analysis, reviewer, academic paper

選択について」講演する機会を得た。統計について論じられるほどの力量はないが、これまでに私が学んだ統計の書籍と統計ソフトを紹介し、次いで正しい統計検定手法選択について述べ、最後に普段論文の査読をしていて気になっている点を示す。

---

## 統計の書籍と統計ソフト

---

自分の臨床研究をまとめて学会で発表したり、論文文化したりする際には、何らかの統計検定手法を使用する必要がある。一方、他者による学会発表や論文を正しく理解するためにも統計学的な知識は必要である。それにもかかわらず統計の体系的な教育は十分ではない。統計の書籍も多数出版されているが、実際に使えるものは数少ない。そこで個人的に推薦できる書籍を紹介する。また、個人で使用しやすい統計ソフトもあわせて紹介する。

初めて統計を勉強するには、浅井隆著「いまさら誰にも聞けない医学統計の基礎のキソ 第1巻」<sup>2)</sup>、「同 第2巻」<sup>3)</sup>がわかりやすく、初心者向けである。次に初級者から中級者には、新谷歩著「今日から使える医療統計」<sup>4)</sup>がおすすめである。

私が使用している柳井久江著「4 Steps エクセル統計第4版」<sup>5)</sup>は、自分のデータ処理を取りあえず何とかしたい時に役に立つ。しかも安価である。これには付録CD-ROMにExcel<sup>®</sup>上で動くアドインソフトStatcel 4<sup>®</sup>が入っており、パソコンにインストールすればExcel<sup>®</sup>上でさまざまな統計処理ができる。例題が多数あり、ばらばらと本を繰るだけで、自分のデータにどの統計検定手法を選択すればよいのかわかってくる。また、得られた統計検定の結果をどう読み取るのかについても丁寧な解説がある。データの入力仕方と検定結果の読み取り方も学習できる優れたものである。このソフトを使って具体的な研究データを解析してみると、少しずつ統計検定手法の本質を理解できるようになる。

Statcel 4<sup>®</sup>では物足りなくなった方には、無料統計ソフトEZR (Easy R)を推薦する。EZRはwww.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.htmlからダウンロードして使用できる。神田善伸著「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2版」<sup>6)</sup>には、CD-ROMが付属しているのでそれをインストールしてもよい。EZRの初心者向けマニュアル<sup>7)</sup>も出版されている。また、EZRの使用マニュアルとして新谷歩著「み

んなの医療統計 12日間で基礎理論とEZRを完全マスター！」<sup>8)</sup>と「みんなの医療統計 多変量解析編 10日間で基礎理論とEZRを完全マスター！」<sup>9)</sup>もお薦めである。EZRを使用した論文を発表する場合には、Bone Marrow Transplantation<sup>10)</sup>を参考文献として引用する。統計解析の上級者は、統計ソフトとしてIBM<sup>®</sup>社SPSS<sup>®</sup>を使いこなしているが、これは高価であるため個人購入には適さない。

---

## 正しい統計検定手法選択

---

正しい統計検定手法選択のためには以下の7つの質問に順に答えるだけでよい。Q1：ランダム化研究か否か？ Q2：差をみるのか、相関をみるのか？ Q3：比較するデータ間の対応の有無は？ Q4：連続変数アウトカムの種類は2値変数？ 順序のあるカテゴリー変数？ 順序のないカテゴリー変数（名義変数）？ 連続変数？ 2値打ち切り？ Q5：アウトカムの正規性はパラメトリック？ ノンパラメトリック？ Q6：比較群の数は2？ 3以上？ Q7：症例数は？

比較群の背景がそろったランダム化研究などでは単変量解析を使用できるが、ランダム化されていない背景のそろっていない群間比較には重回帰分析を使用する<sup>8)</sup>。

がんなどの疾病の発症や死亡などのリスクに及ぼすさまざまな予後因子の影響を検討する場合には多変量解析が用いられる。臨床研究でよく使用される多変量解析には重回帰分析、ロジスティック回帰分析、Cox回帰分析がある。観察研究では、倫理的に無作為割り付けを行うことが難しい場合が多い。そのため、年齢、性別、喫煙、飲酒などの予後因子が疾病の発症や死亡のリスクに及ぼす影響を調べる際に、他の因子の分布に偏りがある場合が想定され、それらの因子が交絡因子として働く可能性がある。そこで、その影響を除いてリスクを調べる必要があり、多変量解析が役に立つ<sup>6)</sup>。

---

## 査読者の観点から

---

### 1. 有効数字

数値を定量的に扱う際には、有効数字を常に意識しなければならない。これは、新潟大学工学部木村勇雄教授の「有効数字の簡便な扱い」<http://www.gs.niigata-u.ac.jp>が参考になる。数値の桁数が多い

ときは丸め処理 (rounding) する。丸め処理には、切り捨て、切り上げ、四捨五入がある。ところで、小学校で教える四捨五入では実際には少しだけ大きめに丸めることになってしまう。そこで JIS (日本工業規格) 式や ISO (国際標準化機構) 式四捨五入を行う。それは「5」の扱い方が小学生式と異なる。つまり、ひとつ上の位が奇数の場合は「5」を切り上げ、ひとつ上の位が偶数の場合は「5」を切り捨てるのである。具体的には、2.4502を有効数字2桁で丸めると2.4になり、0.1750を有効数字2桁で丸めると0.18になる。

## 2. 標準偏差と標準誤差

臨床研究で集めたデータ処理に関して、査読者がどう考えているのかについては、森本剛著「査読者が教える採用される医学論文の書き方」<sup>11)</sup>が参考になる。臨床研究では、ばらつきは基本的に標準偏差 (standard deviation: SD) で表す。測定された平均値のばらつきを示すには標準誤差 (standard error: SE) を使うが、臨床研究では対象集団のばらつきを示すために SD を用いる。臨床研究で集めた、年齢や体重などの数字で表される変数は、平均値と SD で表現されることが多い。

## 3. パラメトリックとノンパラメトリック、平均値と中央値

平均値はデータが正規分布のように左右対称な場合に中心の値と一致するので“分布を想定する”パラメトリックな統計量である。一方、入院日数、症状評価尺度のスコア、アンケートでのスケール (1: ~ない, 2: 少し~ない, 3: どちらでもない, 4: 少し~ある, 5: ~ある) などは、変数によっては中央値 median (四分値, 四分位範囲) を使うことを考慮する。中央値はデータの分布によらずいつでも用いることができるノンパラメトリックな (分布を想定しない) 統計量である<sup>11)</sup>。

## 4. 2値変数

2値変数はどちらか一方を示せばよい。たとえば性別であれば、男性106人 53%、女性94人 47%と両方書わずに、どちらか一方で十分である<sup>11)</sup>。もちろん、男女を併記して正確を期すようにすることもある。

## 5. 百分率 (%)

集めたデータを百分率 (%) で示すこともよく行われる。小数点以下の%は多くの場合、不要である。論文の読者 (査読者も含む) にとって、%で表示する際の小数点以下の数値は重要ではない。分母と分子に相当する観察数を記載しておけばよい。必要があればその数値から計算できるのであるから、1%以下の場合を除くほとんどの場合、小数点以下の数字は必要がないのである。%の計算の対象となる症例数 (分母) が20以下なら%は無意味である。分数 (分子/分母) で記載した方が誤解は生じにくい<sup>11)</sup>。

## 6. P値

P値とは、実際には治療薬 A に効果が全くないにもかかわらず、まるで効果があるかのような結果になってしまう間違いの確率のことである。治療薬 A に全く効果がないならば、治療薬 A の投与群と非投与群の差はゼロになるはずである。臨床研究では、サンプルはランダムに集めるからたまたま大きな差が観察される可能性がある。ミラクルがおこる確率が5%未満ならば、違いのある科学的なエビデンスとして認めてよいという慣習に従い、多くの研究では「 $P < 0.05$ 」で有意差を判定している。通常  $P < 0.05$  ならば治療薬 A は効く、 $P \geq 0.05$  ならば治療薬 A は効かないと判定される。しかし、P値は症例数次第で変わるものである。症例数が多ければ臨床的に無意味な差でも有意差ありとなり、一方、症例数が少なければ、臨床的に意味のある差でも有意差がでないこともある。P値が0.05以上、すなわち「差がない」というわけではない<sup>4)</sup>。有意差が認められなかった場合に、P値の代わりに NS (not significant 有意差なし) と表現せずに、統計処理の結果を  $P = 0.08$  などと明示すべきである。P値よりも、リスク比と95%信頼区間やオッズ比と95%信頼区間を用いることが有用なことも多い<sup>11)</sup>。

## 7. $\chi^2$ 検定とフィッシャーの正確検定

死亡・生存など2値変数のアウトカムを2群間で比較する  $2 \times 2$  分割表の検定では、症例数の総数が40人を超える場合は、 $\chi^2$  検定を用いる。20人未満の場合はフィッシャーの正確検定を使う。20人以上40人未満の場合は、分割表の1つでも5人に満たない場合は、フィッシャーの正確検定を使う<sup>8)</sup>。



## 8. 多重比較の問題

A 群, B 群, C 群の3群で平均値を比較する場合, A 対 B, B 対 C, C 対 A の3回 t 検定を行うと, 実際には差がないにもかかわらず, いずれか1つの検定で偶然に  $P < 0.05$  となる確率は  $1 - (1 - 0.05)^3 = 0.14$  になる. もし, 13回の検定を行うとすると, いずれか1つの検定で偶然に  $P < 0.05$  となる確率は50%に近づく. これを多重比較の問題という<sup>6)</sup>. そのため, 3群以上の多群の差の検定には, まず分散分析をして多重比較が可能か否か検討した上で, Tukey-Kramer 法や Bonferroni 法などの多重比較検定を行う<sup>5)</sup>.

## 9. 相関係数と P 値

相関関係の強弱を示すには, 相関係数の値を示した方がよい. 相関係数の検定において, P 値がきわめて小さいからといって, 強い相関があることを意味するわけではない. 「相関がある」といえるのは相関係数が  $\pm 0.4 \sim \pm 0.7$  の場合である. データ数が100あれば, 相関係数が0.1965でも両側確率の P 値は0.049902となるのである. したがって, 相関係数の有意性を示す P 値に対しては, データ数と相関係数を表記しておく必要がある<sup>5)</sup>.

---

## おわりに

車のエンジンが動く仕組みを知らなくても, ルールを守れば安全に車の運転ができる. それと同様に, 統計検定での数式を知らなくても, 適切な統計検定手法を選択すれば, 学術論文完成という目的地に到着できるのである. この報告が, これから研究内容を論文として完成させたいと願う方々が統計検定手法を選択する際の参考になれば幸いである.

〈本論文は第70回国立病院総合医学会シンポジウム「医療論文を完成させよう - 論文化の意義と押さえておくべきこと -」において「論文化のための統計手法選択について」として発表した内容に加筆したものである.〉

**著者の利益相反:** 本論文発表内容に関連し, 開示すべき利益相反はない.

---

## [文献]

- 1) 榎本哲郎, 伊藤寿彦, 関根慶輔ほか. 治療抵抗性統合失調症55例に対する clozapine 使用経験. 精神経誌 2013; 115: 953-66.
- 2) 浅井 隆. いまさら誰にも聞けない医学統計の基礎のキソ1: まずは統計アレルギーを克服しよう!. 東京: アトムス; 2010.
- 3) 浅井 隆. いまさら誰にも聞けない医学統計の基礎のキソ2: 結果の解釈ができるようになろう!. 東京: アトムス; 2010.
- 4) 新谷 歩. 今日から使える医療統計. 東京: 医学書院; 2015.
- 5) 柳井久江. 4 Steps エクセル統計. 第4版. 東京: 星雲社; 2015.
- 6) 神田善伸. 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2版. 東京: 中外医学社. 2015.
- 7) 神田善伸. 初心者でもすぐにできるフリー統計ソフト EZR (Easy R) で誰でも簡単統計解析. 東京: 南江堂. 2014.
- 8) 新谷 歩. みんなの医療統計 12日間で基礎理論と EZR を完全マスター!. 東京: 講談社. 2016.
- 9) 新谷 歩. みんなの医療統計 多変量解析編 10日間で基礎理論と EZR を完全マスター!. 東京: 講談社, 2017.
- 10) Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013; 48: 452-8.
- 11) 森本 剛. 査読者が教える採用される医学論文の書き方. 東京: 中山書店; 2013.