

〈報告〉

健康医学における多変量・質的データに関する
非母数モデルの体系化

土屋 基*・鈴木 勝彦**・大森大二郎***

A Systemization of the Nonparametric Model concerning Multivariate
and Qualitative Data in Health and Medical Sciences

Motoi TSUCHIYA*, Katsuhiko SUZUKI** and Daijiro OHMORI***

1. はじめに

健康・医学領域で必要な数理統計的な手法は、健康・医学領域以外のそれとは様相をかなり異にする。通常広く用いられている統計検定は、ある特定の分布型を持つ母集団を想定し、そこからの標本を用い、その母集団を特徴付ける母平均、母分散、母比率などに関して検定するパラメトリック検定法 (parametric method) が行なわれている。

ところが健康・医学領域においては、多数の標本を扱う疫学調査や社会調査を除いて、極めて少数例の標本を用いて検討せざるを得ない症例研究や実験研究が行われている。また健康・医学領域では、量的データのみならず名義尺度や順序尺度などに基づく質的データが扱われることが少なくない。

小標本を扱う場合、例えば母平均の差の検定に際して行われるパラメトリック検定法のひとつである t 検定は、標本が正規分布から無作為に抽出

されたものでないかぎり適用することが出来ない。また、質的データの場合にも分布の形を特定することができないため、パラメトリック検定法は使用できない。

しかし、健康や医学の領域においては、サイズの小さい質的データの統計解析をしなければならぬ局面が非常に多く、パラメトリック検定法とは異なる非母数 (ノンまたはセミ) パラメトリック検定法の援用が求められる。

近年、高性能なパーソナルコンピュータの開発普及に伴い、統計ソフトウェアが充実し、標本の特徴を意識しなくとも、とも角データをコンピュータに入力しさえすれば極めて短時間の内に次々に結果が出力される便利な世の中になった。その反面、理論的な背景はブラックボックス化し、手法の適用や結果の解釈を間違えている例が散見されるようになった。

2. 研究目的

本研究においては、健康・医学の領域で用いられる非母数 (ノンまたはセミ) パラメトリック手法にはどのようなものがあるのか、正しい情報を得るためにはどのような方法を用いるべきか、またどのような統計数学理論に基づいているのかについて調査し、整理体系化を図り正しく活用されるための資料を提示することを目的とした。

* 健康管理学研究室
Seminar of Health Care and Administration

** スポーツ栄養学研究室
Seminar of Sport Nutrition

*** 医学部化学研究室
Department of Chemistry

3. 研究方法

ノンパラメトリック検定法に関する文献調査を行い、その長所、短所を拾い出した。その上で各種検定法の①目的、②適用時の制約、③計算方法、④実際のデータ例による計算を行った。さらに、データの形式別・種類別のノンパラメトリック検定方法を一覧表に整理した。

4. 結果報告

4.1 ノンパラメトリック検定法の特徴

ノンパラメトリック検定法を適用する場合に、この検定方法の特徴をよく理解して使っていくことが大切であり、この検定方法の長所短所をまとめると次のようなことがあげられる。

- (1) 長所
- ① 母集団の分布の型にとらわれないために前提条件がかなり緩和される。
 - ② 小標本でも適用可能である。
 - ③ 名義尺度や順序尺度に基づく質的データの検定に活用できる。
 - ④ 検出力は落ちるが、パラメトリック手法が適用できる場合には、常にノンパラメトリック検定法を使うことが出来る。
 - ⑤ 従って、ノンパラメトリック検定法で有意であれば、かなり信憑性が高い。
 - ⑥ サンプルサイズがある程度大きいならば、パラメトリック検定で用いられる数表（例えば標準正規分布表）がそのまま使える。
- (2) 短所
- ① パラメトリック検定法と比較して若干検出力が落ちる。
 - ② 標本サイズが大きくなると計算が煩雑になる。
 - ③ 多くの場合ノンパラメトリック検定手法特有の数表を必要とする。

4.2 一覧表の作成

ノンパラメトリック検定法²⁾⁻¹⁰⁾を整理し、結果を表1のようにまとめた。

4.3 ノンパラメトリック検定法の一例

ノンパラメトリック検定法の種類は数多く、こ

こで全ての手法を取り上げることが出来ない。そこで、一例として成書にはあまり紹介されていないアンサリー・ブラッドレー (Ansari-Bradley) の検定¹⁾を取り上げる。この検定の目的は、2群の母集団のばらつきに違いがあるかどうかを調べるもので、パラメトリックな場合の等分散性の検定に相当する。ただし、2群の位置パラメーターは同じであると仮定できるものとする。

データ

群 A1 は $x_{11}, x_{12}, \dots, x_{1n_1}$, 群 A2 は $x_{21}, x_{22}, \dots, x_{2n_2}$ とする。

解析

- ① 2群のデータを一つにまとめ、小さい方から大きいほうへ順に並べかえる。
- ② 小さい方から1番目と大きいほうから1番目とに順位“1”をつける。
小さい方から2番目と大きいほうから2番目とに順位“2”をつける。
以下同様の作業を続ける。
なお、タイがあるときは中間順位を設定する。
- ③ $n_1 \leq n_2$ ならば群 A1 に、 $n_1 > n_2$ ならば群 A2 につけられた順位の総和 r を求める。
- ④ n_1, n_2 のサイズによって次の2つに分れる。
 - (1) n_1, n_2 が10未満のとき
 - アンサリー・ブラッドレー検定専用の数表から2つの臨界値 $c_1, c_2 (c_1 < c_2)$ を探す。
 - $r < c_1$ または $r > c_2$ ならば、ばらつきに差があるとの結論を出す。
 - (2) n_1, n_2 がともに10以上のとき
 - $n_1 + n_2 (=m)$ とおく) が偶数ならば

$$z_0 = \frac{\left| r - \frac{n_1(m+2)}{4} \right| - \frac{1}{2}}{\sqrt{\frac{n_1 n_2 (m^2 - 4)}{48(m-1)}}}$$

- $n_1 + n_2 = m$ が奇数ならば

表1 ノンパラメトリック統計法一覧

データ形式	測定尺度	検定方法	帰無仮説	
1組の標本	名義尺度	二項検定 χ^2 適合度検定	母比率は想定した一定値に等しい 観測度数は想定した確率分布から求めた理論度数に等しい	
	順序尺度	Kolmogorov-Smirnov の検定 Runs Test (連の検定)	経験分布関数が確率分布の分布関数に等しい 標本系列は観測順序と無関係にランダムである	
2組の標本	対応のある場合	名義尺度	McNemar の検定	2 標本は同じ一つの母集団から抽出されたものである
		順序尺度	Sign Test (符号検定) Wilcoxon の符号付順位和検定	2 標本の中央値は等しい 2 標本の中心位置は等しい 2 通りの処理の結果は等しい
		間隔尺度	確率化検定 Lepage の検定	2 標本の母分散と中央位置は等しい 2 標本の母分散と中心位置は等しい
	対応のない場合	名義尺度	Fisher 直接確率検定 χ^2 検定	期待度数が 5 以下のセルがある場合で、2 因子は独立である 分割表の各セルについて、観測度数は理論度数と等しい 2 因子は独立である
		順序尺度	Wilcoxon の T テスト 中央値検定 Mann-Whitney の U 検定 Kolmogorov-Smirnov の検定 Wald-Wolfowitz の連の検定 Moses Test	2 標本の中心位置は同じである 2 標本の中央値は等しい 2 標本の中心位置は同じである 2 標本の経験分布関数は等しい 2 標本の中心位置は同じである 2 標本の母分散と中央値はそれぞれ等しい ある課題達成の前後の有意差はない
		間隔尺度	Lepage の検定 確率化検定 Ansari-Bradley の検定	2 標本の母分散と中心位置はそれぞれ等しい 2 標本の母分散と中心位置はそれぞれ等しい 2 標本の母分散は等しい
3組以上の標本	対応のある場合	名義尺度	Cochran の Q 検定	3 つ以上の理論度数や母比率は等しい
		順序尺度	Friedman の順位検定 (二元配置) 中央値検定	3 つ以上の標本は同じ一つの母集団から抽出されたものである 3 つ以上の標本の母中央値は等しい
	対応のない場合	名義尺度	χ^2 検定	$k \times 1$ 分割表における因子は独立である
		順序尺度	Kruskal-Wallis の順位検定 (一元配置) 拡張型中央値検定	3 つ以上の標本の母集団の中心位置は同じである 3 つ以上の標本の母集団の中心位置は同じである
		間隔尺度	Bartlett の検定	3 群以上の母分散は等しい
	相関関係	名義尺度	クラメールの C 係数 (関連係数)	2 つの要因間の関係の強さ
順序尺度		Spearman の順位相関係数 Kendall の順位相関係数 Kendall の一致度係数	2 変量に順位の相関はない 2 変量に順位の相関はない k 組の変量間に関連はない	

$$z_0 = \frac{\left| r - \frac{n_1(m+1)^2}{4m} \right| - \frac{1}{2}}{\sqrt{\frac{n_1 n_2 (m+1)(m^2+3)}{48m^2}}}$$

を求める.

- $z_0 > 1.96$ ($z_0 > 2.58$) ならば, 有意水準 5% (1%) で 2 つの母集団のばらつきに違いがあるとの結論を出す.

4.4 データの一例と解析

A1 学科および A2 学科の学生の, ある運動生理学的スコアが次のとおりだったとする. なお, A1 学科生の点数には下線を引いてある.

A1	...	<u>46</u>	<u>65</u>	<u>34</u>	<u>59</u>	<u>74</u>	<u>52</u>	<u>55</u>	<u>41</u>	<u>70</u>
										<u>38</u>
A2	...	39	56	71	60	83	44	37	50	42

このデータに, アンサリー・ブラッドレーの検定を行う. 手順は次のとおりとなる.

- ① 22個のデータをひとまとめにして小さい方から書並べる.
- ② 小さい方から1番目と大きい方から1番目とに順位1をつけ, 小さい方から2番目と大きい方から2番目とに順位2をつけ, ...という作業を続けると, 次の表ができる.

点数	<u>34</u>	37	<u>38</u>	39	<u>41</u>	42	44	45	<u>46</u>	50	51	<u>52</u>	<u>55</u>
順位	①	2	③	4	⑤	6	7	8	⑨	10	11	⑪	⑩

点数	56	<u>59</u>	60	64	<u>65</u>	<u>70</u>	71	<u>74</u>	83
順位	9	⑧	7	6	⑤	④	3	②	1

- ③ A1 の標本サイズは $n_1 = 10$, A2 のそれは $n_2 = 12$ であるから A1 につけられた順位 (上の表中○で囲んだもの) の総和 r を求める.
 $r = 1 + 3 + 5 + 9 + 11 + 10 + 8 + 5 + 4 + 2 = 58$
- ④ n_1, n_2 がともに10以上であることと, $m = n_1 + n_2 = 22$ (偶数) に注意して

$$z_0 = \frac{\left| 58 - \frac{10(22+2)}{4} \right| - \frac{1}{2}}{\sqrt{\frac{10 \times 12(22^2 - 4)}{48(22-1)}}}$$

$$= \frac{|58 - 60| - \frac{1}{2}}{\sqrt{\frac{400}{7}}} = 0.198 (< 1.96)$$

- ⑤ 有意水準 5% で 2 つの母集団のばらつきに違いがあるとはいえない.

5. 結 論

ノンパラメトリック検定法は, 小標本や質的データに対処でき, 特定の確率分布を仮定しなくても済む検定法である. 統計ソフトウェアも数多く散見されるが, 検定法の特徴をよく理解し, データの形式や種類などの確に判断して用いなければ誤った結果や解釈に陥る危険性が高い. 本研究によって正しく検定を進めるための手法を整理し, 体系化を図るとの目的を達することができた.

謝辞

本研究の一部は, 平成13年度順天堂大学学内奨励研究費の助成を受けて行われた.

文 献

- 1) Ansari, A. R, and Bradley, R. A. (1960) *Rank-sum tests for dispersions*. Ann, Math. Stat. 31, 1174-1189.
- 2) 市原清志 (1993) バイオサイエンスの統計学. 第1版, 東京, 南江堂
- 3) 石村貞夫 (2000) すぐわかる統計処理. 第1版, 東京, 東京図書
- 4) 石村貞夫 (2001) SPSSによる統計処理の手順. 第3版, 東京, 東京図書
- 5) 岩原信九郎 (1979) 教育と心理のための推計学. 第23版, 東京, 日本文化科学社
- 6) 中野正孝 (1989) 看護系の統計調査入門. 第1版, 東京, 真興交易医書
- 7) 野田一雄, 三野大来 (1997) やってみよう統計. 第1版, 東京, 共立出版

- 8) 高木廣文(1998)ナースのための統計学. 第1版, 東京, 医学書院
- 9) 脇本和昌, 垂水共之, 田中 豊 (1985) パソコン統計解析ハンドブック I 基礎統計編. 第1版, 東京, 共立出版
- 10) 山崎信也 (2000) なるほど統計学とおどろき Excel統計処理. 奥秋 晟監修, 第2版, 東京, 医学図書出版
- (平成15年11月10日 受付)
(平成16年1月20日 受理)