

平成 30 年 6 月 21 日現在

機関番号：13101

研究種目：基盤研究(C) (一般)

研究期間：2013～2017

課題番号：25380873

研究課題名(和文) 心理学研究における統計的検定手法の見直しと改善

研究課題名(英文) Reconsidering and remediation of statistical testing procedures in psychological research

研究代表者

杉澤 武俊 (Sugisawa, Taketoshi)

新潟大学・人文社会・教育科学系・准教授

研究者番号：30361603

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：心理学の研究において広く利用されている平均値差の検定と質的変数の連関の検定について国内主要学会誌における実態を調査し、方法論的な問題およびそれらを改善するための代替法の検討を行った。前提条件の確認や、有意な効果に対する事後検定の手法選択などにおいて方法論的な問題があるが、他の既存の方法によりそれらを改善することが可能であることが示唆された。また、より望ましい手法を実行するためのRによる関数を作成した。

研究成果の概要(英文)：Statistical testing procedures for both difference of means and contingency tables in psychological research were examined. It was shown that some conventional preliminary and post hoc tests procedures have methodological problems, and alternative procedures were suggested for these purpose. In addition, some functions of R scripts were developed.

研究分野：心理統計学

キーワード：平均値差の検定 連関の検定 事後検定

1. 研究開始当初の背景

心理学分野の学術研究における実証の方法として、統計的データ解析手法による数量的な分析、とりわけ統計的検定は利用頻度が高いものである。統計的手法の方法論に関する研究が発展するに伴い、特に多変量解析を中心に心理学研究における統計的手法の実践にも新しい手法が取り入れられ、ここ近年の間に大きな進歩を見せている。その一方で、統計的検定についても、従来用いられている手法の問題点の指摘や、新しい方法の提案などがなされているにもかかわらず、それらの知見は心理学研究の実践においてほとんど反映されておらず、伝統的・慣習的に用いられてきた手法や手順がルーチンとして選択されたままであるのが現状であると思われる。

このような現状の背景として以下の要因が考えられる。

- 従来の方法のデメリットや問題点、あるいは新しい方法のメリットや存在そのものが一般的な統計ユーザーの間で認識されていない。
- 多くの研究者がデータ解析に利用する主要なソフトウェアで実行できない、もしくは対応していたとしてもデフォルトでは従来の方法を適用するようになっている。
- 一般的な統計学のテキストの解説や公刊された研究論文中で用いられている手法が旧来の知見に基づいており、それらを参考として古い慣習が再生産される。

近年では特に、統計的検定に過度に依存したデータ解析を行うことそのものが見直しをするべきルーチン化された実践であるということもできるが、検定そのものへの批判は古くからなされながらもここまで検定が普及した現状や、すべての研究目的に対して検定に替わる他の望ましい方法が必ずしも整備されているわけではないことを考えると、統計的検定をすぐに廃止するというのは現実的ではない。したがって、現実的に実行可能なより望ましい検定の実践があり得るなら、まずはそこから改善していくことで、心理学研究全体の向上に資するものになると期待される。

2. 研究の目的

本研究では、実際の心理学研究の場面に於いて慣習的に行われている検定の方法を見直し、上記の問題点を可能な限り克服して、ユーザーがより良い検定の実践ができるような手法の提案、指針の呈示を行うことを目的とする。具体的には、検定の中でも心理学の研究で頻りに利用される、平均値の群（条件）間差に関する検定と、質的変数の連関（比率の差）に関する検定の2つに焦点を当て、以下のことを行う。

- (1) 実際に行われた心理学領域の研究にお

いて、どのような検定の手法や手順が選択されているのか、現状の実態を明らかにする。

(2) 通常使われる検定手法や手順について、方法論的な問題点があればそれを明らかにし、それを改善するための代替法を提案する。

(3) 改善された方法を、実際にコンピュータを使って実行できるようなプログラムを開発する。

3. 研究の方法

(1) 統計的検定利用の実態調査

心理学領域における国内の主要な学会誌として、2013年（年度）発行の『心理学研究』（第84巻）、『教育心理学研究』（第61巻）、『発達心理学研究』（第24巻）を選定することとし、それらに掲載された計144編の論文を対象に、論文の中で報告されている検定のうち、量的変数の平均（もしくは分布の位置）の群間差に関する仮説を検定する手法、および、質的変数の連関（もしくは比率の差）に関する仮説を検定するための手法、および、その検定にかかわる、前提条件のチェックなどの予備検定から、主検定で有意な効果が見られた場合に行われる事後検定などの一連の手続きについて調べる。

(2) 検定手法や手順の見直し

(1)の実態調査でみられた検定手法や手順を中心に、方法論的な問題点がないか、各手法の原理等から理論的に、もしくは、人工的に発生させた乱数データを用いたシミュレーションにより検討を行う。方法論的に問題があると考えられる手法・手順について、可能な改善策について検討を行う。

これらの検討の際に、適切な検定手法・手順であるかを判断するための観点・基準としては、

- ① 検定される仮説はリサーチクエスチョンを正しく反映しているか。
 - ② 第1種の誤りの確率が設定された有意水準を超えないように統制され、かつ、検定力が可能な限り高められているか。
 - ③ 母集団分布の仮定など、その手法が適用可能な前提条件を満たしたデータを扱っているか。
- という3点が挙げられる。

(3) 分析用プログラムの開発

(2)で提案されたより適切と考えられる検定手法について、オープンソースの統計解析環境Rで利用可能な既存の関数等がないものについて、新たに開発を行う。

4. 研究成果

(1) 利用される統計的検定手法の現状

対象となった論文において、量的変数の平均の群間差に関する検定、およびそれに準じるノンパラメトリック検定で、実際に用いられていた具体的手法と用いられた論文数について表1に示した。同様に、質的変数の

連関や比率の差に関する検定についての結果を表2に示した。対象となった論文が単年度の限定されたものであるため、偏った結果となった可能性もあるが、一部例外はあるものの、概ね慣習的に用いられていると考えられる手法が多く利用されているといえる。

表1 平均(分布の位置)の群間差に関する検定で用いられた手法

検定手法	論文数
(予備検定・前提条件の確認)	
Kolmogorov-Smirnov 検定	1
Shapiro-Wilk 検定	1
Smirnov-Grubbs 検定	1
F検定(等分散性)	1
Bartlett 検定	1
Fligner-Killeen 検定	1
<i>t</i> 検定	28
Welch 検定と通常の <i>t</i> 検定併用	3
分散分析	
共分散分析	2
一般線型モデル	2
一般化線型混合モデル	1
(主効果の事後検定・多重比較)	
Bonferroni 法	17
Tukey 法	9
Ryan 法	7
Holm 法	2
Shaffer 法	2
最小有意差(LSD) 法	1
Scheffe 法	1
不明	2
(交互作用の事後検定)	
単純効果の検定	32
Kruskal-Wallis 検定	4
Mann-Whitney の <i>U</i> 検定	3
Friedman 検定	2
Wilcoxon の符号付順位検定	1

表2 連関・比率の差に関する検定で用いられた手法

検定手法	論文数
カイ2乗検定	17
(事後検定)	
残差分析	5
Bonferroni 法による多重比較	1
Ryan 法による多重比較	1
Fisher の直接確率法	13
(事後検定)	
残差分析	2
逆正弦変換による分散分析	3
McNemar 検定	2
McNemar-Bowker 検定	1
Kolmogorov-Smirnov 検定	1

(2) 各種手法・手続きの見直し

①ノンパラメトリック検定は研究目的に沿っているか

複数の群間で平均値の比較を行う *t* 検定や分散分析では、分散の等しい正規分布に従う母集団から単純無作為抽出によりデータ収集が行われることが前提となっており、その前提が満たされないときに、第1種の誤りの確率を理論値通りにコントロールできない。これらの前提のうち、正規分布の仮定が満たされない場合に、その仮定が不要なノンパラメトリック検定の利用が推奨されることがあり、表1でも4種のノンパラメトリック検定のべ10件みられる。これらのノンパラメトリック検定は、基本的に標本内のデータの順位に基づいたもので、平均や中央値などの統計量は直接扱ってはいないため、本来、平均や中央値などの分布の中心位置を議論したい場合に、適切な知見が得られるか疑問が残る。そこで、人工データによるシミュレーションで、ノンパラメトリック検定の結果が何を示しているのかを検討した。

2群比較の場合、平均値に関する *t* 検定の代替法として Mann-Whitney の *U* 検定が利用される。そこで、様々な条件下で発生させた同一のデータに対してこの2つの検定を適用した結果を比較した。例えば、母平均が等しいという *t* 分布の帰無仮説が正しい場合として、母集団における平均と分散は等しいが、分布の形状だけ異なるように、一方の群の母集団は β 分布、他方の群はそれと同じ平均と分散を持つ正規分布を設定して標本データを発生させ、それに *t* 検定と *U* 検定を適用するという手順を10000回繰り返す。その結果、*t* 検定ではほぼ設定された有意水準と同じ比率で帰無仮説が棄却されたが、*U* 検定ではサンプルサイズが大きくなるほど帰無仮説が棄却される比率が大きくなった。

また、母集団における平均と中央値が等しく、分散のみが異なる2群の例として、一方の母集団を標準正規分布に、他方は平均を0に固定したまま分散を変えた正規分布を設定して同様のシミュレーションを行った。その結果、*t* 検定は Welch 検定による補正により、いずれの条件でも設定された有意水準とほぼ同じ比率で帰無仮説が棄却されるのに対して、*U* 検定では2群の母分散が大きくなるほど帰無仮説が棄却される比率が大きくなった。

以上より、ノンパラメトリック検定はあくまでも母集団分布が群間で同一といえるかどうかを見ているものであり、平均や中央値などの値の差異を検討する目的においては、必ずしも適切な手法ではなく、正規性の仮定からの逸脱に対して頑健であることが知られている *t* 検定の方がより妥当な結論を得られる場合もあることが示唆された。

②前提条件確認のための予備検定の是非

t 検定や分散分析の前提条件となる正規性

や等分散性の仮定が成立しているかどうかを確認するために、事前にそれらの前提条件に関する検定を実施し、その予備検定の結果次第で、正規性を仮定しないノンパラメトリック検定や等分散を仮定しない Welch 検定などを利用するという手順が用いられることがある。しかし、予備検定に基づく方法では第1種の誤りの確率が統制できない問題があり (Zimmerman, 1996 など)、常に Welch 検定等の制約の少ない方法を利用すべきという考え方もある。そこで、本研究では t 検定と Welch 検定の使い分けの問題について、母集団分布に関する条件を変えながら、第1種の誤りと検定力の両方の観点から検討を行った。その結果、母集団の正規性と等分散性がともに成り立っている場合、Welch 検定は通常の t 検定よりも保守的で、 t 検定では帰無仮説を棄却できるのに Welch 検定では棄却できないという場合があり、同一条件下での検定力は、「常に t 検定」 > 「等分散性の検定結果に基づいた t 検定または Welch 検定」 > 「常に Welch 検定」という大小関係があることが確認できた。実質的な検定力の違いとしてはそれほど大きくないため、第1種の誤りのコントロールの観点から常に Welch 検定を採用するという方法はバランスの取れた望ましいものであるが、等分散の仮定が成り立つことに確証が持て、かつサンプルサイズが十分ではなく、少しでも検定力を高めたい場合には、 t 検定の利用が望ましいと考えられる。

③主効果の事後検定としての多重比較手法の選択

3 群 (条件) 以上の平均値を比較する分散分析において、主効果が統計的に有意である場合に、どの群 (条件) 間で有意な差があるのかを確認するための多重比較を事後検定として実施するのが一般的である。多重比較の検定は非常に多くの方法が提案されており、表1においても、方法名が明示されたものだけでも7種類みられる。最もよく利用されていたのが Bonferroni 法であるが、Bonferroni 法は特に比較する群の数が増えるほど、各対比較における有意水準が低くなるため、有意差の出にくい保守的な方法であることが知られている。Bonferroni 法を修正して段階的に有意水準を変化させる Holm 法は、全体で第1種の誤りを犯す確率を統制しつつ、Bonferroni 法で有意差のみられる対はすべて検出した上で、さらに追加で有意差が得られる可能性のある、一様に検定力の高い方法であり (永田・吉田, 1997)、単に有意差の有無を議論する場合には基本的に Bonferroni 法はすべて Holm 法で置き換えることができ、検定力の観点からも望ましいといえる。

表1で2番目に多い多重比較の方法である Tukey 法は対応のない独立な群間の比較を前提とする (永田・吉田, 1997) が、人工デー

タによるシミュレーションでは、対応のある場合でも特に問題が生じない場合もあり、また、Bonferroni 法を修正した Holm 法や Shaffer 法との比較においては、いずれかの方法でしか検出できない群間差がみられるが、どの多重比較の方法を選択するのが最善であるかを事前に決定するための一般的な基準を得るには至らなかった。これらの多重比較法については、一律に望ましい方法を考えるのではなく、少なくとも複数の選択肢として用意しておくことが望ましいと考えられる。

④交互作用の事後検定手法の選択

2 要因以上の分散分析で統計的に有意な交互作用効果が見られた場合の事後検定として、表1では、ある要因の水準ごとに他の要因の主効果 (または他の要因間の交互作用) を検定する単純効果の検定が唯一みられる方法となっている (ただし、一般線型モデルにおいて、量的変数が関わる有意な交互作用がみられる場合の事後検定の方法がいくつか見られたが、本研究の趣旨から表1では省略している)。しかし、有意な主効果に対してより詳細な情報を求める多重比較とは異なり、単純効果はどのような交互作用があるのかという詳細を調べる方法にはなっていない。例えば、2 要因で各要因が2水準の場合、仮に母集団における各水準の平均が図1のような状態である場合、十分な検定力を確保した状態では、有意な交互作用がみられ、かつ、A1 水準、A2 水準ともに $B1 < B2$ という有意な単純主効果がみられることになるが、この情報からは、どのような交互作用があるのかわからない。また、検定力が十分でない場合は A2 水準でのみ有意な単純主効果がみられる可能性があり、A1 水準における単純主効果の結果との違いが交互作用の詳細と解釈されることもあるが、これは第2種の誤りの結果であり、正しい知見が得られたことにはならない。

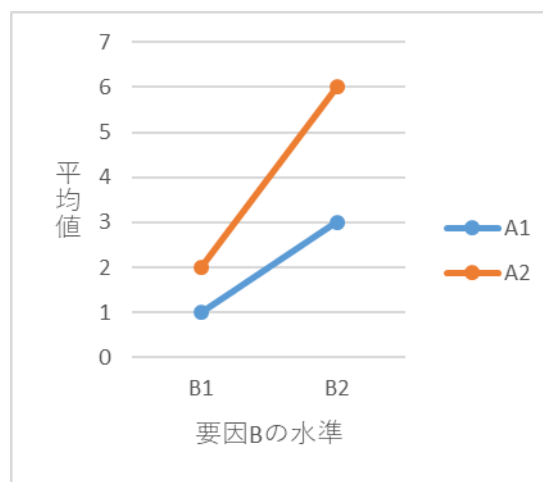


図1 交互作用が存在し、かつ単純主効果の方向性が一致している例

図1の場合に、交互作用の詳細としては、「A1水準よりもA2水準の方が、B1とB2の平均値差が大きい」という結論を得る必要がある。このような結論を直接得ることのできる手法として、処理対比交互作用(treatment-contrast interaction)の検定(Marascuilo, & Levin, 1970)があり、有意な交互作用の詳細を調べる事後検定としては、このような方法を用いるべきである。

⑤分析時において独立変数と従属変数が逆転することによる影響

例えば、「ある量的変数 x の得点が高い人ほど、ある質的変数 y で特定のカテゴリを選択する傾向がある」という仮説を検証したい場合、仮説上では量的変数 x が独立変数、質的変数 y が従属変数として想定されているにもかかわらず、データ分析の際には、 y の各カテゴリを選択した群間で x の平均値を比較する検定 (t 検定や分散分析) を行うというように、仮説と分析モデルで独立変数と従属変数が逆転している場合がある。このような変数の役割の逆転が検定結果に及ぼす影響について人工データによるシミュレーションにより確認した。

独立変数が1つの量的変数、従属変数が質的変数である統計モデルであるロジスティック回帰分析モデルに従って発生させた乱数データに、 t 検定等の平均値差の検定とロジスティック回帰分析の両方を適用し、両手法の結果を比較した。その結果、従属変数のカテゴリ数が2つであるときは、ほとんどの場合において t 検定とロジスティック回帰分析の p 値がほぼ等しく、少なくとも「変数間の有意な効果の有無」については実質的に同じ結論が得られたが、データ発生モデルにおける独立変数の効果が大きい場合などで、両手法の p 値が大きく食い違うケースがしばしば生じることがわかった。また、質的変数のカテゴリ数が3つ以上の場合については、分散分析や多重比較による平均値差の検定との一貫性は見られなかった。

⑥質的変数の連関の検定における事後検定

質的変数に関する連関の検定(カイ2乗検定またはFisherの直接確率法)で有意な連関がみられた際の事後検定として、表1より、ほとんどの場合においてクロス集計表のセルごとに調整済み残差の有意性を検定する残差分析が用いられていることがわかる。しかし、残差分析はセルごとに独立になされるものであり、セル間の比較を直接行う検定ではない。したがって、分散分析における多重比較のような「どのカテゴリ間で有意な差異があるか」という結論を検定結果から導くことはできない。

クロス集計表全体で有意な連関がみられた場合に、事後検定として、どのカテゴリ間で有意な違いがみられるかを調べる多重比較の方法として、廣津(2007)の方法や松田

(2004)の方法などが提案されている。カテゴリ間の比較が目的の研究においては、事後検定としてこれらの手法が候補となる。

(3) 分析用プログラムの開発

本研究において、従来慣習的に用いられてきた事後検定の代替法として望ましいものとして取り上げた、分散分析における交互作用に対する処理対比交互作用の検定および、連関の検定における松田(2004)の閉検定手順に基づく多重比較を、R上で実行するための関数を作成した。本検定との連携や、適応可能な群の個数の制約などの点において、より汎用性のある使いやすいツールとなるよう、今後さらに改良を続けていく必要があると考えられる。

<引用文献>

①廣津千尋(2007). 分割表における行の多重比較法について 応用統計学, 36(2), 1-17.

②Marascuilo, L. A. & Levin, J. R. (1970). Appropriate post hoc comparisons for interactions and nested hypotheses in two-factor analysis of variance designs, *American Educational Research Journal*, 7, 397-421.

③松田眞一(2004). 名義尺度の分割表に対する多重比較法 南山大学紀要「アカデミア」数理情報編, 4, 29-37.

④永田靖・吉田道弘(1997). 統計的多重比較法の基礎 サイエンティスト社

⑤Zimmerman, D. W. (1996). Some properties of preliminary tests of equality of variances in the two-sample location problem, *The Journal of General Psychology*, 123, 217-231.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計1件)

① 杉澤武俊・吉田寿夫・荘島宏二郎・南風原朝和(2016). 研究法におけるルーチンの見直し 教育心理学年報, 55, 234-242. 査読無

[学会発表] (計3件)

① 杉澤武俊(2017). 「気になる統計解析」あれこれー理想的な統計解析と現実的な制約との間のジレンマー 日本教育心理学会第59回総会 2017年10月9日 名古屋国際会議場

② 杉澤武俊(2015). 統計的検定の実践に関わる諸問題(準備委員会企画シンポジ

ウム「研究法におけるルーチンの見直し」) 日本教育心理学会第 57 回総会
2015 年 8 月 26 日 朱鷺メッセ新潟コン
ベンションセンター

- ③ 杉澤武俊(2013). 研究で利用できるデータ分析と統計的検定－検定力分析の実践を通して－ 日本行動計量学会岡山地域部会第 48 回研究会 (招待講演)
2013 年 7 月 6 日 岡山理科大学

[図書] (計 2 件)

- ① 杉澤武俊(2017). 検定力分析に基づくサンプルサイズ設計 (村井潤一郎・橋本貴充 (編著) 心理学のためのサンプルサイズ設計入門) 講談社 総頁数 166 (Pp. 21-41)
- ② 山田剛史・村井潤一郎・杉澤武俊(2015). R による心理データ解析 ナカニシヤ出版 総頁数 272 (Pp. 35-37 (コラム), 135-210 (第 3 章))

6. 研究組織

(1) 研究代表者

杉澤 武俊 (SUGISAWA, Taketoshi)

新潟大学・人文社会・教育科学系・准教授

研究者番号 : 30361603