

令和 4 年 6 月 20 日現在

機関番号：30110

研究種目：若手研究

研究期間：2019～2021

課題番号：19K13044

研究課題名(和文)統計的仮説検定の歴史的源流と哲学的意義に関する研究

研究課題名(英文)A historical and philosophical study on null hypothesis significance testing

研究代表者

森元 良太 (MORIMOTO, Ryota)

北海道医療大学・リハビリテーション科学部・准教授

研究者番号：70648500

交付決定額(研究期間全体)：(直接経費) 1,900,000円

研究成果の概要(和文)：統計的仮説検定は仮説を検定する一般的な方法だが、近年この検定法の誤解や誤用が氾濫し、さまざまな分野で警鐘が鳴らされている。統計的仮説検定はフィッシャー流の有意性検定とネイマンピアソン流の仮説検定を混成した検定法である。フィッシャーとネイマンピアソンは検定法に関する科学哲学的な論争を繰り広げた。この論争は解決をみずに終結し、現在でも科学哲学的問題は棚上げにされたままである。本研究では、統計的仮説検定の誤解や誤用される要因を究明するため科学史・科学哲学の観点から考察する。

研究成果の学術的意義や社会的意義

生物学や心理学などの統計解析を使用する分野では、科学哲学的な問題に踏み込むよりも実用的な要請への対応が優先される。本研究は、統計的仮説検定の誤解と誤用に正面から向き合い、史的考察と哲学的分析を通じて、統計的仮説検定の適切な使用法を示す点が創造的であり、科学史・科学哲学のみならず統計教育の領域にも貢献できることが期待できる。

研究成果の概要(英文)：Null hypothesis significance testing (NHST) is the most widely used statistical method. Nevertheless, misuse and misinterpretation of NHST are widespread, and severe criticisms have been levelled against NHST. NHST is an incoherent hybrid of two different testing theories: Fisher's significance testing and Neyman-Pearson hypothesis testing. Fisher on one side and Neyman-Pearson on the other were engaged in a heated controversy as to philosophy of testing theory. In this study, I review the primary textbooks on NHST and conduct a historical and philosophical survey to clarify the source of misuse and misinterpretation of NHST.

研究分野：科学哲学

キーワード：科学哲学 統計的仮説検定 有意性検定 仮説検定

1. 研究開始当初の背景

現在、生物学や心理学、医学や社会学など様々な分野において、仮説を検定するときに「統計的仮説検定」(統計的帰無仮説検定や統計的検定とも呼ばれる)が用いられる。統計的仮説検定では、検討する仮説そのものではなく、その仮説を否定した帰無仮説が評価される。帰無仮説のもとで得られたデータが統計的に確からしいかどうかで、帰無仮説やもとの仮説を棄却するかどうかを決める。統計的仮説検定は現在の統計学ではごく一般的な手法であり、統計学の教科書では必ずこの検定法が紹介されている。

しかし近年、統計的仮説検定の誤解や誤用があまりに多く、その使用について注意喚起が促されている。2015年の*Nature*誌では、統計的仮説検定を用いた心理学研究の再現性が低いことが指摘され、同年*Basic and Applied Social Psychology*誌の冒頭で編集長が統計的仮説検定だけ使用した論文を受理しないと宣言し、2016年にアメリカ統計学会が統計的仮説検定の誤解と誤用に対する警告を声明した。2019年に同学会は統計的仮説検定についての特集を組み、統計的仮説検定の使用に対するより過激な警鐘を鳴らしている。統計的仮説検定の誤解と誤用は、 p 値という統計量を仮説の棄却基準としていることから「 p -ハッキング」と呼ばれたり、実験の再現性の低さから「再現性の問題」などと呼ばれ、様々な分野で深刻な問題として取り上げられている。それに対し、統計的仮説検定の使用を制限したり、ベイズ統計学などの別の統計解析法を推奨するなど、場当たりの対処法が提案されているのが現状である。

このような誤解や誤用が生じる根源の一つは、統計的仮説検定が登場する歴史的経緯にあると考えられる。統計的仮説検定は、R. フィッシャー流の有意性検定とJ. ネイマン E. ピアソン流の仮説検定を混成したものである。フィッシャーは1920年代に科学の方法論として有意性検定を考案し、ネイマンとピアソンは科学だけでなく商業などにも応用できるよう有意性検定を拡張し、また数学的な基盤固めにも努めた。ところが、二種類の検定法は異なる手法であり、実際フィッシャーとネイマン ピアソンは検定の前提や目的、哲学的意義などについて科学的方法論や科学の本性といった科学哲学的問題にも踏み込み、激しい論争を繰り広げた(Lehmann 2011、芝村 2004)。この論争はフィッシャーの死により決着をみないまま終結する。その一方で、統計ユーザーは目下の研究や調査のために統計解析法を使用せざるをえず、折衷案として二種類の検定法を混成した統計的仮説検定が提案され、科学哲学的問題は棚上げにされてしまう。統計的仮説検定が登場するのは両陣営が論争中の1940年代で、E. リンドキスト(1940)の*Statistical Analysis in Educational Research*やQ. マクネマー(1949)の*Psychological Statistics*などの教科書に掲載されている。それ以降現在に至るまで、統計的仮説検定は統計学の教科書に掲載され、統計教育では標準とされている。このように、現在の統計的仮説検定は実用上の要請により折衷案として提出され、棚上げにした科学哲学的問題を下すことなく、現在まで仮説の一般的な検定法として教育されてきたことが、統計的仮説検定の誤解と誤用を生む根源の一つと考えられる。そこで本研究では、「統計的仮説検定はなぜ誤解や誤用されるのか」という問いに科学史・科学哲学の観点から考察を加える。

2. 研究の目的

本研究の目的は、統計的仮説検定の誤解と誤用を生む根源を究明することである。そのため、まずフィッシャー流の有意性検定とネイマン ピアソン流の仮説検定の前提、目的、哲学的意義を比較する。次に、この二種類の検定法が統計的仮説検定に混成された歴史的源流を辿り、統計的仮説検定の誤解や誤用の根源を究明する。最後に、統計的仮説検定の前提、目的、適用範囲、哲学的意義を明らかにし、適切な使用法を提示する。そのため、統計的仮説検定を提唱し普及させたリンドキストやワルドらの検定法や科学についての考え方を究明し、統計的仮説検定の前提や目的を明示する。統計学の歴史研究は、D. マッケンジーの*Statistics in Britain, 1865-1930*やA. ハルドの*A History of Mathematical Statistics from 1750 to 1930*のように、有意性検定が登場する1930年代までのものが多く、統計的仮説検定として混成される1940年代以降の研究は少ない。本研究は1940年以降のリンドキストやワルドらの科学的方法論や科学観を分析し、統計的仮説検定を哲学的観点から考察する。

生物学や心理学などの統計解析を使用する分野では、科学哲学的な問題に踏込むよりも実用的な要請への対応が優先される。そのため、統計的仮説検定の誤解や誤用の氾濫に対する警鐘が鳴らされると、その使用の制限やベイズ統計学などの別の検定法の使用が推奨されている。しかし、統計的仮説検定が濫用されていることと、その検定法自体が誤っていることは異なる。誤解や誤用への解決策はそうした濫用を正すことであり、使用を回避することではない。本研究は、統計的仮説検定の誤解と誤用に正面から向き合い、史的考察と哲学的分析を通じて、統計的仮説検定の誤解や誤用の根源を探ることで、科学史・科学哲学のみならず統計教育の領域にも貢献できることが期待できる。

3. 研究の方法

本研究では、フィッシャー流の有意性検定、ネイマン・ピアソン流の仮説検定、統計的仮説検定についての一次文献調査や関連文献を精読することにより、次の3つの点を究明した。

(1) フィッシャー流の有意性検定とネイマン・ピアソン流の仮説検定の前提や目的、学的意義などの異同を分析するとともに、フィッシャー、ネイマン、ピアソンの科学についての考え方の違いを明らかにする。そのため、R. フィッシャー、J. ネイマン、E. ピアソンの一次文献、および統計学者 E. レーマン(2011)の *Fisher, Neyman, and the Creation of Classical Statistics* や芝村(2004)の『R. A. フィッシャーの統計理論』などの文献を用い、フィッシャー流の有意性検定とネイマン・ピアソン流の仮説検定の比較をおこなった。

(2) 二種類の検定法が統計的仮説検定として混成された歴史的源流を辿り、統計的仮説検定が導入された経緯を探る。統計的仮説検定が最初に登場したのは、1940年出版のリンドキストの統計学の教科書 *Statistical Analysis in Educational Research* とされる。まずはリンドキストの文献を精読し、リンドキストの科学や統計的仮説検定についての思想を明らかにする。また、リンドキストの他の教科書も検討する。次に、リンドキスト以降 1950年代にかけての統計学の代表的な教科書である H. ウォーカー(1943)の *Elementary Statistical Methods* や Q. マクネマー(1949)の *Psychological Statistics* などの文献をもとに、統計的仮説検定が普及していく歴史的経緯を辿る。

4. 研究成果

フィッシャー流の有意性検定、ネイマン・ピアソン流の仮説検定、統計的仮説検定についての一次文献調査や関連文献を精読することにより、以下の成果が得られた。

(1) まずは、フィッシャー流の有意性検定とネイマン・ピアソン流の仮説検定の数学的前提と目的の相違点を整理した。二つの検定の大きな違いは、(a) 仮説への態度、(b) 検定についての考え方、(c) 対立仮説の有無について、(d) 抽出についての考え方にある。(a) について、有意性検定では仮説の棄却しか認められない。フィッシャーは、科学において真偽の定かでない仮説を採択するという大胆な態度を採るべきではなく、疑わしい仮説を棄却するだけにとどめておくべきだと考えているからである。一方、仮説検定では仮説の棄却だけでなく、採択も認められる。(b) について、フィッシャーは有意性検定を科学的な帰納的推理だと捉えるのに対し、ネイマン・ピアソンは過去の経験から行動を学習する帰納的行動と捉えた。(c) について、有意性検定では仮説として帰無仮説(とその否定)しか設定されないのに対し、仮説検定では帰無仮説のほかに対立仮説が設けられる。ただし、ネイマンたちは帰無仮説という表現を避け、単に仮説と呼んでいる。また、対立仮説は1つだけでなく、複数設定することを認めていた。(d) について、フィッシャーは同じ母集団から繰返しの抽出はできないと考える。これは、検定を適切におこなうには母集団を明確に定める必要があり、実験が異なれば母集団も変わるからである。一方、ネイマン・ピアソンは同じ母集団から繰返し抽出することを想定している。ネイマンは研究者個人の経験を何度も繰返せば同じ分布に収束すると考えた。

こうしたフィッシャーとネイマン・ピアソンの違いは科学観の違いに依拠することを明らかにした。フィッシャーは科学的推論を中心に科学を捉え、検定は帰納的推論だと主張した。そして、科学は誤った仮説を棄却するだけで、真偽の不明な仮説の採択は認めない。検定は誤りを排除するためのもので、科学は仮説を棄却していく作業というポパーの反証主義的な考え方をフィッシャーは採っている。一方、ネイマンは、科学者の意思決定を中心に科学を捉えている。仮説検定は、科学者が経験を積上げて行動を決めるために用いられ、検定は帰納的行動だと主張した。また、ピアソンも科学者の意思決定を中心に科学の本質を捉えていた。ピアソン自身は検定を帰納的行動だとは主張しないが、科学者が観察結果をもとに学習する方法だと理解している。ネイマンとピアソンは仮説検定を科学者の意思決定の理論と捉えている。以上の分析から、フィッシャーとネイマン・ピアソンの科学観の違いは、フィッシャーが科学的推論という理論的側面を重視しているのに対し、ネイマン・ピアソンは科学理論というよりも科学者という人に注目していることに起因する点が明らかになった。

(2) 2つの検定法が統計的仮説検定として混成された歴史的源流を辿るため、主に1940年代から1950年代の統計学の教科書の調査をおこなった。有意性検定と仮説検定を最初に混成した教科書は1940年出版のリンドキストによる *Statistical Analysis in Educational Research* である。この教科書では、検定の理論をフィッシャー流の有意性検定として解説しているが、そのなかにネイマンやピアソンに言及せずに仮説検定の二種類の過誤の説明を混入させている。そのため、有意性検定であるにもかかわらず、フィッシャーが拒絶した「仮説の採択」や「対立仮説」の設定を導入してしまっている。また、二種類の過誤については、ネイマンとピアソンの偉業である「ネイマン・ピアソンの補題」にも触れられていない。しかも、リンドキストの教科書には参考文献表がなく、またフィッシャーへの言及はあるが、ネイマンとピアソンについては言及がない。さらに両陣営の激しい論戦についても言及されてなく、二種類の検定法の違いも考慮されていない。このように、有意性検定と仮説検定の目的や哲学的意義を棚上げにしたまま、2つの検定法は統計的仮説検定として混成されることになった。リンドキストは1953年にも *Design and Analysis of Experiments in Psychology and Education* という教科書を出版しているが、

混成理論の説明については1940年の著書とあまり変わりがない。この2冊の著書は教育学や心理学の研究者を対象としている。リンドキストは科学での使用を想定していたフィッシャーの有意性検定として紹介していたが、対象としたのは教育学や心理学であり、フィッシャーの想定した科学とは異なっていた。

リンドキストの教科書出版後、1943年にH. ウォーカーが *Elementary Statistical Methods* を出版した。この教科書は、日常生活の一般的特徴の理解を目的として統計手法が紹介されている。この著書の読者対象は、心理学、社会学、経済学などであり、フィッシャーの想定した科学ではない。また、ウォーカーは統計的検定の目的を行為の選択だと主張する。この教科書は記述統計学と統計的推定の解説が大半で、統計的仮説検定については後半の1章分でしか扱っていない。そのため、フィッシャーとネイマン - ピアソンの論争は取り上げられてなく、哲学的問題は棚上げにされている。統計的検定法については、基本的にフィッシャーの有意性検定ではなく、ネイマン - ピアソン流の仮説検定が解説されている。1949年にはQ. マクネマーが *Psychological Statistics* という教科書を出版する。ハプランとスタム(2006)は、この教科書によって統計的仮説検定の混成理論が生み出されたとしている。この教科書では、混成理論の手続きが簡潔に説明されていた、またその手順が表として示されているため、研究者に利用しやすくなっている。ただし、リンドキストの教科書と同じく参考文献表がないため、読者にはフィッシャーとネイマン - ピアソンの考え方の違いを判別できない。また、彼らの論争の中身についても言及がなく、統計的仮説検定という混成理論が1つの完成された理論のように紹介されている。他の教科書も程度の差はあれ、同様の解説が加えられている。このような文献調査から、1940年代以降の統計学の教科書では、フィッシャー流の有意性検定とネイマン - ピアソン流の仮説検定の違いには触れず、哲学的な問題を棚上げにしたまま、あたかも1つの確立した検定理論であるかのように統計的仮説検定が導入されたことを明らかにした。また、フィッシャーは科学理論として検定法を構築したが、1940年代以降の統計的検定理論は科学理論というよりは、科学以外含む広範分野を射程にする意思決定の理論として普及したことも明らかにした。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 4件）

1. 著者名 森元良太	4. 巻 97
2. 論文標題 「科学哲学からみた統計解析法 シングルケースデザインを中心に」・	5. 発行年 2020年
3. 雑誌名 日本行動分析学会ニューズレター	6. 最初と最後の頁 3-7
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 森元 良太	4. 巻 35
2. 論文標題 行動分析学の科学哲学的一試論 最節約性の観点から	5. 発行年 2021年
3. 雑誌名 行動分析学研究	6. 最初と最後の頁 165～176
掲載論文のDOI（デジタルオブジェクト識別子） 10.24456/jjba.35.2_165	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 MORIMOTO Ryota	4. 巻 30
2. 論文標題 <i>Stop and Think About p-Value Statistics: Fisher, Neyman, and E. Pearson Revisited&/i>	5. 発行年 2021年
3. 雑誌名 Annals of the Japan Association for Philosophy of Science	6. 最初と最後の頁 43～65
掲載論文のDOI（デジタルオブジェクト識別子） 10.4288/jafpos.30.0_43	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 丹野貴行・山本淳一・武藤崇・澤幸祐・森元良太・竹内康二、山岸直基	4. 巻 106
2. 論文標題 徹底的行動主義の現代的な位置づけをめぐる諸論 討論会	5. 発行年 2022年
3. 雑誌名 日本行動分析学会ニューズレター	6. 最初と最後の頁 2～39
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 森元良太	4. 巻 36
2. 論文標題 合理的な理論選択の一基準としてのオッカムのかみそり 島宗 (2022) へのリプライ	5. 発行年 2022年
3. 雑誌名 行動分析学研究	6. 最初と最後の頁 217 ~ 225
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件 (うち招待講演 2件 / うち国際学会 0件)

1. 発表者名 森元良太
2. 発表標題 擬人主義がまじな科学的研究プログラムかどうかを統計学の考え方に基づいて考察する
3. 学会等名 日本進化学会
4. 発表年 2019年

1. 発表者名 森元良太
2. 発表標題 科学哲学からみた統計解析法 シングルケースデザインを中心にー
3. 学会等名 日本行動分析学会 (招待講演)
4. 発表年 2019年

1. 発表者名 森元良太
2. 発表標題 科学哲学からみたp値統計学の思考法
3. 学会等名 公立はこだて未来大学 行動経済学セミナー (招待講演)
4. 発表年 2022年

〔図書〕 計2件

1. 著者名 エリオット・ソーバー、森元 良太	4. 発行年 2021年
2. 出版社 勁草書房	5. 総ページ数 368
3. 書名 オッカムのかみそり	

1. 著者名 日本科学史学会	4. 発行年 2021年
2. 出版社 丸善出版	5. 総ページ数 758
3. 書名 科学史事典	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------