

論文の内容の要旨

論文題目 集団単位でのデータ収集によって生じる観測値の非独立性が
統計的検定の危険率および検定力に与える影響

氏名 安田 佳代子
(栗田)

教育心理学などの研究においては、データ収集を行う際に学級などの既存の小集団をそのままデータの収集単位として用いることが多い。このデータ収集方法では、統計的検定の重要な前提条件の一つである観測値の独立性が損なわれる。しかも統計的検定は、一般にこの観測値の独立性という前提条件からの逸脱の影響を受けやすいこと、すなわち頑健でないことが知られており、統計的検定の適用上重要な問題の一つとなっている。そこで、本論文では、統計的検定の中でももっとも広く利用されている手法の一つである t 検定をとりあげ、小集団を単位としたデータ収集が原因となっている観測値の独立性からの逸脱が、統計的検定におよぼす影響について、様々な角度から検討することを目的とした。

第1章 序論では、本論文において扱う問題の背景についてまとめ、これまでの研究を

概観し、本論文における議論の準備を行った。まず、統計的検定の理論について整理し、研究方法としての現状を述べるとともに、特に本論文で扱う t 検定について詳述した。次に、教育心理学などの研究において特徴的である学級などの既存の集団を単位としたデータ収集について整理し、このデータ収集方法が統計的検定の前提条件である観測値の独立性を損なう要因となることを論じた。さらに、本論文の研究の基礎となる、小集団を単位としたデータ収集によって得られる観測値のモデルを示した。また、観測値の独立性からの逸脱の問題を中心に統計的検定の頑健性についてのこれまでの研究を概観し、明らかになっている部分と残されている課題について指摘した。さらに、観測値の独立性から逸脱している場合の代替法について論じた。最後に本論文の目的についてまとめた。

観測値の独立性が満たされない場合における統計的検定の代替法の一つとして、修正検定統計量を用いる方法がある。これは、非独立性の度をあらわす指標を検定統計量に組み込むことによって、危険率に関する頑健性を高めるという方法であり、この検定統計量（以下、修正検定統計量と呼ぶ）を利用すれば、設定したとおりの危険率で統計的検定を行うことができる。しかし、この修正検定統計量に対応した検定力については明らかとなっていない。そこで、第2章 検定統計量の修正と検定力では、修正検定統計量を用いたときの検定力の算出方法について提案を行った。修正検定統計量に対応した検定力は、非心 t 分布において非心度に関して修正を行うことにより得られることを示した。

また、観測値の独立性が満たされていない場合の危険率および検定力（それぞれ実測危険率、実測検定力と呼ぶ）は、従来コンピュータ・シミュレーションによって推定されてきたが、ここでは、修正検定統計量および修正検定力の考え方を利用して、実測危険率および実測検定力を推定する方法についても提案を行った。この方法により、ある限定された条件下ではコンピュータ・シミュレーションによる方法よりも正確にかつ簡便に統計的検定の頑健性を評価することが可能となった。

観測値の非独立性の問題を扱う場合には、まず、観測値のモデル化が行われる。このモデルはコンピュータ・シミュレーションにおけるデータ発生モデルや代替法を数学的に導出する際のモデルとして用いられるなど、統計的検定の頑健性に関する研究の要となる重要な役割を持つ。しかし、このモデルが、現実の状況で生じる観測値の非独立性の適切なモデルになっているかどうかについては、これまでほとんど検討されることが

なかった。そこで、第3章 人工データと実際データとの対応において、モデルと現実データの対応についての検討を行った。検討方法としては、特定の数学的なデータ発生モデルに従って発生させた人工データによる結果と、現実の大きなデータセットから小集団単位で標本抽出することによって得られる結果の比較を行い、人工データによる結果の妥当性の検討を行った。二種類の大きなデータセットについて検討を行ったところ、いずれについても実際のデータを用いて得られた結果は人工データを用いた結果から十分正確に予測できていた。したがって、観測値の非独立性を表すモデルのもとで得られた結果は、モデルで設定された分布の正規性などの条件からはずれているデータに対しても一般化が可能であるということが示唆された。

教育心理学などの研究では、学力や性格特性など直接的な測定の難しい構成概念がしばしば研究対象とされるが、その測定精度を高める工夫として、それらの構成概念を測定していると思われる複数の変数から合成変数を作成するという手続きが広く行われている。あるいは、情報を要約する方法の一つとして類似した項目を事後的にまとめることで合成変数が作成されることも一般的である。これらの手続きは項目の尺度化と呼ばれるが、この手続きが統計的検定の頑健性にどのような影響をおよぼすかについてはほとんど明らかにされていない。一般に小集団単位のデータ収集による観測値の非独立性は、小集団内の類似性および小集団間の差異を反映するものである。複数の変数を合成した場合、そうした類似性や差異がより顕著になり、結果的に統計的検定の頑健性を低下させる可能性がある。合成変数が使用されることの多い現状を考えると、これは実際上重要な問題であるといえる。そこで、第4章 項目の尺度化の影響では複数の項目得点の和によって定義される尺度得点のように、いくつかの変数を合成した変数が t 検定に用いられるとき、そのことが頑健性にどのような影響をおよぼすかを検討した。実際のデータを用いた検討の結果、項目の尺度化は、やはり観測値間の相関を大きくする方向に作用し、統計的検定の頑健性に影響をおよぼしうることが明らかとなった。項目の尺度化は、危険率については単純に頑健性を低める方向に働くが、検定力については複雑な影響をおよぼすことがわかった。

以上の章では、観測値の独立性が満たされない場合の代替法についての考察を行い、さらに観測値の独立性から逸脱している場合に t 検定を適用するときの実際的な問題を中心に検討を行った。そして、 t 検定の危険率および検定力への影響について一般的な傾向をとらえることができた。しかし、一方、実際に検定を適用する立場の研究者の関

心は、自分のデータに対してその検定法がどれだけ頑健かということにある。個々の研究者が扱うデータの性質はきわめて多様であるため、これまでの限られた数の頑健性研究でそのすべての可能性を網羅することは不可能であった。第5章 個別のデータ収集状況に合わせた危険率および検定力の推定プログラムでは頑健性研究のこうした限界をふまえ、研究者の個別のデータ収集状況に合わせて頑健性を検討することのできるソフトウェアの開発を行った。本章で開発された LocalSim は、観測値の独立性をはじめとして正規性および等分散性など様々な前提条件からの逸脱状況における t 検定の実際の危険率および検定力の推定を行うプログラムである。まず、LocalSim の機能および構成について解説し、具体的な実行例を2つ挙げて操作をたどり、柔軟な設定が可能な有用なプログラムであることを示した。LocalSim は、コンピュータに不慣れな人でも使いやすいうようにインターフェイスに十分配慮し、個別の研究における危険率および検定力に関する t 検定の頑健性について簡単に確かめることができるようになっている。本ソフトウェアは、実際に統計的検定のユーザーが自身の研究に役立てることができる他、統計教育においても有効に活用できるものである。

最後に、第6章 まとめと今後の展望において、本論文のまとめと今後の展望を述べた。本論文では、主として修正検定統計量に対応した検定力の算出法の提案、実測危険率および実測検定力の算出法の提案、人工データと実際データとの対応の検討、項目の尺度化が t 検定の頑健性におよぼす影響の検討、および個別のデータ収集状況において危険率および検定力を推定するソフトウェアの提案等を行った。今後に残された課題としては、(1)集団単位以外の原因で生じる観測値の非独立性に対する頑健性の検討、(2) t 検定以外の検定法の頑健性の検討、(3)観測値の非独立性を表す指標の推定に関わる問題、(4)ソフトウェアの拡張の4点を挙げた。