

論文の内容の要旨

論文題目 実際の研究場面におけるデータ収集および分析の手続きが
統計的検定の検定力に及ぼす影響

氏 名 杉 澤 武 俊

心理学の分野において、統計的検定はデータ解析の中心的な役割を担うことが多い。通常、研究者は帰無仮説を棄却して有意な結果を得ることを目的とするから、誤った帰無仮説を正しく棄却できる確率である検定力は研究の結論そのものを左右することになる。理論的に検定力を規定する要因として、有意水準と母集団効果量と標本の大きさの3つが挙げられるが、実際の研究で検定が適用される場面では、具体的なデータ収集の方法や分析の進め方など、研究者のさまざまな判断が検定力に影響する可能性がある。本論文は、現実に検定が適用される場面でのそうした具体的なデータ収集および分析の手続きが検定力にどのような影響を与えるのかを明らかにすることを目的とする。

第1章「序論」では、本論文で扱う題材に関して基礎となる理論的事項を整理した。まず、検定力に関する研究について概観し、心理学研究において検定力があまり重視されてこなかったことに対する問題点について述べた。次に検定力に関する理論的基礎、特に、検定力を理論的に扱う上で重要な概念となる効果量について、その定義および対立仮説のもとでの検定統計量の標本分布との関連について解説した。また、検定力を規定する要因について整理を行い、具体的なデータ収集および分析の手続きの選択など、

研究者の裁量が及びうる範囲においても検定力に影響するものがあることを述べた。さらに、第5章および第6章で取り上げる潜在曲線モデルについて予備的解説を行った。最後に本論文の目的をまとめた。

第2章「日本の心理学研究における検定力と標本効果量」では、1992年から1996年までの5年間に発行された『教育心理学研究』誌に掲載された250編の論文を対象として研究で用いられた標本の大きさを調査し、小・中・大の各効果量のもとで、どの程度の検定力をもつのかを計算した。

対象となった全論文について集計した結果、中効果量のもとでの検定力の中央値は、欧米で行われた先行研究と比べるとやや高いことがわかった。また、研究領域ごとに集計したところ、同一の効果量を仮定したときの検定力の値が領域間で大きく異なっていた。その背景としてそれぞれの研究領域において比較的良好に用いられる研究法に違いがあり、一般的に大規模な標本を得やすい「調査」が多用される領域は、「実験」「観察」が多用される領域に比べて同一効果量のもとでの検定力が高くなることが示唆された。また、有意でない結果をもって帰無仮説を支持するためには特に高い検定力を必要とするが、帰無仮説を研究仮説としている研究の検定力は十分に高いとはいえなかった。

各研究において実際に得られた標本効果量を検定手法ごとに求めたところ、標本効果量と標本の大きさの逆数との間に正の相関があることがわかった。このことは、研究者は予想される効果量に応じて標本の大きさを調節することで検定力をある一定の水準に保とうとする「直感的検定力分析」を行っている可能性を示唆するものである。

第3章「連続変数のカテゴリ化による検定力への影響」では、連続的な値を取る2変数間の関係を調べる際に、一方の変数の値に基づいて「高群」「低群」などの群分けをして群間の平均値差を検定したり、両方の変数をカテゴリ化して無相関検定やクロス集計表における独立性の検定（カイ2乗検定）を行ったりした場合に、連続変数のまま無相関検定を行う場合と比べて検定力にどのような影響があるのかを調べた。

まず、2変数正規分布に従う変数の一方の値によって、母集団における比率が等しくなるように切断した分布を考え、その切断分布における他方の変数の周辺分布に基づいて、分散分析を行う場合の母集団効果量を算出した。次に、双方の変数について同様に切断してできる各群が母集団全体に占める比率から、カテゴリ化した変数の無相関検定およびカイ2乗検定を行う場合の母集団効果量を算出した。その結果、2変数正規分布の相関係数が0でない場合は、いずれの検定においても分割するカテゴリ数が多くなる

ほど母集団効果量は大きくなることが示された。しかし、算出された母集団効果量に基づいて検定力を計算すると、いずれの検定も連続変数の無相関検定の検定力を上回ることがないほか、全体の標本の大きさが一定ならば、分散分析やカイ2乗検定では分割するカテゴリ数が多くなるほど検定力が低下することがわかった。この結果は人工データを発生させるシミュレーションでも確認された。また、標本データの分布に基づいて事後的に変数のカテゴリ化を行った場合でも概ね同程度の検定力となることがわかった。

さらに、検定力の観点から見たカテゴリ化に伴う損失の大きさを、母集団効果量および標本の大きさに換算した値を求めた。

第4章「2段抽出法によるデータ収集が無相関検定に及ぼす影響」では、母集団全体における相関係数に関する検定を行う際に、便宜的に2段抽出法を採用した場合の影響について検討した。

まず、母集団全体の相関係数（母集団効果量）、「クラスタリング効果」を表す各観測変数の級内相関係数、1次抽出単位（クラスタ）内における観測変数間の相関係数、クラスタ間における観測変数の平均値の相関係数という5つの母数の関数関係を導出し、母数に関する条件設定を行う際の制約を確認した。

上記5つの母数とクラスタ数およびクラスタサイズの値を変えて、人工データを発生させるシミュレーションを行った。その結果、2変数ともクラスタリング効果がない場合には、危険率、検定力ともに2段抽出による影響はみられなかった。1変数のみにクラスタリング効果がある場合には、危険率への影響はみられなかったが、クラスタ数が少ないときに変数の級内相関係数が高くなると検定力も高くなるという関係が見られた。2変数ともクラスタリング効果がある場合は、条件によっては級内相関係数の値が大きくなると危険率も大きく上昇し、不当に有意になりやすい「誤った検定」が行われることがわかった。それに伴い検定力にも複雑な影響を及ぼすことが示された。

2変数ともクラスタリング効果がある場合において、実際の危険率が有意水準として設定した値を超えないように補正した検定を行うと、特にクラスタ数が少ないときに級内相関係数が大きくなるにしたがって検定力が低下することがわかった。

第5章「反復測定の実験データにおける処理効果の検定」では、2回以上の反復測定を行った実験データにおいて処理効果の検定を行うために提案された、潜在曲線モデルを用いた検定の手法を、初期状態において等質でない2群間で処理効果の比較を行う実験データに適用する方法を取り上げた。

まず、準実験データの分析の際に、実験データの場合と同様に「初期状態因子の平均・分散が群間で等しい」という制約を課したモデルで検定を行ったときの影響を調べた。その結果、初期状態の平均値に関する誤った制約を課すと、危険率、検定力ともに真の初期状態の差異による影響を受け、正しい検定が行われないことがわかった。それに対して、初期状態の分散に関する誤った制約に対しては概ね頑健であることがわかった。

そこで、初期状態因子に関する制約を置かないモデルを適用したところ、初期状態における群間の差異にかかわらず処理効果の検定が正しく行われるようになることが示された。つまり、潜在曲線モデルを用いた処理効果の検定は準実験データにおいても有効な手法であることがわかった。

第6章「潜在曲線モデルを用いた事前・事後データの検定における効果量の指標」では、潜在曲線モデルを用いた処理効果の検定において、先行研究および本論文の第5章で採用された効果量の指標が、実際に検定力分析を行って標本の大きさを決める際にはあまり望ましくない性質を持っていることを指摘し、事前・事後データ（2時点の反復測定データ）において望ましい性質を持つ効果量の提案を行った。

他の効果量の指標と同様に、対立仮説のもとで検定統計量が従う分布の非心度パラメータから標本の大きさに関する要素を取り除くという方針により、潜在曲線モデルにおけるパラメータを用いて非心度パラメータの展開を行った。その結果、効果量の指標が共分散分析で用いられる効果量の指標と同一の形式で表せることが示された。そこで、同一効果量を持つ事前・事後データに潜在曲線モデルを当てはめた検定を行った場合と、従来から用いられている共分散分析を適用した場合の検定力を比較した。その結果、共分散分析モデルの前提条件が満たされる場合には、共分散分析を適用した方が高い検定力を確保できることが示された。

第7章「まとめと今後の展望」では、本論文のまとめと今後の展望を述べた。本論文では、統計的検定において検定力に影響を及ぼす要因として、標本の大きさの決定、連続変数のカテゴリ化、2段抽出によるデータ収集、反復測定による群間比較という、実際の研究場面におけるデータ収集および分析の手続きに焦点を当てて検討した。本研究の今後の発展の方向として、2段抽出データの無相関検定に関する数理的なメカニズムの解明や、3時点以上の反復測定データに関する効果量の指標の開発、そして、近年注目されている多水準モデルにおける検定力の検討などが考えられることを述べた。