

科学的研究や意思決定において、テスト情報が重要な役割を果たすことは論を俟たない。しかし、物理実験によって得られたデータとは異なり、心理学・教育学研究におけるテスト情報は、物理的実体を持たない構成概念を計測したものであることが多く、その測定精度は必ずしも明確ではないし、そもそも測定対象がいかなる因子から構成されるのかについても一意の解があるわけではない。従って、テスト情報を利用するためには、その妥当性、すなわち計測対象と理論、測定方法との間の整合性が保証されていることが必要である。妥当性には様々な要素があるが、その中でも予測値と実測値との整合性は最も重要なものの一つである。本論文は、テスト情報の妥当性に関する研究をまとめたものである。特に、テストの素データに適切な統計学的変換を施すことによって、テスト情報の妥当性を高めるための統計的手法をベイズ的アプローチによって開発し、その効果を検証することに焦点が当てられている。

本論文では問題をテスト妥当性の評価基準となる変数が

- 1 1変数である場合、
- 2 複数の変数からなる場合、
- 3 直接観測することが出来ない場合、

の三つのケースに分け、それぞれについて新しい統計手法を開発・提案している。論文は五つの章からなる。上記の各ケースは第2章から第4章までの三つの章で扱われ、これが本論文の中核部分である。また、第1章では序論として、教育測定分野における妥当性の定義や分類についての既存研究が紹介されている。特に、内容妥当性、基準関連妥当性、構成概念妥当性が議論されている。

第2章では、複数の予測変数と1つ基準変数が観測される場合におけるテストの妥当性評価の問題が扱われている。具体的には、重回帰モデルのリッジ回帰推定法の改良に関する提案をおこなっている。周知の通り、重回帰モデルの回帰係数の推定において、最小2乗推定量 (least squares estimator, LSE) は線形不偏推定量のクラスの中で最小の分散を持つ (ガウス=マルコフの定理)。ところが、心理・教育データでは予測変数の間に強い相関が存在することが多く、それによって予測変数の積和行列が特異行列に近くなり、しばしばLSEが数値的に不安定となる。これを多重共線性問題という。この問題への対処として広く用いられている手法の一つがリッジ推定法である。これは積和行列に適切な正値定符号行列を加えることによって、LSEの不安定性を緩和する方法である。この方法によって得られる推定量 (リッジ推定量) は、リッジ定数 (正則化係数) が適切に選択されれば、平均2乗誤差の意味でLSEを優越することが知られている。しかし、そのようなリッジ定数の範囲は一般に未知母数に依存するため、未知母数に関する事前情報がない限り、実際の選択は容易ではない。また、リッジ推定量は線形バイアス推定量であり、この推定量のクラスそのものに興味がないという場合も少なくない。他方、未知母数に関する事前情報が利用出来る場合には、事前分布を適切に選ぶことにより、ベイズ推定量が利用可能である。ベイズ推定量はベイズリスク (平均2乗誤差を事前分布で平均したもの) を最小にする推定量であり、緩やかな条件の下で、他の如何なる推定量にも優越されないという性質

(許容性、admissibility)を有する。また、ベイズ推定量のクラスはリッジ推定量のそれに比べてはるかに広く、事前情報を取り入れる際の柔軟性に優れている。

本論文では回帰係数の事前分布として、有限個の正規分布の尺度混合分布を含むような階層ベイズモデルを考案し、この事前分布に対するベイズ推定量を用いることを提案し、その計算法も導出し、これをベイズ的リッジモデルと呼んでいる。このベイズ的リッジモデルは、先行研究であるLindley and Smith (1970)で提案された階層ベイズモデルの一般化とみることが出来、事前情報を利用する上でより柔軟性を持つものである。本研究ではシミュレーションによってベイズ的リッジモデルがLSEと階層ベイズモデルよりも推定精度において優れていることを明らかにしている。また、事例研究として、中高年被験者の認知テスト結果を予測変数とし、被験者の年収を基準変数とする重回帰モデルを分析し、その妥当性を検証している。

第2章の後半では、適応的リッジ推定量の推定精度を検証している。適応的リッジ推定量とは、リッジ推定量のリッジ定数をデータに基づいて選択するというものである。通常のリッジ推定量とは異なり、データベクトルに対して非線形な推定量となるため、その推定精度を理論的に評価することは難しく、LSEを優越するか否かを調べることも容易ではない。その中であって、Wang and Chow (1990)は正規分布の仮定の下で、ある種の適応的リッジ推定量がLSEを優越するための十分条件を与えており、これは大きな貢献と言える。心理・教育データではしばしば正規分布よりも高い頻度で外れ値が発生するため、彼らの結果が非正規分布の下でも依然として成り立つのかという問いは重要である。本論文ではこの問題に取り組み、正規分布のクラスを含むようなより一般的な分布のクラスとして、球対称分布族を取り上げ、この下で適応的リッジ推定量がLSEを優越するための十分条件を導いている。また、適応的リッジ推定量がミニマックスでもあることや、Maruyama and Strawderman (2005)で議論された一般ベイズ推定量との関数形の類似も指摘している。事例研究では、認知テスト結果を予測変数とし、被験者の学歴を基準変数とするモデルについての妥当性の検証が行われており、適応的リッジ推定量と一般ベイズ推定量は同等の妥当性、汎化性を持つことが結論付けられている。

第3章では、複数の予測変数と複数の基準変数が観測される場合でのテストの妥当性評価について考察されている。具体的には、正準相関分析モデルの改良に関する2つの研究結果がまとめられている。前半部では、予測変数群間または基準変数群間に強い相関が存在している場合、一般化固有値問題の解の値が不安定となる問題を扱う。この論文ではこれを正準相関分析における多重共線性問題と呼んでいる。その際、先行研究としてKlami and Kaski (2007)によるベイズ的正準相関分析モデルに注目する。彼らのモデルは、予測変数と基準変数の背後に共通の潜在変数を仮定した正準相関分析モデルに、階層的な事前分布を追加したものである。本研究では、彼らの用いた事前分布の中で、潜在変数の寄与の度合いを定めるベクトルに対する部分が制約的であることを指摘し、この点を改良した、より一般的なベイズ的正準リッジモデルを提案している。このモデルは従来のKlami and Kaskiのものに比べて変数群内の変動構造により一般的な構造をあてはめることが出来るという特徴を持つ。本研究はこの拡張モデルの下でのベイズ解の計算法を導出し、同時にシミュレーションにより、古典的正準相関モデル、従来のベイズ的正準相関分析モデルとの比較を行っている。その結果、提案モデルが、他の二つよりも推定精度において優れていること、推定精度が群内相関の強弱に大きく左右されない頑健なものであることを確認している。事例研究においても、中高年被験者の認知テスト結果を予測変数とし、家族や

友人、配偶者との結びつきに関する指標を基準変数とする、正準相関分析モデルが推定され、妥当性が検証されている。

後半部では、正準相関と因子分析の統合モデルが提案されている。このモデルは、予測変数と基準変数がそれぞれ因子分析モデルの構造を持つだけでなく、(前半部と同様に) 共通の潜在変数も有するというものである。各変数群内の変動を因子分析モデル部分に吸収させ、正準相関分析部分に波及させないという狙いである。モデルの構造は数学的には前半部のそれに近いものであり、実際、シミュレーションも事例研究においても前半部と同様に良好な結果が得られている。

第4章では、予測変数が複数存在し、かつ基準変数が観測されない場合におけるテストの妥当性評価の問題が扱われている。特に、局所独立性を持たないテストレットに対する項目反応理論 (item response theory, IRT) モデルの改良についての貢献がまとめられている。ここで、テストレットが局所独立性を持つとは、複数の項目への反応ベクトルが観測される時、ある潜在変数を所与としたときに各反応が互いに条件付き独立となることである。例えば、大問形式のテストなどで、各小問を解くことが次の小問を解く際のヒントとなるようなものは局所独立性を持たないと考えられる。局所独立性が成り立たないときにIRTモデルを適用すると、一部の母数が過大推定されることが知られている。この問題への対処として、先行研究では同一テストレット内の潜在得点の間に相関構造を仮定したIRTモデルが提案されている。本研究では、先行研究において相関係数を一様とするモデルが中心であることを指摘し、近接する問題や科目については高い相関を持ち、近接しないものに対しては低い相関を持つような構造を持ったモデルが重要であるとする。そしてそのようなモデルとして、潜在得点の相関構造に自己回帰共分散構造 (Toeplitz行列) をあてはめたベイズモデルを提案している。更に、シミュレーションによって、提案モデルを、局所独立性を仮定したIRTモデルや各種のテストレットIRTモデルと比較し、テストレット群内の相関が無視できない場合も提案モデルが精度の高い推定を行うことを確認している。事例研究として、北アイルランドの学力テストデータ、及び日本の国立大学法人の教育研究評価データに適用され、妥当性についての検討もなされている。

最終章である第5章では全体のまとめと今後の課題について述べられている。

本論文の内容は以上の通りである。本論文では、ベイズ階層モデルに基づいて、様々な状況の下で、素データをより有効な情報源に変換するための多変量解析的手法が、統一的視点から提案されており、その意義は大きい。また、リッジ推定、正準相関分析、項目反応理論などといった、既に広く応用され、膨大な研究蓄積を持つ分野に敢えて挑んでいる点も注目に値する。先行研究を丹念に調べ、着実に前進させてゆく研究姿勢が窺われる。ベイズ階層モデルは数値解析の発展なくしては理論的価値を認められても実践的な技法の開発が困難であったが、本研究では理論的精密さだけではなく、実践的価値を志向している点も評価できる。実際、各章で展開されている事例研究は価値の高いものである。

このように、本論文は、基準変数の数に基づき複数の統計手法を体系的かつ網羅的に扱っており、理論的に興味深い進展を含むと同時に、教育測定分野におけるテスト妥当性研究への貢献をしていると言える。本審査委員会は博士(学術)の学位を授与するにふさわしいものと認定した。