

# 論文の内容の要旨

論文題目 テスト情報評価モデルの妥当性に関する研究

氏名 森 一 将

テスト情報は、個人や組織の意思決定のために重要な役割を果たす。テスト情報が的確に意思決定に寄与するためには、そのテスト情報が妥当でなければならない。ここで述べている妥当性とはテストの素データより加工された変数（予測変数）が年収、学歴、社会での成功といった外的変数（基準変数）を整合的に説明する度合いのことを示している。妥当性の高いテスト情報を得るためには、テストの作成、実施などの諸手順の改善と共に的確な評価モデルが必要とされる。

そこで、本研究では、テストの素データが得られたことを前提とし、その素データを変換する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開発し、その効果を検証することを目的とする。

本研究は以下の3つの下位研究に分かれる。

## (1) テスト妥当性の評価基準が1変数で観測される場合

将来の年収等の外的な基準によって学力テストの妥当性を評価する場合を考える。このとき、学力テストの得点を予測変数とし、回帰分析によってテスト得点の情報をまとめることが可能となる。ところが予測変数群間に高い相関が生じているときには前述の多重共線性の問題が発生する。従来よりこの問題に対する解決としてはリッジ回帰分析による推定の安定化という方法が用いられてきた。(1)ではこのリッジ回帰分析の拡張を提案する。

## (2) テスト妥当性の評価基準が変数群で観測される場合

複数の下位尺度を持つ知能テストを予測変数とし、基準が学力テストであり、学力が複数で観測される場合を考える。このとき、正準相関分析と呼ばれる統計手法を用い、テスト変数群と基準変数群の相関関係を評価する。この場合もテスト変数群、基準変数群の群内で高い相関が生じている場合は多重共線性により係数の推定値が安定

しない。(2)では階層ベイズを用いたより頑健な正準相関分析モデルを提案する。

### (3) テストの妥当性の基準が観測されない場合

実際の学力テストにおいては、受験者の学力の真値が得られない場合が多い。このようにときに利用される統計手法の1つが項目反応理論 (Item Response Theory;IRT) である。IRTモデルではテスト問題の項目反応 (正答・誤答) を予測変数群として考え、潜在変数を基準変数と捉え推定する。この潜在変数は局所独立という統計的な相関に関する仮定の下で推定されている。ところがこの仮定は大問形式テストに代表されるテストレットと呼ばれる形式の下では成立せず、母数や潜在変数の推定が正しく行われなことが知られている。ここではこのテストレットに適応するIRTモデルをベイズ的に拡張し、より真値に近い推定を行うモデルを提案する。

主な結果は以下の通りである。

1. 予測変数が複数、かつ基準変数が1つの場合における回帰分析モデルの適用を検討した。最初に、混合分布を事前分布に設定し、従来の階層ベイズを用いたリッジ回帰分析モデルを拡張したベイズ的リッジモデルを森、繁樹 (2007)に基づき提案した。拡張されたモデルは従来のモデルと比較し、事前情報の取り込みにおいてより高い柔軟性を持つ。ベイズ的リッジモデルはシミュレーション研究により、多重共線性を持つデータに対し、従来の統計手法である最小2乗推定 (LSE)、階層ベイズモデルよりより推定精度が高いことが検証された。事例分析においては、心理学分野で一般的に利用される MIDUS II データセットを用い、認知テスト変数群-収入変数の関係を分析した。その結果、ベイズ的リッジモデルは従来の統計手法よりもより妥当性の高い推定を行うことが検証され、提案手法の有効性が確認された。
2. 回帰分析モデルの適用において、テストデータが非正規である状況を想定し、正規分布より一般的な確率分布クラスである球面对称分布下におけるリッジ推定量を扱った。まず、Maruyama and Strawserman(2005)で提案された一般化ベイズリッジ推定量とミニマクスを満たす条件について概観した。次に、この推定量の設定する定数が多く、テストデータの解析において最適な定数設定を行うことが難しいことを指摘し、より定数設定が易しい球面对称分布下における適応的リッジ推定量を Mori and Kurata (2010)に基づき検討した。加えて、一般化リッジ推定量と適応的リッジ推定量が関数形の近い類似したものであることを数理的に示した。事例研究では MIDUS II データセットを用い、認知テスト変数群-学歴変数の関係について、LSE、一般化ベイズリッジ推定量、適応的リッジ推定量の比較を行った。その結果、一般化リッジ推定量 (一般化ベイズリッジ推定量、適応的リッジ推定量)は LSE と比較してよりテスト情報の

妥当性を高める統計手法であること、テストデータの適用時において適応的リッジ推定量は一般化ベイズリッジ推定量と同等の汎化性を持つことが確認された。

3. 予測変数が複数、かつ基準変数が複数の場合における正準相関分析モデルの適用を検討した。まず、ベイズ的正準相関分析モデル (Klami and Kaski, 2007) を拡張したベイズ的正準リッジモデルを提案した。この拡張モデルは従来のモデルと比べてより変数群内の変動に柔軟なモデル構造を許容するという特徴を持つ。この拡張モデルはシミュレーション研究により、古典的正準相関モデル、ベイズ的正準相関分析モデルより精度の高い推定を行うことが検証された。事例研究においては MIDUS II データセットを用い、認知テスト変数群・関連者変数群の関係が分析された。その結果、ベイズ的正準リッジモデルはベイズ的正準相関モデルと合わせて妥当性の高い推定を行う統計手法であることが確認された。これらの結果より、ベイズ的正準リッジモデルは多重共線性の発生下で精度の高い推定を行い、かつテスト情報の妥当性を高める統計手法であるといえる。
4. 同様に正準相関分析モデルの適用において、正準相関分析と因子分析の統合モデルを提案した。このモデルは変数ごとの変動要因を独自の潜在変数で説明し、その影響を正準相関分析部分に波及させないという特徴を持つ。このモデルは、シミュレーション研究により、古典的正準相関モデルより精度の高い推定を行うことが検証された。事例研究においては MIDUS II データセットを用い、認知テスト変数群・社会生活変数群の関係が分析された。その結果、正準相関分析と因子分析の統合モデルは古典的正準相関モデルと比較し妥当性を高める統計手法であることが確認された。
5. 予測変数が複数、かつ基準変数が観測されない場合における項目反応理論 (IRT) の適用において問題になるのは局所独立性の仮定である。ここでは、テストレット群内の各項目間の背後に無視できない相関が存在する大問形式のテストにおけるデータ構造を想定し、これに対する母数推定を適切に行うモデルを提案した。提案されたモデルは局所独立性を仮定した IRT モデル、及び従来のテストレット IRT モデルとシミュレーション研究により比較評価され、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が検証された。また提案モデルは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、従来研究におけるテストレットを仮定しない IRT モデルと比較して、妥当性を高める統計手法であることが確認された。

これらの結果から本研究で開発された統計手法はテスト情報の妥当性を高めており、有効なものであると結論付けることが出来る。