

博士(工学) 鈴川晶夫

学位論文題名

多変量解析における変量選択に関する研究

学位論文内容の要旨

多変量解析の各手法は複数の変量間の相関構造を考慮して分析を行なうことができるという意味において非常に有用なデータ解析法ではあるが、そこで扱われる変量の数が多いときには次のような問題が生じる。データ収集に対する負担が大きい。分析結果の解釈が難しい。統計的推測を行なう際の未知パラメータ数が多くなり、推測の効率が悪化する。重回帰分析を例にとってこれらの問題点を述べると、多くの説明変量を用いることによって、得られる回帰式のあてはまりが良くなる(すなわち、重相関が大きくなる)ことは期待できるが、その一方では多くの変量についてデータを収集する必要が生じると共に得られる回帰式のもつ意味についての解釈が難しくなる。また、未知パラメータの推定という観点からは、標本数を固定したとき、説明変量の個数の増加により回帰係数、誤差分散の推定精度が悪化するという問題が生じる。したがって、重回帰分析においては少數個の説明変量によって十分大きな重相関が得られるならば、そのような説明変量を用いるべきであると考えられる。

多変量解析の各手法が上記の問題点をもつことから、各手法において有効な変量を選択すること(変量選択問題)は非常に重要な問題である。そこで、多変量解析の各手法における変量選択問題を仮説検定問題として定式化し、その検定問題に対する具体的な検定方式を与えることによって、その分析手法を用いてデータ解析を行なうにあたっての有効な変量を選択することが可能となる。しかし、このような立場から変量選択問題を論じた文献のはほとんどは、変量選択問題に関する仮説に対する尤度比検定方式を与えるものであり、その性質(検出力、許容性等)やその他の合理的と考えられる検定方式との比較については十分に検討されていない。

以上のような背景から、本論文では、多変量回帰分析、正準判別分析、正準相関分析における変量選択問題に関する仮説検定問題に対して、尤度比検定方式およびその他の合理的と考えられる検定方式について考察し、それらの検定方式に基づく効率的な変量選択法を提案した。また、通常の(二組の変量に対する)正準相関分析を変量の組が二組以上の正準相関分析に拡張し、その分析における有効正準変量の選択法を与えた。

本論文では変量選択を行う目的を明確にするために、変量選択問題を次のように二つに分類したうえで、この分類に基づいて各手法における変量選択法について論じた。分類(1):考慮されている変量の中に冗長的変量が含まれているかどうかを検証する問題。分類(2):正準判別分析や正準相関分析のような次元縮小を目的とする正準分析において、有効な正準変量の個数を決定する問題。上で述べた重回帰分析における説明変量の選択問題は分類(1)に属す問題である。分類(2)に属す問題の例として、正準相関分析における有効正準変量の選択問題、すなわち第何番目までの正準変量を有効とみなすかを決定する問題があげられる。しかし、分析手法が与えられるとこの分類が一意に定まるものではない。例えば正準相関分析

においても分類(1)に属す問題を考える必要がある。正準相関分析において考慮されている変量の中に、全ての正準変量に対して影響をもたない変量が存在するならばその変量は正準相関分析において冗長である考えることができ、このような意味で冗長な変量が含まれているかどうかを検証する問題は分類(1)に属す。このような分類に従えば、分類(1)に属す変量選択問題は変量の冗長性を定義したとき、考慮されている変量の一部が冗長であるという仮説の検定問題として定式化することが可能であり、分類(2)に属す変量選択問題については、正準変量の有効性を定義したうえで、ある与えられた非負整数 k に対して有効正準変量の個数が k であるという仮説の検定問題として定式化することができる。

本論文は以下の6章から構成されている。

1章では、本研究の背景と多変量解析における変量選択に関する現状について概説し、本研究の目的について述べた。

2章では、重回帰分析、多変量回帰分析における説明変量の冗長性仮説の検定問題(分類(1)に属す問題)は多変量線形モデルにおける期待値行列に対する線形仮説検定問題と考えることができることから、この検定問題について既に知られている結果を詳述した。また、この検定問題に対する代表的な検定統計量の凸結合に基づく検定方式を考え、その検出力、許容性、不偏性について考察した。

3章では、正準判別分析における有効判別関数の個数に関する検定問題(分類(2)に属す問題)を考える上で基本的な問題である多変量線形モデルにおける期待値行列のランクに関する仮説検定問題について議論した。この検定問題に対する検定方式として提案されている三通りの検定方式について、検出力、漸近効率の二つの観点から考察することによって、それらの検定方式のもつ特徴を明確にした。

4章では、正準判別分析モデルは多変量線形モデルの特別な場合であることから、3章で得られた結果をもとに、正準判別分析における有効正準判別変量の効率的な選択法を提案した。

5章では、正準相関分析における変量選択問題について議論した。二組の変量間の相関構造の簡略化を目的とする正準相関分析においては、分類(1)に属す変量選択問題と分類(2)に属す変量選択問題の両方について考える必要がある。そこで、これら各々の問題に対する変量選択法を詳述し、それらの性質を考察した。また、通常の(変量の組が二組の)正準相関分析を多群の(変量の組が二組以上の)正準相関分析に拡張し、その分析における有効正準変量の個数に関する仮説検定問題に対して、尤度比基準を導出し、その帰無分布の漸近展開を与えた。

6章では、結論として、正準相関分析における変量選択問題や正準判別分析における有効判別変量の選択問題へのアプローチに対する指針を与えたこと、通常の正準相関分析を多群の正準相関分析に拡張したうえで有効正準変量の選択法を与えたことなど、本研究の有効性を述べた。また、モデルに対して必ずしも正規性が仮定できない場合に対する本研究の適用可能性など、本研究において残された問題について論じた。

学位論文審査の要旨

主査 教授 佐藤 義治
副査 教授 新保 勝
副査 教授 伊達 悅

学位論文題名

多変量解析における変量選択に関する研究

多変量解析の有効性は、複数の変量を同時に扱うことにより、それらの相関構造を考慮したデータ解析を行なうことが可能であることである。しかし、そこで扱われる変量の数が多くなるにつれて、統計的推測の効率の悪化、データ収集に関する負担の増大、さらには、分析結果の解釈が困難となるという問題が生じてくる。したがって、多変量解析の各手法において、その分析目的に応じて冗長な変量を取り除き有効な変量を選択すること(変量選択問題)は重要な問題である。

多変量解析における変量選択問題へのアプローチとしては、AIC等の適合度と信頼度のトレード・オフを加味したモデル選択の立場と仮説検定に基づいて変量の冗長性を検証する立場とがある。本論文は後者の立場から変量選択問題を論じている。それは各分析手法の目的に応じて変量の冗長性を定義し、変量の冗長性仮説を定式化し、その仮説を検定することにより変量の冗長性を検証するという立場である。このような立場から変量選択問題を論じた従来の研究の多くは、冗長性仮説に対する尤度比検定方式を与えるものであり、その性質やその他の合理的と考えられる検定方式との比較については十分に検討されていない。

本論文はこのような背景に基づき、多変量解析における変量の冗長性検定問題に対する尤度比検定およびその他の合理的と考えられる検定方式についてそれらの性質を様々な視点から考察し、その結果に基づいて効率的な変量選択法を提案している。

本論文では、多変量回帰分析や正準判別分析における変量の冗長性検定問題が多変量線形モデルにおける線形仮説検定問題または次元検定問題として論じられることを示し、これらの問題に対するいくつかの具体的な検定方式について検出力や漸近効率等の観点からそれらの特徴を明確にすることによって、多変量回帰分析や正準判別分析における効率的な変量選択法を与えていた。また正準相関分析における観測変量および正準変量の選択問題に関して、これらの問題が分割された共分散行列の非対角ブロック行列のランクに関する検定問題として論じられることを示したうえで、この検定問題に関する考察結果に基づいて正準相関分析における効率的な変量選択法を提案している。

また、本論文において、二組の変量に対して適用される通常の正準相関分析を二組以上の変量に対する正準相関分析に拡張し、その分析における有効正準変量の選

択法を与えていた。

これを要するに、著者は、多变量解析手法を用いたデータ解析における变量選択問題に対する数多くの新知見を得ており、情報解析学の発展に寄与するところ大なるものがある。

よって著者は、北海道大学博士(工学)の学位を授与される資格あるものと認める。