

線形回帰における観測値の影響力評価

竹内 秀一

論文の内容の要旨

本論文では、回帰分析の代表的な研究課題の1つである影響力評価について論ずる。影響力評価は、線形回帰において重要となる推定値や予測値あるいは残差などのさまざまな指標に、観測値が及ぼす影響力を調べることを中心課題とする。また、推定法などに含まれるパラメータ値の変化が線形回帰の推定値等へ及ぼす影響力についても問題とする。

影響力評価は、実際にデータ解析を行う際の必要不可欠な手段であるにもかかわらず、理論面あるいは適用段階における問題点が数多く残されている。本論文では、こうした問題点を解明し、影響力評価を実際に適用するために必要となる実用的対応策や具体的指針を提案し、それらの評価を行う。

第1章では、従来の研究および本論文の目的・構成について述べる。

第2章では、最小2乗推定における観測値の影響力評価を行うために、代表的な診断統計量であるCookの距離を中心に新しい表現を提案する。従来は個々の観測値の影響力評価(single-case)が主流であったが、複数の観測値の影響力評価(multiple-case)のための診断統計量を取り上げ、それらのsingle-caseと結び付けた新しい表現を与え、それによりmultiple-caseの重要な問題であるマスク効果を検出することが容易になることを示す。

第3章では、最小2乗推定に替わるリッジ推定における観測値の影響力評価を取り上げ、Cookの距離を中心とした診断統計量の新しい表現を与える。これらの新表現から、影響力の原因となる要素を詳細に探ることが可能になることを明らかにする。また、リッジパラメータの変動による影響力を調べるため、この新表現に基づき補助的な診断指標を導き、その有効性を例示する。

第4章では、変数選択問題における観測値の影響力評価を取り上げ、従来の研究とは異なる視点に立ち、予測平均2乗誤差にたいする影響力を測る新たな診断統計量を提案し、数値実験等によりその有効性を確認する。その診断統計量とCookの距離やCp統計量との関係も明らかにする。

第5章では、multiple-caseで用いられるCookの距離を、残差の意味での外れ値を検出するための検定統計量として利用するときの検出力を問題とし、実データ等を用いてCookの距離の検定統計量としての不自然さおよび不安定さを明らかにする。

第6章では、多変量単回帰モデルでの線形校正問題において、古典的推定量および逆回帰推定量の2つを折衷した一般化逆回帰推定量のクラスを提案し、2つの推定量よりも優れた推定量が存在することを示し、さらに、推定量のパラメータの平均2乗誤差への影響力評価を論じる。

最後に、第7章において線形回帰における観測値の影響力評価について総括する。

以上