

整合性回帰における回帰データの選択について

愛知学院大学経営学部 田中浩光

1.はじめに

最小二乗法に基づく推定回帰式には、解釈上の問題が指摘されている。説明変数値間の相関係数の影響を受ける所謂多重共線性問題(あるいは準共線性問題)が代表的である。技法上の対策として、統計的性能の改良を意図する主成分回帰・リッジ回帰の縮小回帰法など多様な接近が提示されているが、所与の回帰データに対し必ずしも整合的な解釈を与えるものでない。一方、解釈上の問題に留意して、回帰標本上の整合性を確保することを第1として、その上で回帰性を得る推定回帰式の構成が考えられる(田中(2014)など)。

本報告では、説明変数が2個である場合に限定して、整合性を確保する推定方式の構成において基盤となるデータ集合に着目する。推定回帰式が有用となる回帰標本を探るため、データ集合の選択と、回帰標本の崩れの原因(たとえば、異常値を有する観測個体の混入、関連の問題を有する説明変数集合など)の関わりについて、数値実験を通して考察・吟味する。

2.推定回帰式の構成と回帰データの選択

以下、高い相関を示す回帰標本の場合を取り上げる。全観測個体のうち特定の個体集合に注目して、推定回帰式を構成する例について示す。

A.推定回帰式の構成

A1.データ集合の生成：全個体(n)から k (≥ 1) 個の観測個体を順次除いた観測値集合(回帰データ)を回帰分析の対象とする。

A2.整合性領域の生成：解釈上の整合性を確保する整合性領域には、例えば、悪条件と謂われる準共線性、あるいは **suppression** の各現象が生起する領域の排反を用いる。

A3.最適な回帰データの選択：A1 で生成されたデータ集合に基づく推定回帰式が A2 の整合性領域を満たすとき、このデータ集合を最適な回帰データの候補とする。候補の中から、例えば、SD 基準を満たした上で決定係数が最大となるデータ集合を選択する。

B.回帰データの選択

B1.異常値(個体)を原因とする場合：A1 でのデータ集合の生成では、 $k=1,2$ が主として対象となる。留意点として、はずれ値(個体)が準共線性現象を惹起させることがある。

B2.説明変数間の相関を原因とする場合：とくに、隠れた変数の影響を受けて、高い相関を惹起することがある。A1 でのデータ集合の生成では、比較的大きい k 値が対象となる。

参考文献

1. 田中浩光(2003).回帰分析における **Suppression** と準共線性, 2003 年度統計関連学会連合大会講演報告集.
2. 田中浩光(2014).回帰データと推定回帰式の整合性について, 2014 年度統計関連学会連合大会講演報告集