



慶應義塾大学ビジネス・スクール

回帰分析シリーズ 5

— 回帰分析の制約 —

今までの統計ノートでは、いくつかの仮定の下での回帰分析を紹介してきたが、現実にはこれらの仮定をすべて満たすことはまずあり得ない。したがって回帰分析を用いて現実の世界をモデル化しようとする、これらの仮定のいくつかを緩めたり修正しなければ、誤った結果を導き出すことも十分に考えられる。このノートでは、修士論文などで陥りやすい回帰分析の限界や制約についてまとめている。それらは、誤った回帰モデルの構築についてのスペック・エラー、説明変数間の相関関係についての多重共線性、誤差項の分散不均一性、そして誤差項の自己相関という4つのトピックである。

1：スペック・エラー

これまで、回帰モデルには必要な変数がすべて含まれており、また不必要な変数が一切除外されていることを仮定してきた。しかし現実に回帰モデルを構築する場合、このような仮定が成立しない方がはるかに多い。

たとえば真のモデルが $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ であるとしよう。ここで $E(\mathbf{u}) = \mathbf{0}$ 、また $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$ と仮定すると、 $\boldsymbol{\beta}$ の不偏推定値 \mathbf{b} は、 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ によって求めることが可能であった。しかし、もし分析者が誤ったモデル $\mathbf{y} = \mathbf{X}\cdot\boldsymbol{\beta} + \mathbf{u}$ を構築してしまったならば、 $\mathbf{b}\cdot$ は $(\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{y}$ によって推定されることになる。ここで $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ を代入すると、 $\mathbf{b}\cdot$ は、

$$\begin{aligned}\mathbf{b}\cdot &= (\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{u} \\ &= (\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{X}\boldsymbol{\beta} \quad \therefore E(\mathbf{u}) = \mathbf{0}\end{aligned}\tag{1}$$

となる。これは誤ったモデルによって推定された回帰係数が、もはや真のパラメータではないことを示すものである。このように行列 \mathbf{X} の失敗のことをスペック・エラーという。このようなスペック・エラーには、(1) 必要な変数をモデルから除外してしまったケースと、(2) 不必要な変数を組み込んだケースとが考えられる。ここでは、その2つのケースを解説することにより、スペック・エラーが推定にどのような影響をもたらすか示す。