慶應義塾大学ビジネス・スクール

回帰分析シリーズ3

— 行列 —

これまで、モデルに説明変数が２つ含まれる場合の重回帰分析を考察してきた。しかし単回帰であっても、説明変数が複数個存在する場合であっても、その基本的な仮定や仕組みは同じである。このことを理解するために、このノートでは行列や行列式を用いて回帰分析の仕組みを解説する。これは行列を用いることによって、回帰モデルの一般化が可能になるからである。さらに、多重共線性やスペック・エラー、分散の不均一性、自己相関といった様々な回帰モデルの制約や限界を考察する場合、行列を用いる方が容易に理解することができる。ただし、行列に対する知識が十分でない学生に対する配慮から、重要な証明や難解な計算については、その都度解説を行ってゆくことにする。

一般に k 個の未知数を持つ回帰モデルは、

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots.. + \beta_k X_{ki} + u_i \quad \therefore\ i = 1, 2, \ldots., n \tag{1}$$

もしくは、

$$Y_i = \sum_{g=1}^{k} \Sigma \beta_g X_{gi} + u_i \quad \therefore\ i = 1, 2, \ldots., n;\ X_{1i} = 1 \tag{2}$$

のように示すことができる。しかし行列式を用いると、このモデルは次のように記述することが可能である。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{3}$$

ここで $\mathbf{y}$ と $\mathbf{u}$ は $(n \times 1)$、$\boldsymbol{\beta}$ は $(k \times 1)$ の列ベクトル、$\mathbf{X}$ は $(n \times k)$ の行列である。ただし $\mathbf{X}$ の第１列は、切片 (β_0) の係数を示すことから１になっている。

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},\quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},\quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix},\quad \mathbf{X} = \begin{pmatrix} 1 & X_{12} & \cdots & X_{1k} \\ 1 & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & \cdots & X_{nk} \end{pmatrix}$$

そして単回帰の場合と同様に、誤差項の分布に関していくつかの仮定をおくが、それらの仮定を行列式で表せば、

(1) 誤差項の期待値はゼロである。

このノートは、慶應義塾大学ビジネス・スクールにおける補助教材として、同ビジネス・スクール教授矢作恒雄と博士課程磯辺剛彦が作成した。（1995年5月作成）