慶應義塾大学ビジネス・スクール

回帰分析シリーズ6

— ロジット・モデル —

これまでの回帰分析ノートでの従属変数（被説明変数）には、量的データを採用してきた。このようなデータは、たとえば投資収益率であったり、売上高、消費額などが含まれる。しかし分析目的が二者択一的なもの、たとえば設備を自製すべきか、もしくは購入すべきかといった意思決定であるとすれば、従属変数に量的なデータは使用できない。このような例は、他にも倒産分析、企業内での特定部門の設置などがあり、従属変数には「イエス」か「ノー」、または「成功」か「失敗」を示す質的データを採用しなければならない。従属変数に質的データ（ダミー変数）をもつ分析方法には、おもにロジット、プロビット、トービットがあが、このノートではロジット・モデルにのみ注目する。

ここでは次のような線形単回帰モデルを考えてみる。

$$Y_i = \alpha + \beta X_i + u_i \quad Y_i = 0, 1 \tag{1}$$

ここで従属変数 Y_i は0か1の二元変数 (binary variable) である。(1) 式の期待値をとれば、

$$E(Y_i) = \alpha + \beta X_i \tag{2}$$

であらわすことができる。いま Y_i が Bernoulli 変数であるとすれば、Y_i の確率分布は次のようになる 。

$$\begin{aligned} &Y_i = 1 \qquad P(Y_i = 1) = \pi_i \\ &Y_i = 0 \qquad P(Y_i = 0) = 1 - \pi_i \end{aligned} \tag{3}$$

Bernoulli 変数とは、もしある変数Zが1か0という2つの値をとり、それぞれが発生する確率が p (Z = 1)、$1 - p$ (Z = 1) のとき、変数Zは Bernoulli 分布にしたがうといわれる。Bernoulli 変数の確率密度関数は次のように示される。

$$f(z \mid p) = \begin{cases} p^z(1-p)^{1-z} & x = 0, 1 \\ 0 & \text{その他} \end{cases}$$

たとえば $z_1 = 1,\ z_2 = 1,\ z_3 = 0$ という3つのデータを与えられたとき、このランダムなサンプルの確率は次式によって計算される。

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）