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）

$$f\left(z_1 = 1,\ z_2 = 1,\ z_3 = 0 \right) = \prod_{i=1}^{3} p^{z_i}(1-p)^{1-z_i} = p \cdot p \cdot (1-p)$$

ここで (3) 式の Y_i の期待値は、

$$E\left(Y_i\right) = 1(\pi_i) + 0(1-\pi_i) = \pi_i \tag{4}$$

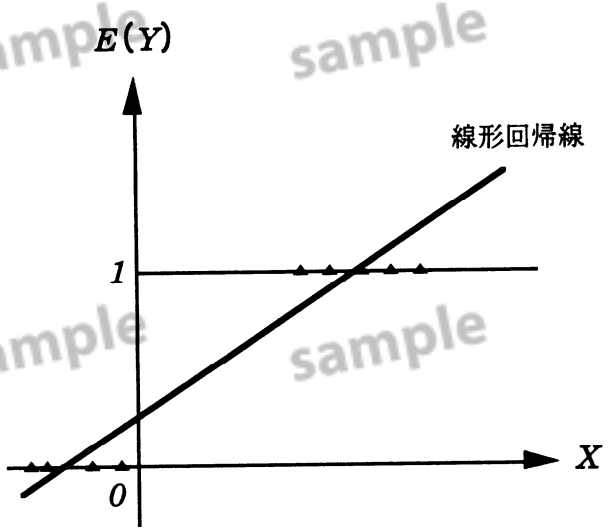
となるので、(2) 式と (4) 式とにより、

$$E\left(Y_i\right) = \alpha + \beta X_i = \pi_i \tag{5}$$

が導かれる。しかしこのような線形回帰モデルのダミー従属変数への適用は、いくつかの無視できない問題を引き起こす。

まず $E\left(Y\right) = \pi$ のとりうる範囲は 0 と 1 との間であるが、線形回帰モデルを適用すると、図 1 のように予測値が $(0, 1)$ の範囲の外側に出てしまうことから予測誤差が大きくなる。

（図 1 ）線形確率モデルによる推定



次に、ダミー従属変数の誤差項は、

$$\begin{aligned} Y_i = 1 &: \quad u_i = 1 - \alpha - \beta X_i \\ Y_i = 0 &: \quad u_i = -\alpha - \beta X_i \end{aligned} \tag{6}$$

となるが、これは誤差項の正規分布というＯＬＳの重要な仮定がもはや適用できないことを示している。

最後に、(6) 式と (5) 式を用いて誤差項の分散を計測すると、

$$
\begin{aligned}
E(u_i^2) &= (1-\alpha-\beta X_i)^2\pi_i + (-\alpha-\beta X_i)^2(1-\pi_i) \\
&= (1-\alpha-\beta X_i)^2(\alpha+\beta X_i) + (-\alpha-\beta X_i)^2(1-\alpha-\beta X_i) \\
&= (\alpha+\beta X_i)(1-\alpha-\beta X_i) \\
&= E(Y_i)[1-E(Y_i)]
\end{aligned} \tag{7}
$$

(7) 式は誤差項の分散が$E(Y_i)$に依存するために、分散の不均一性をもたらすことを意味するものである。したがって以上３つの理由により、ダミー従属変数に線形回帰モデルを適用することが適切でないことが理解できる。

ロジット・モデルの特性

線形確率モデルの不適切性は、非線形モデルの適用を示唆する。直観的には$E(Y_i)$が$(0, 1)$の範囲におさまるようなＳ字型カーブが思いつくであろう。このような非線形カーブの一つに「ロジスティック・カーブ」と呼ばれるものがあり、この曲線を用いたモデルを「ロジット・モデル」という。通常このモデルは、

$$
E(Y_i) = \frac{1}{1+e^{-\alpha-\beta X_i}} = \frac{e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} \tag{8}
$$

によって定義される。(8) 式の両辺を自然対数化すると$E(Y)=\pi$ により、

$$
\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha+\beta X_i \tag{9}
$$

ここで気づくのは、(9) 式の左辺が $Y_i=0$ に対する $Y_i=1$ のオッズ (odd) の対数比であり、またこの対数オッズ比が $\alpha+\beta X_i$ の線形関数になっていることである。したがって、ざまざまな水準の X_i を観察することにより、(9) 式を推定することが可能になる。

しかし実際には観察データが究めて少ない場合の方が多く、(9) 式による推定はもはや不可能になる。そのような問題を解決できる手法が最尤法である。最尤法は個々の Y_i によってロジット・モデルを推定することを可能にする。

もし Y_i が Bernoulli 変数であるならば、(3) 式が導き出された。ここで確率密度関数は、

$$
f(Y_i) = \pi_i^{Y_i}(1-\pi_i)^{1-Y_i} \qquad Y_i = 0, 1\ ; i = 1, 2, \ldots, n \tag{10}
$$

によって表すことができる。ここで $f(1)=\pi_i$ 、 $f(0)=1-\pi_i$ である。つまり $f(Y_i)$ は単純に $Y_i=0$ 、または $Y_i=1$ となる確率を表している。

Y_i は独立なので、同時確率関数 (joint probability function) は、

$$f(Y_1, Y_2, \ldots, Y_n) = \prod_{i=1}^{n} f(Y_i)\, \pi_i^{Y_i} (1-\pi_i)^{1-Y_i} \tag{11}$$

のようになる。最尤法とは、(11) 式を最大化するようなパラメータを求めることであるが、対数変換によって積 (product) から和 (sum) の式になる。ここで (11) 式を対数変換すると、

$$\begin{aligned} \ln f(Y_1, Y_2, \ldots, Y_n) &= \ln \prod_{i=1}^{n} f(Y_i)\, \pi_i^{Y_i} (1-\pi_i)^{1-Y_i} \\ &= \sum_{i=1}^{n} \left[Y_i \ln \left(\frac{\pi_i}{1-\pi_i} \right) \right] + \sum_{i=1}^{n} \ln (1-\pi_i) \end{aligned} \tag{12}$$

すでに $E(Y) = \pi$ であることを示しており、(8) 式の両辺を 1 から引くと、

$$1 - \pi_i = 1 - \frac{e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} = \frac{1}{1+e^{\alpha+\beta X_i}} \tag{13}$$

したがって、

$$\ln (1-\pi_i) = \ln \left(\frac{1}{1+e^{\alpha+\beta X_i}} \right) = [1 + \ln (e^{\alpha+\beta X_i})]^{-1} \tag{14}$$

ここで (12) 式に (9) 式と (14) 式を代入し、 $L(\alpha, \beta)$ を最尤法による推定されるパラメータとすれば、

$$\ln L(\alpha, \beta) = \sum_{i=1}^{n} Y_i (\alpha + \beta X_i) - \sum_{i=1}^{n} \ln [1 + e^{\alpha+\beta X_i}] \tag{15}$$

最尤推定値とは、(15) 式の最尤方程式を最大にするような α と β である。そして最尤方程式を最大化するためには、未知数 α と β に関して(15) 式を偏微分し、その方程式をゼロにすればよい。

$$\begin{aligned} \frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} &= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \frac{e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} \\ \frac{\partial \ln L(\alpha, \beta)}{\partial \beta} &= \sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} \frac{X_i e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} \end{aligned} \tag{16}$$

ここで α と β を推定値 a, b に置き換えて、(16) 式をゼロに設定すると、

$$\begin{aligned} \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \frac{e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} &= 0 \\ \sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} \frac{X_i e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} &= 0 \end{aligned} \tag{17}$$

このような方程式を机上で解くことは究めて困難であるが、現在では統計パッケージによって瞬時に計算してくれる。読者は(17)式を解く方法までは理解する必要はなく、ロジット・モデルがどのような仮定をもっているのか、導き出されたパラメータのもつ意味といった基本的な事項だけを理解すればよい。

(17)式で計算されたβについて解説しよう。すでに(9)式で示したように、ロジット・モデルは回帰係数を通常の線形モデルと同様に理解することはできない。ロジット・モデルの左辺（被説明変数）は、$Y_i = 0$ に対する $Y_i = 1$ のオッズ $[\pi_i/(1-\pi_i)]$ の対数比であるから、(9)式を指数化すると $\pi_i/(1-\pi_i) = e^{\alpha+\beta X_i}$ となる。したがって説明変数 X 1単位の増加は、オッズ比を e^{β} 倍引き上げることになる。たとえば事業の成功が経営者の経験年数に依存するという仮説をロジット・モデルによって検証したとしよう。分析の結果、もし統計パッケージによるβの数値が0.187であれば、$e^{0.187} = 1.206$ により経験年数が1年長くなるにつれ成功の確率が約20%増加することを示すのである。

またロジットモデルの分散は、情報行列を利用することによって求めることが可能である。情報行列 $R(\alpha, \beta)$ とは(15)式の2階の偏微分のことである。

$$R(\alpha, \beta) = -\left[\begin{array}{cc} \dfrac{\partial^2 ln\, L(\alpha,\beta)}{\partial \alpha^2} & \dfrac{\partial^2 ln\, L(\alpha,\beta)}{\partial \alpha\, \partial \beta} \\ \dfrac{\partial^2 ln\, L(\alpha,\beta)}{\partial \beta\, \partial \alpha} & \dfrac{\partial^2 ln\, L(\alpha,\beta)}{\partial \beta^2} \end{array}\right]^{-1}\Bigg|_{\alpha=a,\,\beta=b} \tag{18}$$

ここでαとβの分散は(18)式の対角線上に位置し、それぞれ次のようになる。

$$Var(\alpha) = -\left[\frac{\partial^2 ln\, L(\alpha,\beta)}{\partial \alpha^2}\right]^{-1}\Bigg|_{\alpha=a,\,\beta=b} \tag{19a}$$

$$Var(\beta) = -\left[\frac{\partial^2 ln\, L(\alpha,\beta)}{\partial \beta^2}\right]^{-1}\Bigg|_{\alpha=a,\,\beta=b} \tag{19b}$$

プロビット・モデルとの関係

ロジット・モデルとプロビット・モデルの違いは、ロジット・モデルがロジスティック・カーブを仮定してるのに対して、プロビット・モデルは累積正規分布を前提とする点にある。したがってロジット・モデルとプロビット・モデルは、それぞれ次のようになる。

ロジット・モデル：　$F(\alpha+\beta X_i)=\dfrac{e^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}}$

プロビット・モデル：　$F(\alpha+\beta X_i)=\displaystyle\int_{-\infty}^{\alpha+\beta X_i}\frac{1}{\sqrt{2\pi}}e^{-z_i^2/2}\,dz_i$

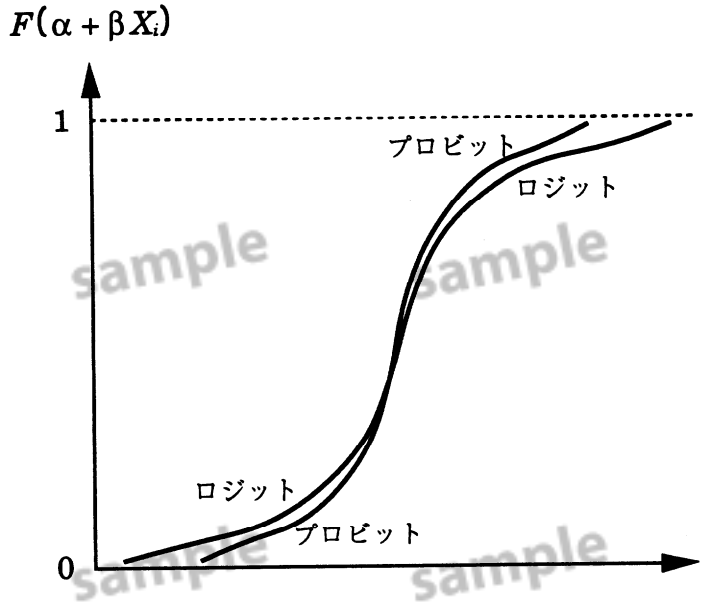
またプロビット・モデルに最尤法を適用すれば、

$$\ln L(\alpha,\beta)=\sum_{i=1}^{n}\{Y_i\ln F(\alpha+\beta X_i)+(1-Y_i)\ln[1-F(\alpha+\beta X_i)]\}\quad(20)$$

となる。

図２はロジット・モデルとプロビット・モデルを比較したものであるが、両モデルは互いに究めて類似しており、データが十分に大きければ、その結果はそれほど違わない。しかしロジット・モデルはプロビット・モデルよりも裾野のデータに敏感である。一般にはロジット・モデルの方が簡単に計算可能で、また係数の意味も理解しやすいことも指摘できる。

（図２）ロジット・モデルとプロビット・モデル



(Kmenta, J. "Elements of Econometrics", 2nd. ed., MacMillan Publishing, 1986; pp.555)

ここで注意すべきことは、ロジット・モデルとプロビット・モデルの回帰係数の意味が異なっており、単純に両者を比較できないということである。ただ Amemiya は、次の変換を行うことにより線形確率モデルとロジット・モデル、プロビット・モ

デルとが比較できると述べている[1)]。

$$a_{LP} \simeq 0.25\, a_L + 0.5 \simeq 0.4\, a_P + 0.5$$
$$b_{LP} \simeq 0.25\, b_L \simeq 0.4\, b_P$$

ここで a_{LP} は線形確率モデルの定数項、a_L はロジット・モデルの定数項、a_P はプロビット・モデルの定数項である。もし分析者が複数のモデルの回帰係数を比較しようとするならば、いずれかのモデルの結果を基準化する必要がある。

例題

次にロジット・モデルを用いて表１の例題を考えてみよう。表１は男性に対する百貨店カードの調査であり、もし百貨店カードを所持していれば１、そうでなければ０である。そしてこの購買行動を説明する変数として、１カ月間に自由になる所得を採用している。

（表１）百貨店の利用調査

カード	所得（千円）	カード	所得（千円）
1	28	1	38
0	48	1	33
0	50	0	57
1	31	0	62
1	28	1	43
0	52	0	48
1	34	0	68
1	25	1	22
0	60	0	39
1	38	1	33
1	21	1	34
1	29	1	52
0	52	0	60
0	72	0	33
0	67	1	54

このデータを線形確率モデル、ロジット・モデル、プロビット・モデルで分析し

---

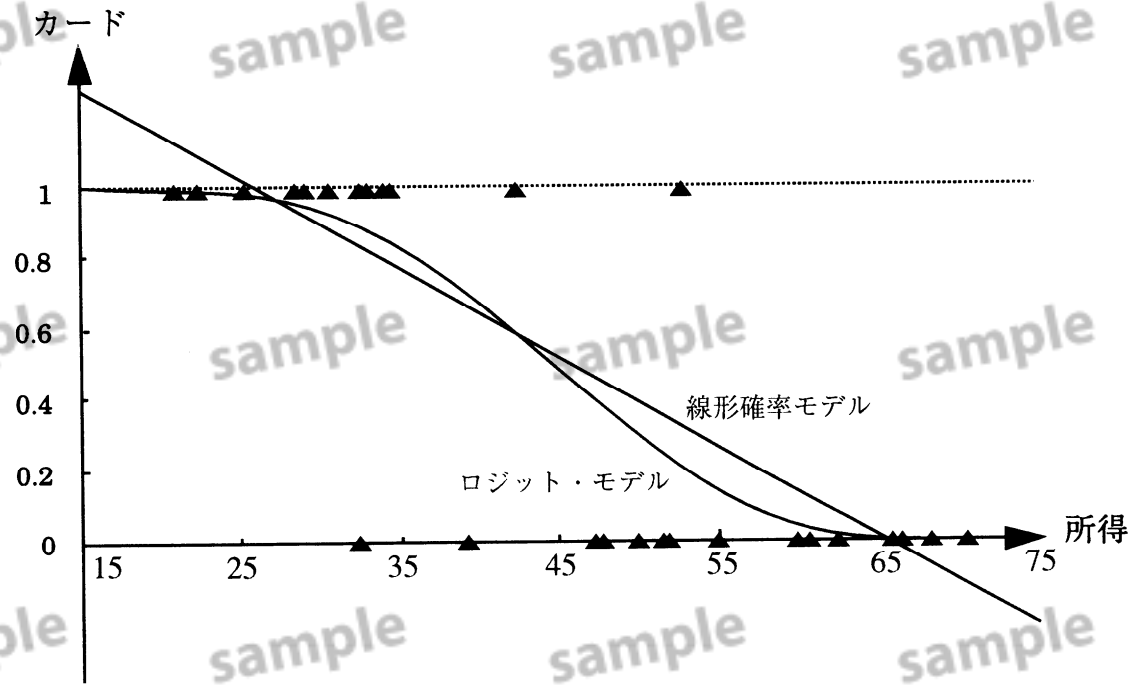
1) Amemiya, T., "Qualitative Response Model: A Survey", *Journal of Economic Literature,* 19, Dec. 1981, pp.1488.

（表２）出力結果

	線形確率モデル	ロジット	プロビット
Current Predictions	—	86.7%	86.7%
R-square	0.528	0.538	0.533
α	**1.63566**	**7.6370**	**4.49776**
(standard error)	0.207	2.517	1.341
(t-value)	7.884	3.034	3.354
β	**-0.02523**	**-0.171167**	**-0.100763**
(standard error)	0.004	0.056	0.030
(t-value)	-5.593	-3.053	-3.366

ロジット・モデルに関しての出力結果の意味は、$e^{-0.171167} = 0.842$ により所得が１単位（千円）増えるにつれて、カードの所有率が 15.8％減少するというものである。さらに図２は、導き出された線形モデルとロジットモデルをグラフ化したものである。線形モデルの場合、裾野のデータの推定値 $\hat{Y}_i$ が０と１の範囲の外側に出てしまっていることが分かる。このことは、もし予測のために線形モデルを適用するのであれば、データによってはその予測（確率）がマイナスであったり１以上であることも生じる。

（図２）線形モデルとロジットモデル



次に、Amemiya の近似値を使って基準化したのが表 3 である。この表から理解されるように、ロジットとプロビットのパラメータが非常に近い数値になっている。つまり両モデルによる推定が、それほど違ったものにならないことを示すものである。

（表 3）ロジットとプロビットの基準化

	線形確率モデル	ロジット	プロビット
α	1.6357	2.4093	2.2991
β	-0.0252	-0.0428	-0.0403