

第 11 章 被説明変数が (0, 1) 変数の回帰分析

James H. Stock 著・Mark W. Watson 著・宮尾 龍蔵 訳
『入門計量経済学』（共立出版、2016）

<https://www.kyoritsu-pub.co.jp/book/b10003746.html>

2026-03-01

1. 被説明変数が $(0, 1)$ 変数の場合

(0, 1) 被説明変数とは

- ▶ 結果が「はい/いいえ」「成功/失敗」のように2つの値しか取らない変数。
- ▶ 例:
 - ▶ 住宅ローンの申請が承認される (0) か、否認される (1) か。
 - ▶ 大学に進学する (1) か、しない (0) か。
 - ▶ 喫煙する (1) か、しない (0) か。
- ▶ このような場合、回帰モデルの目的は、説明変数 X が与えられた下で $Y = 1$ となる確率 $P(Y = 1|X)$ をモデル化することである。

確率としての回帰式

- ▶ 2.2 節で学んだように、(0, 1) 変数 Y の期待値は、 $Y = 1$ となる確率に等しい。

$$E(Y|X) = P(Y = 1|X)$$

- ▶ したがって、(0, 1) 変数を被説明変数とする回帰式は、説明変数の値に基づいた「予測される確率」と解釈される。

2. 線形確率モデル (LPM)

線形確率モデル (Linear Probability Model)

- ▶ 被説明変数が (0, 1) 変数である場合に、通常の変数回帰 (OLS) をそのまま適用するモデル。
- ▶ モデル: $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i$
- ▶ 解釈:
 - ▶ \hat{Y}_i は、 X_i の下で $Y_i = 1$ となる予測確率。
 - ▶ β_1 は、 X_1 が 1 単位変化したときの $Y = 1$ となる確率の変化量。

LPM の長所と短所

▶ 長所:

- ▶ 解釈が非常に単純で直感的。
- ▶ OLS で簡単に推定できる。

▶ 短所:

- ▶ 予測確率が 0 未満や 1 を超えることがある（現実にはそぐわない）。
- ▶ 傾きが常に一定であると仮定しているが、確率は 0 や 1 に近づくほど変化しにくくなるのが一般的。
- ▶ 本質的に不均一分散を伴う。

3. プロビットとロジット回帰

非線形な確率モデルの必要性

- ▶ 予測確率が必ず 0 と 1 の間に収まり、 X の変化に対して S 字型の曲線を描くモデルが必要。
- ▶ **累積分布関数 (CDF)** を用いてモデル化する：

$$P(Y = 1|X) = F(\beta_0 + \beta_1 X_1 + \dots)$$

- ▶ F の選び方によって、プロビットとロジットの 2 つの代表的なモデルがある。

プロビット回帰 (Probit Regression)

- ▶ F として標準正規分布の累積分布関数 (Φ) を用いる。

$$P(Y = 1|X) = \Phi(\beta_0 + \beta_1 X_1 + \dots)$$

- ▶ **解釈:**

- ▶ β_j が正であれば、 X_j の増加は $Y = 1$ となる確率を増加させる。
- ▶ 具体的な確率の変化量は、 X の初期値に依存するため、 Φ を通じて計算する必要がある。

ロジット回帰 (Logit Regression)

- ▶ F としてロジスティック累積分布関数を用いる。

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots)}}$$

- ▶ プロビットと非常によく似た結果を与えることが多く、計算の利便性から広く使われる。

4. 推定と適合度

最尤推定 (Maximum Likelihood Estimation, MLE)

- ▶ プロビットやロジットは非線形モデルであるため、OLS は使えない。
- ▶ **最尤法:** 実際に観測されたデータが発生する確率（尤度）を最大にするようなパラメーター β を推定する方法。
- ▶ 大標本の下で、最尤推定量は不偏・一致であり、正規分布に従う。

適合度の指標

- ▶ 通常の R^2 は (0, 1) 変数の場合には適切ではない。
- ▶ 疑似 R^2 (Pseudo- R^2): 尤度に基づいた指標。
- ▶ **正答率 (Fraction Correctly Predicted)**: 予測確率が 0.5 以上なら $Y = 1$ 、0.5 未満なら $Y = 0$ と予測し、実際のデータとどれだけ一致したかを測る。

5. 結論

住宅ローン差別の分析 (ボストン HMDA データ)

- ▶ 被説明変数：ローンの否認 (1) か承認 (0) か。
- ▶ 説明変数：返済・所得比率、および人種ダミー。
- ▶ 発見:
 - ▶ 返済能力などの経済的要因をコントロールした後でも、黒人であることはローンの否認確率を統計的に有意に高める。
 - ▶ 線形確率モデル、プロビット、ロジットのいずれを用いても、同様の結論が得られる。
- ▶ まとめ：被説明変数が (0, 1) 変数の場合、予測確率を適切に扱えるプロビットやロジットが、より洗練された分析手法となる。