

離散選択モデル入門

LPM, Logit, Probitによる二項選択モデル

北川梨津

1 離散選択モデル

- 離散的な選択肢の中から何かを選ぶこと（何かが選ばれること）を**離散選択**（discrete choice）と呼ぶ。
 - 例：人事考課（ $\{S, A, B, C\}$ 。）
- その統計モデルを**離散選択モデル**（discrete choice model）と呼ぶ。
 - 例： $\Pr(S | X), \Pr(A | X), \Pr(B | X), \Pr(C | X)$ 。
- つまり、従属変数が離散型の変数の回帰モデル。
- 今日は2つの選択肢がある場合限定して議論する。
 - **二項選択モデル**（binary choice model）。

2 二項選択モデル

- 選択肢が2つの例：ハイパフォーマー，昇進，離職など
.
- カテゴリーが2つなので，ダミー変数として扱える.
- $\Pr(Y = 1 \mid X) = 1 - \Pr(Y = 0 \mid X)$ なので，一方の確率を考えるだけでよい.
- すると， $\Pr(Y = 1 \mid X) = f(X \mid \beta)$ という回帰モデル $f(X \mid \beta)$ を推定するという問題ということになる.
- $f(X \mid \beta)$ にどのような関数を仮定するか？

3 主要な3つのモデル

- 線形確率モデル (LPM; linear probability model)
 - $\Pr(Y = 1 | X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$
- ロジット・モデル (logit model)
 - $\Pr(Y = 1 | X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}$
- プロビット・モデル (probit model)
 - $\Pr(Y = 1 | X) = \Phi(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)$

4 線形確率モデル

- $\Pr(Y = 1 \mid X)$ を説明変数の線形結合 $\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$ で近似するモデル.
- 単に従属変数をダミー変数として線形モデルをOLS推定すればよい.

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

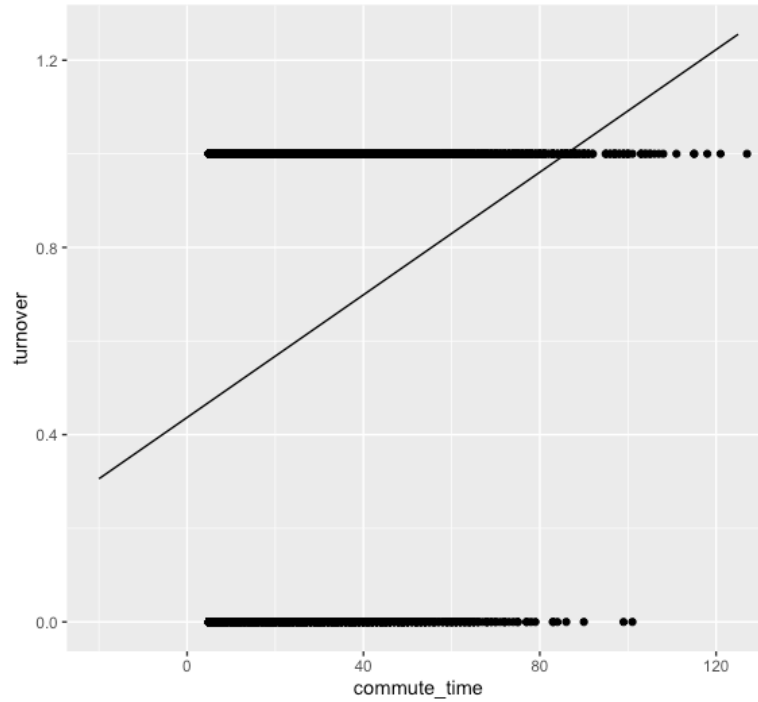
式(1)の両辺の条件付き期待値をとると,

$$E[Y \mid X] = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k.$$

$Y \in \{0,1\}$ なら, $E[Y \mid X] = \Pr(Y = 1 \mid X)$ なので,

$$\Pr(Y = 1 \mid X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \quad (2)$$

5 LPMのイメージ



6 LPMの良いところと悪いところ

- 悪いところ
 - 確率の性質を満たさない。（論理的整合性がない）
 - 必ず不均一分散になる。（頑健な標準誤差を使えばよい）
- 良いところ
 - 解釈しやすい。
 - 固定効果法や操作変数法などもそのまま適用できる。

7 LPMの実装

```
library(modelsummary);
library(dplyr)
mydata <-
  readr::read_csv("turnover.csv")

mydata %>%
  lm(turnover ~ tenure +
    commute_time, data = .) %>%
  modelsummary(vcov = "robust",
    gof_map = c("nobs",
      "r.squared"),
    stars = TRUE)
```

	Model 1
(Intercept)	0.481***
	(0.015)
tenure	-0.003***
	(0.001)
commute_time	0.007***
	(0.000)
Num.Obs.	5000
R2	0.085

8 LPMの解釈

- 推定されたモデル：
- $\widehat{\Pr}(Y = 1 | X) = 0.481 - 0.003 \times tenure + 0.007 \times commute_time$
- 左辺は確率なので、係数の値は説明変数が1単位上昇したときの確率の増加分として解釈できる（セテリスパリブス）。
- 勤続年数が10年増えると、離職確率が0.03減少する。
 - 「離職確率が0.03減少する」＝「離職確率が3%ポイント減少する。」
 - 「離職確率が3%減少する」は間違い。
- ある説明変数の効果の大きさを示すときに、従属変数の平均値と比べて、そこから何%増えたかを計算することがある。
 - $\text{mean}(\text{mydata}\$turnover)$ は0.6296なので、 $\frac{0.5996 - 0.6296}{0.6296} = -0.047$ 、約5%の減少。
- 今回のサンプルデータは違うが、事象がそもそも稀だと係数が小さくても、かなり大きな効果であることがある。

9 ロジット・モデル

- $\Pr(Y = 1 | X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}$ とするモデル.
- 右辺は必ず0と1の間の値をとるので、確率の性質を満たす.
- **最尤法**を使ってパラメータを推定する.
- 最尤法のイメージ：手元のデータが尤もらしくなるようなパラメータの値は何だろう～？
- 尤度： $L(\beta) = \prod_{i=1}^n \left[\Lambda(\beta_0 + \beta_1 X_{1i})^{Y_i} \times (1 - \Lambda(\beta_0 + \beta_1 X_{1i}))^{1-Y_i} \right].$

10 ロジット・モデルの推定

- 尤度を最大化するパラメータの値を求める = 最尤推定 (MLE; maximum likelihood estimation) .

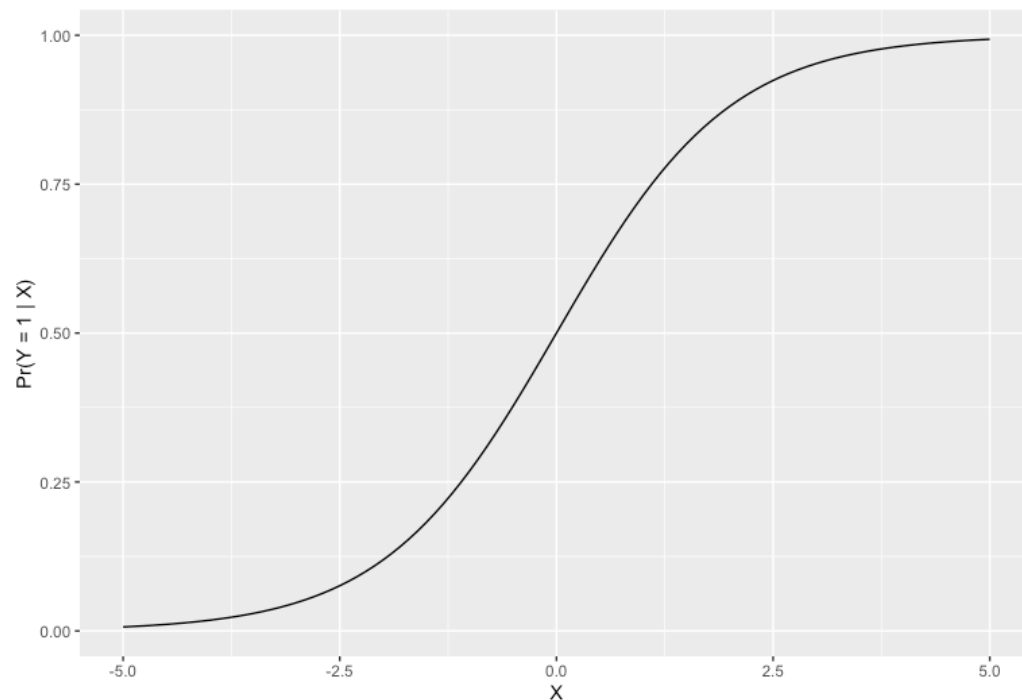
$$\hat{\beta}^{\text{MLE}} = \operatorname{argmax}_{\beta} L(\beta)$$

- 実際には計算しやすくなるため, 対数尤度を最大化する .

$$\hat{\beta}^{\text{MLE}} = \operatorname{argmax}_{\beta} \log(L(\beta))$$

- ニュートン法などの数値計算でこれを求める. (Rがやってくれる)

11 ロジット・モデルのイメージ



12 ロジット・モデルの実装

```
library(modelsummary);
library(dplyr)
mydata <-
  readr::read_csv("turnover.csv")

mydata %>%
  glm(turnover ~ tenure +
    commute_time,
    data = .,
    family = binomial(link =
      "logit")) %>%
  modelsummary(gof_map = c("nobs",
    "r.squared"),
    stars = TRUE)
```

	Model 1
(Intercept)	-0.185**
	(0.068)
tenure	-0.013***
	(0.003)
commute_time	0.034***
	(0.002)
Num.Obs.	5000

13 ロジット・モデルの解釈

- 係数はそのまま効果として解釈することはできない.
- 非線形なモデルなので, 元の X_1, X_2, \dots, X_k の水準によって, 説明変数の効果の大きさも変わる.
- 経済学では, 効果の評価のために**平均限界効果** (AME; average marginal effect) を計算することが多い.

$$AME_{X_1} = \sum_{i=1}^n \frac{\partial \Pr(Y_i = 1 \mid X_i)}{\partial X_{1i}}$$

- ただし, 説明変数がダミー変数の場合は, 微分できないので, 差をとる.

$$AME_{X_1} = \sum_{i=1}^n \Pr(Y_i = 1 \mid X_i, X_{1i} = 1) - \Pr(Y_i = 1 \mid X_i, X_{1i} = 0)$$

14 平均限界効果の求め方

```
library(marginaleffects)
```

```
mydata %>%  
  mutate(female =  
as.factor(female)) %>%  
  glm(turnover ~ tenure +  
commute_time,  
      data = .,  
      family = binomial(link =  
"logit")) %>%  
  marginaleffects() %>%  
  modelsummary(gof_map = c("nobs"),  
              stars = TRUE)
```

	Model 1
tenure	-0.003***
	(0.001)
commute_time	0.007***
	(0.000)
Num.Obs.	5000

15 プロビット・モデル

- $\Pr(Y = 1 \mid X) = \Phi(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k).$
- 他は, ロジット・モデルと同様の議論が成り立つ.
- Rでは, `glm(y ~ x, family = binomial(link = "probit"))`とすればよい.

16 LPM, ロジット, プロビットの比較

LPMの係数, ロジット, プロビットのAMEはだいたい同じ値になることが多い.

	LPM	Logit (AME)	Probit (AME)
(Intercept)	0.48091***		
	(0.01454)		
tenure	-0.00290***	-0.00283***	-0.00286***
	(0.00060)	(0.00059)	(0.00059)
commute_time	0.00653***	0.00717***	0.00705***
	(0.00031)	(0.00032)	(0.00031)
Num.Obs.	5000	5000	5000

17 潜在変数モデル

- ロジットやプロビットを潜在変数に基づいて導出することもできる.
（ここではプロビットを例に.）
- 個人の意思決定の場合は, 経済学的には効用として解釈できる.

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i$$
$$u_i | X \sim N(0,1)$$

として,

$$Y_i = \begin{cases} 0 & (Y_i^* \leq 0) \\ 1 & (Y_i^* > 0) \end{cases}$$

つづき

- すると,

$$\begin{aligned}\Pr(Y_i = 1 \mid X) &= \Pr(Y_i^* > 0 \mid X) \\ &= \Pr(u_i > -\beta_0 - \beta_1 X_{1i} - \cdots - \beta_k X_k \mid X) \\ &= 1 - \Phi(-\beta_0 - \beta_1 X_{1i} - \cdots - \beta_k X_k) \\ &= \Phi(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_k)\end{aligned}$$

18 内生性

線形回帰と同じように，内生性があると，推定量が一致性を持たない．

- LPMで操作変数法や固定効果法を使う．
- 固定効果ロジットを使う．
- IVプロビットを使う．

19 被説明変数が2値でないとき

- 3つ以上のレベルがあるとき（順序がある）.
 - 例：「不満」「やや不満」「やや満足」「満足」
 - 単に, 1, 2, 3, 4としてOLS.
 - どこかで区切って2カテゴリーに変換：「満足」「不満」
 - $\Pr(Y = 4 | X)$, $\Pr(Y \geq 3 | X)$, $\Pr(Y \geq 2 | X)$ をそれぞれ推定.
 - 順序ロジット, 順序プロビット.
- 3つ以上のカテゴリーがあるとき（順序がない）.
 - 例：希望部署
 - ある選択肢とそれ以外の0-1のダミー変数 D_1, \dots, D_J を作って, $J - 1$ 個の回帰モデルを推定する.
 - 多項ロジット, 多項プロビット.