# 第 13 回：2 値応答モデルの推定（2）

北村　友宏

2021 年 1 月 8 日

# 本日の内容

1. 2 値プロビット・モデルの限界効果

2. gretl での限界効果推定

# 2 値プロビット・モデル

2 値プロビット・モデルは,

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$y_i^* = \beta_0 + \beta_1 x_i + u_i,$$

$$u_i \mid x_i \sim N(0, 1).$$

2 値プロビット・モデルの別の表現は,

$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_1 x_i).$$

$\Phi(.)$ は標準正規分布の累積分布関数.

2 値プロビット・モデルが

$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_1 x_i),$$

と書ける理由は以下の通り．
$x_i$ を所与として，$y_i = 1$ となる条件付き確率は，

$$\begin{aligned} P(y_i = 1 \mid x_i) &= P(y_i^* > 0 \mid x_i) \\ &= P(\beta_0 + \beta_1 x_i + u_i > 0 \mid x_i) \\ &= P\left(u_i > -(\beta_0 + \beta_1 x_i) \mid x_i\right). \end{aligned}$$

標準正規分布は 0 で対称な分布なので，

$$P\left(u_i > -(\beta_0 + \beta_1 x_i) \mid x_i\right) = P\left(u_i < \beta_0 + \beta_1 x_i \mid x_i\right).$$

よって，

$$\begin{aligned} P(y_i = 1 \mid x_i) &= P\left(u_i < \beta_0 + \beta_1 x_i \mid x_i\right) \\ &= \Phi(\beta_0 + \beta_1 x_i). \end{aligned}$$

# 限界効果

2 値プロビット・モデルにおける，$x_i$ の限界効果（marginal effect）は，

$$
\begin{aligned}
\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_i} &= \frac{\partial \Phi(\beta_0 + \beta_1 x_i)}{\partial x_i} \\
&= \phi(\beta_0 + \beta_1 x_i)\beta_1 .
\end{aligned}
$$

$\phi(.)$ は標準正規分布の確率密度関数.

$$\Downarrow$$

$\beta_1$ そのものではなく $\phi(\beta_0 + \beta_1 x_i)\beta_1$ が，「$x_i$ が 1 単位増加したときに $y_i = 1$ となる確率がどの程度変化する傾向があるか」を表す.

- $\phi(.)$ は確率密度関数なので 0 以上.
  $\Rightarrow$ 限界効果の符号は $\beta_1$ の符号と同じ.

- 説明変数 $x_i$ の値は各個体によって異なる.
  ➥ 限界効果

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_i} = \phi(\beta_0 + \beta_1 x_i)\beta_1,$$

  の値も各個体によって異なる.
  $\Rightarrow$ 説明変数 $x_i$ をその平均 $\bar{x}$ で置き換えた，平均における限界効果（marginal effect at mean, slope at mean）を計算する.

- 平均における限界効果（marginal effect at the mean）は，

$$\phi(\hat{\beta}_0 + \hat{\beta}_1 \bar{x})\hat{\beta}_1.$$

  - $\hat{\beta}_0, \hat{\beta}_1$ はそれぞれ $\beta_0, \beta_1$ の最尤推定値．
  - $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$
  - 定数項以外に説明変数が複数個ある場合は，それらを全てそれぞれの標本平均で置き換える．

# 線形確率モデルを適用すべき場合

以下の場合は被説明変数 $y_i$ がダミー変数であっても線形確率モデルを仮定して推定する.

- モデルの右辺に交差項（変数の積）を含み，その係数を解釈したい.
  - e.g., $y_i = \beta_0 + \beta_X x_i + \beta_Z z_i + \beta_{XZ} x_i z_i + u_i$ のような線形確率モデルを推定.
- パネルデータを用いて固定効果モデルを仮定したい.
  - e.g., $y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \varepsilon_{it}$ のような線形確率モデルを推定.

# gretl での限界効果推定

2 値プロビット・モデルの場合は,

- メニューバーから「モデル」→「制限従属変数」→「プロビット」→「二項 (Binary)」と操作.
- ラジオボタンの中から「平均での限界効果（slope at mean）を表示する」を選ぶ.

⇒ 出力結果に各説明変数の,「平均における限界効果」が表示される.

gretl では，平均における限界効果を出力すると $p$ 値および有意性を示すアスタリスク記号が出力されない.

$\Downarrow$

メニューバーから「モデル」→「制限従属変数」→「プロビット」→「二項 (Binary)」と操作して出てくるウィンドウで,「p 値を表示する」を選んだ結果と「平均での限界効果（slope at mean）を表示する」を選んだ結果の両方を出力して，前者の結果から $p$ 値とアスタリスク記号の個数を確認し，後者の結果から限界効果を確認するとよい.

# サッカー選手のチーム移籍

いま整理・加工・分析しているデータセットを用いて，以下の２値プロビット・モデルを推定し，各説明変数の限界効果を推定する．

$$Transfer_i = \begin{cases} 1 & \text{if } Transfer_i^* > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$Transfer_i^* = \beta_0 + \beta_1 Timerate_i + \beta_2 Goalrate_i + u_i,$$

$$u_i \mid x_i \sim N(0, 1).$$

この２値プロビット・モデルの別の表現は，

$$\begin{aligned} P(&Transfer_i = 1 \mid Timerate_i, Goalrate_i) \\ &= \Phi(\beta_0 + \beta_1 Timerate_i + \beta_2 Goalrate_i). \end{aligned}$$

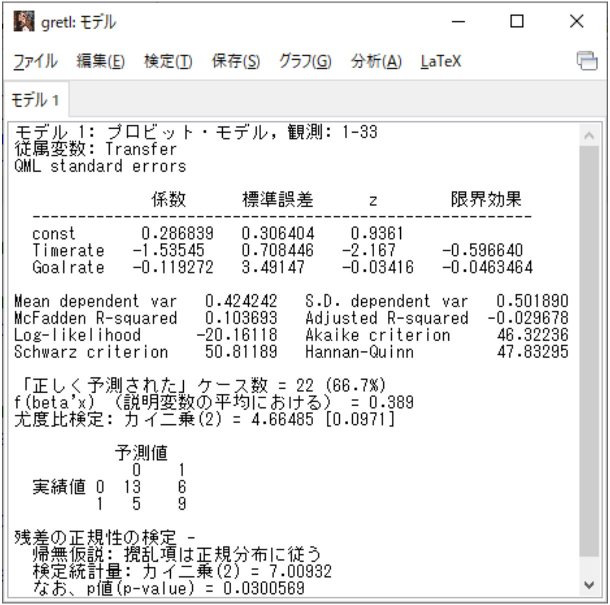
- $Transfer_i$: 移籍ダミー
  - 翌年（2012 年）に移籍した＝ 1
  - 翌年（2012 年）に移籍しなかった（残留した）＝ 0
- $Timerate_i$: 出場時間率
- $Goalrate_i$: 得点率
- $\Phi(.)$: 標準正規分布の累積分布関数

# 実習 1

1. gretl を起動.
2. 「ファイル」→「データを開く」→「ユーザー・ファイル」と操作.
3. jleaguekobe2011.gdt を選択し,「開く」をクリック.
4. gretl のメニューバーから「モデル」→「制限従属変数」→「プロビット」→「二項 (Binary)」と操作.

5. 出てきたウィンドウ左側の変数リストにある Transfer をクリックし，3 つの矢印のうち上の青い右向き矢印をクリック.
    - 推定式の左辺の変数（被説明変数，従属変数）が「『Transfer』が 1 になる確率（移籍する確率）」となる.
6. ウィンドウ左側の変数リストにある Timerate をクリックした後，Ctrl キーを押しながら Goalrate をクリックして，3 つの矢印のうち真ん中の緑の右向き矢印をクリック.
    - 推定式の右辺の変数（説明変数，独立変数）が Timerate（出場時間率）と Goalrate（得点率）となる.
    - 最初から説明変数リストに入っている const は推定式の切片（定数項）のこと.

7. 「頑健標準誤差を使用する」にチェックする．このデータは横断面データのため，不具合は発生しないと考えられる．
    - モデルの定式化に対して頑健な標準誤差が計算される．
8. ラジオボタンの「平均での限界効果（slope at mean）を表示する」をクリック．
    - 各説明変数の，「平均における限界効果」が表示されるようになる．
9. 「OK」をクリックすると，結果が表示される．




```
モデル 1: プロビット・モデル, 観測: 1-33
従属変数: Transfer
QML standard errors

                 係数        標準誤差        z          限界効果
  ---------------------------------------------------------------
  const        0.286839    0.306404     0.9361
  Timerate    -1.53545     0.708446    -2.167      -0.596640
  Goalrate    -0.119272    3.49147     -0.03416    -0.0463464

Mean dependent var   0.424242   S.D. dependent var   0.501890
McFadden R-squared   0.103693   Adjusted R-squared  -0.029678
Log-likelihood      -20.16118   Akaike criterion     46.32236
Schwarz criterion    50.81189   Hannan-Quinn         47.83295

「正しく予測された」ケース数 = 22 (66.7%)
f(beta'x) (説明変数の平均における) = 0.389
尤度比検定: カイ二乗(2) = 4.66485 [0.0971]

           予測値
             0    1
  実績値 0  13    6
         1   5    9

残差の正規性の検定 -
  帰無仮説: 撹乱項は正規分布に従う
  検定統計量: カイ二乗(2) = 7.00932
  なお、p値(p-value) = 0.0300569
```

このような画面が表示されれば成功.

# 限界効果推定結果

- 出場時間率の限界効果
  - −0.596640
    ➡ 出場時間率が 0.01 高くなると（1 パーセントポイント高くなると），チームを移籍する確率が 0.0059664 低くなる（0.59664 パーセントポイント低くなる）.
- 得点率の限界効果
  - −0.0463464
    ➡ 得点率が 0.01 高くなると（1 パーセントポイント高くなると），チームを移籍する確率が 0.000463464 低くなる（0.0463464 パーセントポイント低くなる）.

前回の仮説検定で，出場時間率のみ，係数ゼロの帰無仮説が棄却されたことから，出場機会に恵まれないサッカー選手がチームを移籍する傾向がある.

# 実習 2

1. 「gretl: モデル 1」のウィンドウのメニューバーから「ファイル」→「名前を付けて保存」と操作.
2. 「標準テキスト」を選び,「OK」をクリック。
3. プロビットモデル推定結果 2.txt という名前で「2020 ミクロデータ分析 2」フォルダに保存. すると,表示された推定結果をそのままテキストファイルで保存できる.

# 2 値プロビット・モデルの推定結果表

- レポートや論文に載せるための 2 値プロビット・モデルの推定結果表を作成する際には，最低限，以下の情報を載せればよい．
  - 係数推定値
  - 限界効果推定値
  - 「係数ゼロ仮説の検定のための $z$ 値または $p$ 値」または「係数の標準誤差」のどれか
  - 対数尤度
  - 観測値数
- 定数項の限界効果は存在しない．

レポートや論文には，例えば以下のような表を載せればよい.

表 1：2 値プロビット・モデル推定結果

| | 偏回帰係数 | 限界効果 | $z$ 値 | |
|---|---|---|---|---|
| 出場時間率 | −1.54 | −0.60 | −2.17 | ** |
| 得点率 | −0.12 | −0.05 | −0.03 | |
| 定数項 | 0.29 | | 0.94 | |
| 対数尤度 | −20.16 | | | |

（注 1）表中の**は有意水準 5%で統計的に有意であることを表す.
（注 2）モデルの定式化に対して頑健な標準誤差を用いている.
（注 3）観測値数は 33 である.

本日の作業はここまで.
今回は gretl のデータセットに変更を加えていないので，gretl のデータセット（jleaguekobe2011.gdt）を上書き保存する必要はない.