

二値アウトカム用推定コマンド

従属変数が二値（通常は 0/1）のアウトカムを表す場合、そのモデル式は次のように表現されます。

$$\Pr(y_j \neq 0) = F(\mathbf{x}_j\beta) \quad (5.1)$$

ただし $F()$ は何らかの累積分布関数を表します。 $F()$ として正規分布の累積分布関数を用いるものがプロビット (probit) モデル、ロジスティック分布の累積分布関数を用いるものがロジット (logit) モデル、すなわちロジスティック (logistic) 回帰モデルです。この他に complementary log-log 関数を用いた cloglog モデルもよく使用されます。

プロビットモデル、ロジットモデルに対する最尤推定法として実装されているのがそれぞれ `probit` ([R] `probit` 参照) コマンド、`logit` ([R] `logit` (*mwp-039*) 参照) コマンドです。ロジットモデルについては `logistic` ([R] `logistic` (*mwp-039*) 参照) というコマンドも選択できますが、`logit` と `logistic` の間に本質的な違いはありません。係数推定値のデフォルト表示様式に差があるに過ぎないからです。ロジットモデルに関しては正確推定法 (exact estimators) に基づく `exlogistic` ([R] `exlogistic` 参照) という選択肢も用意されています。

“logit”と“logistic”という用語の定義については分野によってまちまちです。このため Stata では `logit` と `logistic` という 2 つのコマンドを提供していますが、選択は好みの問題と言えます。`logit` の場合、係数推定値 $\hat{\beta}$ がそのままの形で表示されるのに対し、`logistic` の場合には指数化された $\exp(\hat{\beta})$ の形、すなわちオッズ比 (odds ratio) として表示されるという違いがあります。しかしオプションを指定すれば `logit` であってもオッズ比の出力が可能であり、`logistic` であっても係数の直接表示が可能です。推定内容自体に違いがあるわけではありません。

(1) ロジット系コマンド

`logit` と `logistic` に加え、`blogit`, `glogit`, `binreg`, `scobit` というコマンドも用意されています。

1. `blogit` ([R] `glogit` 参照) はグルーピング (ブロッキング) されたデータに対して最尤法に基づくロジスティック回帰を適用します。`logit`/`logistic` の場合、個々の observation (観測データ) は個体の情報を表しているわけですが、グルーピングされたデータの場合には個々の observation はグループごとの success 数/failure 数を記録する形となります。
2. `glogit` ([R] `glogit` 参照) は `blogit` と同様、グルーピングされたデータを操作対象としますが、推定法として重み付き最小 2 乗法 (weighted least-squares) が用いられる点が異なります。

3. `binreg` ([R] `binreg` 参照) は一般化線形モデル (GLM: generalized linear model) の概念に基づき推定を行うもので、対象とするデータは個体レベル、グループレベルのいずれでも構いません。分布族としては二項分布が仮定されますが、リンク関数としては次のような選択が可能です。

| オプション | リンク関数 | パラメータ |
|-----------------|-----------------------------|-------------------------------|
| <code>or</code> | <code>logit</code> | オッズ比 = $\exp(\beta)$ |
| <code>rr</code> | <code>log</code> | リスク比 = $\exp(\beta)$ |
| <code>hr</code> | <code>log complement</code> | Health ratios = $\exp(\beta)$ |
| <code>rd</code> | <code>identity</code> | リスク差 = β |

4. `scobit` ([R] `scobit` 参照) は skewed logit model のフィットを行います。この場合、(5.1) 式における $F(z)$ は

$$F(z) = 1 - \frac{1}{\{1 + \exp(z)\}^\alpha} \quad (5.2)$$

のように表現されます。パラメータ α が加わる点が通常のロジットと異なる点です。

(2) プロビット系コマンド

標準的な `probit` に加え、`ivprobit`, `bprobit`, `gprobit`, `biprobit` 等のコマンドも用意されています。

- `ivprobit` ([R] `ivprobit` 参照) は回帰変数中に内生変数が含まれる場合にプロビットモデルをフィットさせます。
- `bprobit` ([R] `glogit` 参照) はグルーピング (ブロック) されたデータに対して最尤法に基づくプロビット回帰を適用します。
- `gprobit` ([R] `glogit` 参照) は `bprobit` と同様、グルーピングされたデータを操作対象としますが、推定法として重み付き最小 2 乗法が用いられる点が異なります。
- `hetprobit` ([R] `hetprobit` 参照) は分散が不均一 (heteroskedastic) であることを前提にプロビットモデルをフィットさせます。この場合、誤差項の分散はパラメータ化されることになります。
- `heckprobit` ([R] `heckprobit` 参照) は標本選択 (sample selection) を伴うプロビットモデルをフィットさせます。
- `biprobit` ([R] `biprobit` 参照) は 2 変量プロビットモデル (bivariate probit models) をフィットさせます。この場合、相関を持った 2 つのアウトカム変数が存在することになります。また (0, 0) と (1, 1) という組合せのみ観測し得るという partial-observability models の扱いも可能です。

■