est-001

線形回帰系の推定コマンド

最小 2 乗法 (OLS: ordinary least squares) 推定法が前提とする線形回帰モデル式は次のように記述されます。

$$y_j = \mathbf{x}_j\boldsymbol{\beta} + \epsilon_j \tag{1.1}$$

ただし次の条件が付くので注意してください。

- y は連続変数である。
- σ_ϵ^2 は observation j によらず一定である。

この OLS 推定法の機能を提供する基盤となるコマンドが `regress`（[R] **regress** (*mwp-037*) 参照）です。ただし `regress` にはロバストな推定法も実装されており、分散の均一性 (homoskedasticity) が仮定できないモデルにも対応できるようになっています。またフィットされたモデルの診断等に関わる多彩な postestimation 機能が `regress` には用意されています（[R] **regress postestimation** (*mwp-038*) 参照)。

線形回帰の機能は以下のコマンドによっても提供されますが、それぞれ独自の機能・特徴を有しています。

1. `ivregress`（[R] **ivregress** (*mwp-082*) 参照）は一部の回帰変数が内生変数 (endogenous variables) であるようなモデルを対象にフィットを行います。基本的には操作変数 (instrumental variables) 法による推定を行いますが、一般化モーメント (GMM: generalized method of moments) 推定法を選択することもできます。

2. `areg`（[R] **areg** 参照）は

$$y_j = \mathbf{x}_j\boldsymbol{\beta} + \mathbf{d}_j\boldsymbol{\gamma} + \epsilon_j \tag{1.2}$$

というモデルをフィットさせます。ここに $\mathbf{d}_j$ は observations を区分するためのダミー変数群を表します。ただし実際に変数群 $\mathbf{d}_j$ が構成されるわけではなく、また $\boldsymbol{\gamma}$ の推定値が出力されるわけでもありません。なお、パネルごとの固定効果 (fixed-effects) をモデル化したいのであれば `areg` ではなく、より汎用的な `xtreg`（[XT] **xtreg** (*mwp-017*) 参照）の使用を推奨します。なお、*mwp-017* 中には `areg` の用例に関する記述も含めてありますので参考にしてください。

3. `boxcox`（[R] **boxcox** 参照）は

$$y_i^{(\theta)} = \beta_0 + \beta_1 x_{i1}^{(\lambda)} + \beta_2 x_{i2}^{(\lambda)} + \cdots + \beta_k x_{ik}^{(\lambda)} + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \cdots + \gamma_l z_{il} + \epsilon_i \tag{1.3}$$

というモデルをフィットさせます。ただし ϵ は $N(0,\sigma^2)$ に従うものとします。ポイントとなるのは、従属変数 y、及び独立変数 $x_1, x_2, \ldots, x_k$ に対して Box-Cox 変換が施されるという点にあります。その際、y に対してはパラメータ値を θ とする変換が、$x_1, x_2, \ldots, x_k$ に対してはパラメータ値を λ とする変換が適用されます。これに対し $z_1, z_2, \ldots, z_l$ は通常の独立変数で変換の対象とはなりません。な

お、(1.3) 式は一般形ですが、$\lambda = \theta$, $\lambda = 1$, $\theta = 1$ という特殊ケースに対応したモデルも用意されています。

4. `tobit`（`[R]` **tobit** 参照）は y_i に対して左打切り (left-censoring) や右打切り (right-censoring) がある場合に使用されます。例えば $y_i \geq 1,000$ のときに実際の値ではなく打切り値 $1,000$ が記録されているようなケースがこれに該当します。
   `ivtobit`（`[R]` **ivtobit** 参照）は内生回帰変数を許容した形で `tobit` と同様の機能を提供します。

5. `intreg`（`[R]` **intreg** 参照）は区間回帰 (interval regression) の機能を提供するコマンドで、`tobit` を一般化したものと言えます。すなわち `intreg` は開区間のみならず閉区間をも許容します。観測値 y_j の値域が $y_{0j} \leq y_j \leq y_{1j}$ であるとしたとき、`intreg` では y_{0j} と y_{1j} が観測されるものと仮定します。サーベイデータにおいて月収が$1,500-$2,500 の範囲にあると記録するようなケースがこれに該当します。`intreg` を使うとこのような区間データを対象とした回帰モデルのフィットが行えます。

6. `truncreg`（`[R]` **truncreg** 参照）は標本が母集団のある限られた範囲から抽出されているようなケースに対して回帰モデルをフィットさせます。`tobit` の場合には従属変数の値域に制限があったわけですが、`truncreg` の場合には独立変数側の値域に制限が加わることが仮定されます。母集団の分布を正規分布としたとき、切断回帰 (truncated regression) モデル中の誤差項は切断正規分布 (truncated-normal distribution) に従うことになります。

7. `cnsreg`（`[R]` **cnsreg** 参照）を使用すると係数に関する線形の制約を課した状態で線形回帰モデルをフィットさせることができます。

8. `eivreg`（`[R]` **eivreg** 参照）は独立変数中に測定誤差が含まれるとする変数誤差回帰モデル (errors-in-variables regression models) をフィットさせます。

9. `nl`（`[R]` **nl** 参照）は
   $$y_j = f(\mathbf{x}_j, \boldsymbol{\beta}) + \epsilon_j \tag{1.4}$$
   のような一般式で記述される非線形回帰モデルを最小 2 乗法によりフィットさせます。

10. `rreg`（`[R]` **rreg** 参照）はロバスト回帰モデル (robust regression models) のフィットを行います。これは標準誤差のロバスト推定を行うための `robust` オプションとは別個のもので、点推定値をよりロバストな形で推定しようとするものです。`rreg` は外れ値 (outliers) の影響を軽減するためのデータ依存の手法を実装しています。

11. `qreg`（`[R]` **qreg** 参照）は分位点回帰 (quantile-regression) の機能を提供します。モデル式は
    $$y_j = \mathbf{x}_j\boldsymbol{\beta} + \epsilon_j \tag{1.5}$$
    のように表現されますが、線形回帰とは別物です。このモデルの基本形は中央値回帰 (median regression) と呼ばれることがあります。通常の線形回帰ではベクトル $\mathbf{x}_j$ が与えられたときに予測される y_j の平均値を $\mathbf{x}_j\boldsymbol{\beta}$ は表現するわけですが、中央値回帰の場合には y_j の中央値を表現することになります。このため `qreg` は ϵ_j に関する分散均一性が仮定できない場合にしばしば用いられます。より一般的な分位点回帰といった場合には分位点の指定が可能となるため、パーセンタイルの値 (1st, 2nd, . . . , 99th percentile) に関する線形推定値を求めることが可能です。

bsqreg（[R] **qreg** 参照）も qreg と同等の機能を提供しますが、標準誤差はブートストラップ法によって推定されます。

sqreg（[R] **qreg** 参照）は複数の分位点に関する推定を同時に行います。その際、標準誤差の推定にはブートストラップ法が用いられます。

iqreg（[R] **qreg** 参照）は 2 つの分位点間の差を推定します。標準誤差の推定には同様にブートストラップ法が用いられます。

12. vwls（[R] **vwls** 参照）は分散重み付き最小 2 乗法 (variance-weighted least squares) の機能を提供します。モデル式は

$$y_j = \mathbf{x}_j \boldsymbol{\beta} + \epsilon_j \tag{1.6}$$

のように表現されますが、分散均一性に関する条件は前提としません。代りに ϵ_j は推定に先立ち、グループデータから事前に算出されることになります。

■