

## 推定機能 【 評価版 】

Stata には regress, logistic, xtreg, stcox といった数多くの推定コマンドが用意されています。それら推定コマンドの集合体に若干の補足機能を加えたものが Stata の実体であると言っても過言ではありません。一方、これらの推定コマンドには例外なく postestimation 機能が付帯しています。実際マニュアル上には regress とか logistic といったエントリとは別に、regress postestimation や logistic postestimation といったエントリが用意されており、多数のページが割かれていることが確認できると思います。本 whitepaper では推定、及び postestimation 機能全般に関する基本的な事項について解説を行います。個々の推定コマンド、個々の postestimation 機能に関する情報については個別の whitepaper、またはマニュアルエントリをご参照ください。

1. e クラスコマンド
2. 係数の参照
3. predict コマンド
4. estat コマンド
5. 推定結果の保存と回復
  - 5.1 メモリ上への退避/回復
  - 5.2 ディスク上への保存/回復
6. test コマンド
7. vce オプション
8. 重み付き推定
  - 8.1 fweight
  - 8.2 aweight
  - 8.3 pweight

## 1. e クラスコマンド

Stata のコマンドは summarize 等の一般的な r クラスコマンドと推定機能を司る e クラスコマンドの 2 群に区分されます\*<sup>1</sup>。この区分はコマンドの実行結果が格納される領域の種類や構造に依存したものであるわけですが、試しに線形回帰用の regress コマンドを例に取ってその内容を覗いてみましょう。まず Example データセット auto.dta をロードします。

```
. sysuse auto.dta *2
```

(1978 automobile data)

このデータセット中には 1978 年に米国で販売された 74 車種のデータが収納されていますが、ここでは走行燃費性能を表す mpg (miles per gallon) を車重 weight と国産/非国産の区分を表す指標変数 foreign とによって説明する線形回帰モデルをフィットさせてみます。

```
. regress mpg weight foreign *3
```

| . regress mpg weight foreign |             |           |            |               |                      |           |
|------------------------------|-------------|-----------|------------|---------------|----------------------|-----------|
| Source                       | SS          | df        | MS         | Number of obs | =                    | 74        |
| Model                        | 1619.2877   | 2         | 809.643849 | F(2, 71)      | =                    | 69.75     |
| Residual                     | 824.171761  | 71        | 11.608053  | Prob > F      | =                    | 0.0000    |
|                              |             |           |            | R-squared     | =                    | 0.6627    |
|                              |             |           |            | Adj R-squared | =                    | 0.6532    |
| Total                        | 2443.45946  | 73        | 33.4720474 | Root MSE      | =                    | 3.4071    |
| mpg                          | Coefficient | Std. err. | t          | P> t          | [95% conf. interval] |           |
| weight                       | -.0065879   | .0006371  | -10.34     | 0.000         | -.0078583            | -.0053175 |
| foreign                      | -1.650029   | 1.075994  | -1.53      | 0.130         | -3.7955              | .4954422  |
| _cons                        | 41.6797     | 2.165547  | 19.25      | 0.000         | 37.36172             | 45.99768  |

foreign に対する  $p$  値が 0.130 と 0.05 よりも大きいため、説明変数としての有意性には疑問がありますが (実際、95% CI の中に値 0 が含まれており、係数値が 0 である可能性が否定できないことが示されています) ここでは目をつぶって先に進むことにします。フィットされたモデル式は次のように記述できます。

$$\begin{aligned}\widehat{\text{mpg}} &= \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{weight} + \hat{\beta}_2 \cdot \text{foreign} \\ &= 41.68 - 0.0066 \cdot \text{weight} - 1.65 \cdot \text{foreign}\end{aligned}\tag{1}$$

この regress コマンドの実行結果は e() という領域中に格納されています。一応その内容を確認しておきましょう。

\*<sup>1</sup> 細かく言えばこの他に s クラス、n クラスといった区分があります。

\*<sup>2</sup> メニュー操作: File ▸ Example datasets ▸ Example datasets installed with Stata と操作しロードする。

\*<sup>3</sup> メニュー操作: Statistics ▸ Linear models and related ▸ Linear regression

```
. ereturn list
```

```
. ereturn list

scalars:
      e(N) = 74
    e(sum_w) = 74
    e(df_m) = 2
    e(df_r) = 71
      e(F) = 69.74846262000308
    e(r2) = .6627029116028815
    e(rmse) = 3.407059285651584
    e(mss) = 1619.287698167387
    e(rss) = 824.1717612920727
    e(r2_a) = .6532015851691599
    e(ll) = -194.1830643938065
    e(ll_0) = -234.3943376482347
    e(rank) = 3

macros:
    e(cmdline) : "regress mpg weight foreign"
    e(title) : "Linear regression"
    e(marginsok) : "XB default"
    e(vce) : "ols"
    e(depvar) : "mpg"
    e(cmd) : "regress"
    e(properties) : "b V"
    e(predict) : "regres_p"
    e(model) : "ols"
    e(estat_cmd) : "regress_estat"

matrices:
      e(b) : 1 x 3
      e(V) : 3 x 3
    e(beta) : 1 x 2

functions:
    e(sample)
```



r クラスコマンドの出力内容は return list コマンドによって確認できます。

Stata 環境でプログラミングを行う場合を除くと  $e()$  領域の内容を直接意識することは余りありませんが、例えば推定結果である係数ベクトル  $\beta$  は  $e(b)$  の中に格納されています。

```
. matrix list e(b)
```

```
. matrix list e(b)

e(b)[1,3]
      weight      foreign      _cons
y1  -.00658789  -1.6500291  41.679702
```

また  $e(V)$  の中には推定された分散共分散行列の値が格納されています。

```
. matrix list e(V)
```

```
. matrix list e(V)

symmetric e(V)[3,3]
      weight      foreign      _cons
weight  4.059e-07
foreign .0004064  1.1577633
_cons  -.00134646 -1.5713159  4.6895946
```

Postestimation 機能として区分されるコマンド群はこれら  $e()$  領域中に格納されている情報に基づき処理を行います。

## 2. 係数の参照

評価版では割愛しています。

## 3. predict コマンド

セクション 2 では係数参照という少々ベーシックな方法を用いて予測値  $\widehat{\text{mpg}}$  を算出したわけですが、postestimation コマンドの代表格である predict コマンドを使用すれば苦もなく予測値を算出することができます。

- Statistics ▷ Postestimation ▷ Predictions
  - ▷ Predictions and their SEs, leverage statistics, distance statistics, etc. ▷ Launch と操作
- predict ダイアログ: Main タブ: New variable name: mpghat2
  - Produce: Linear prediction (xb) (デフォルト)

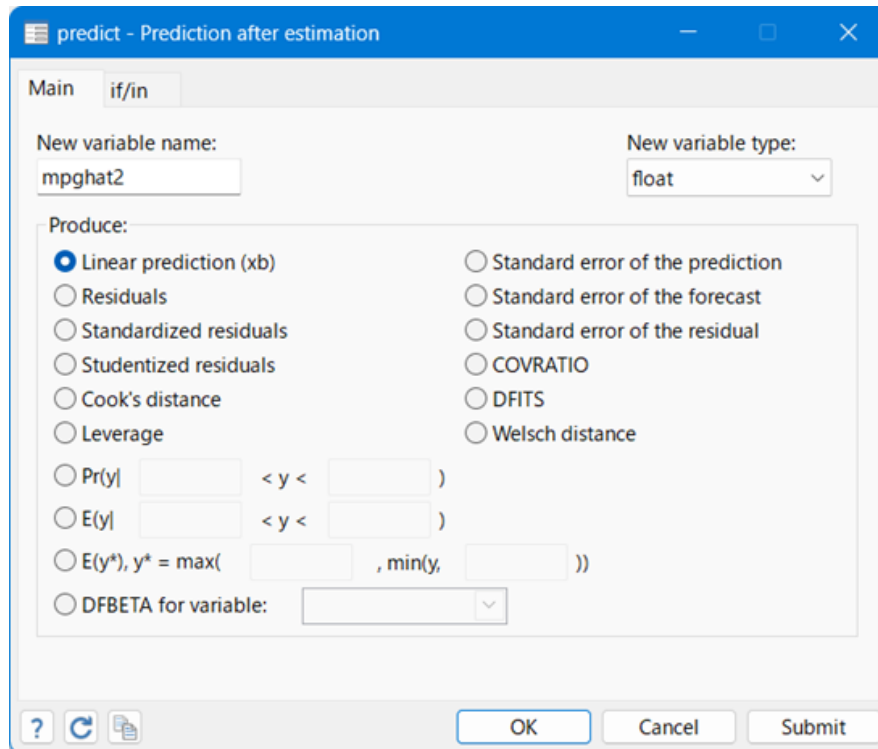


図 1 predict ダイアログ - regress 実行後

```
. predict mpghat2, xb
```

predict によって生成された新変数 mpghat2 の内容を確認しておきます。

```
. list make weight foreign mpg mpghat1 mpghat2 in 1/5, nlabel
```

|    | make          | weight | foreign | mpg | mpghat1  | mpghat2  |
|----|---------------|--------|---------|-----|----------|----------|
| 1. | AMC Concord   | 2,930  | 0       | 22  | 22.3772  | 22.3772  |
| 2. | AMC Pacer     | 3,350  | 0       | 17  | 19.61028 | 19.61028 |
| 3. | AMC Spirit    | 2,640  | 0       | 22  | 24.28768 | 24.28768 |
| 4. | Buick Century | 3,250  | 0       | 20  | 20.26907 | 20.26907 |
| 5. | Buick Electra | 4,080  | 0       | 15  | 14.80113 | 14.80113 |

セクション 2 で算出した mpghat1 の値と一致していることがわかると思います。

評価版では割愛しています。

#### 4. estat コマンド

評価版では割愛しています。

#### 5. 推定結果の保存と回復

評価版では割愛しています。

#### 6. test コマンド

評価版では割愛しています。

#### 7. vce オプション

評価版では割愛しています。

#### 8. 重み付き推定

評価版では割愛しています。

■