

glm - 一般化線形モデル 【 評価版 】

glm は一般化線形モデルのフィットを行います。推定の手法としては IRLS と Newton-Raphson 法が選択できます。

- | | |
|-------------|------------------------|
| 1. 一般化線形モデル | |
| 2. glm の用例 | Example 1 Example 2 |
| 3. 分散推定法 | Example 3 Example 4 |
| 4. ユーザ定義関数 | |

1. 一般化線形モデル

glm が対象とする一般化線形モデル (generalized linear models) のモデル式は次のように表現されます。

$$g\{E(y)\} = \mathbf{x}\boldsymbol{\beta}, \quad y \sim F \quad (1)$$

$g()$ はリンク関数 (link function) を、 F は分布族 (distributional family) を表します。種々の定義を $g()$ と F に当てはめることにより多様なモデルが構成されることとなります。例えば y が正規分布に従い、 $g()$ を恒等関数 (identity function) とした場合には

$$E(y) = \mathbf{x}\boldsymbol{\beta}, \quad y \sim \text{Normal} \quad (2)$$

すなわち線形回帰が表現されます。また y がベルヌーイ分布に従い、 $g()$ をロジット関数とした場合には

$$\text{logit}\{E(y)\} = \mathbf{x}\boldsymbol{\beta}, \quad y \sim \text{Bernoulli} \quad (3)$$

すなわちロジスティック回帰が誘導されます。さらに y の分布をポアソン分布とし、 $g()$ を自然対数とした場合には

$$\ln\{E(y)\} = \mathbf{x}\boldsymbol{\beta}, \quad y \sim \text{Poisson} \quad (4)$$

という形でポアソン回帰が導かれます。もちろんこれら以外の組合せも想定可能です。

原理的には `glm` を使って線形回帰を実行することも可能ですが（実際デフォルトでは線形回帰を実行します）、これはあくまで形式的な話に過ぎません。 `regress` の方が計算時間は短く、かつそれには数多くの `postestimation` コマンドが付帯しているからです。

`glm` を使用する場合、リンク関数は `link()` オプションで、分布族は `family()` オプションで指定します。指定可能なリンク関数は表 1 の通りです。

表 1 リンク関数

| リンク関数 | glm オプション |
|-----------------------|------------------------------|
| identity | <code>link(identity)</code> |
| log | <code>link(log)</code> |
| logit | <code>link(logit)</code> |
| probit | <code>link(probit)</code> |
| complementary log-log | <code>link(cloglog)</code> |
| odds power | <code>link(opower #)</code> |
| power | <code>link(power #)</code> |
| negative binomial | <code>link(nbinomial)</code> |
| log-log | <code>link(loglog)</code> |
| log-complement | <code>link(logc)</code> |

評価版では割愛しています。

一方、分布族として指定できるのは表 2 の通りです。

表 2 分布族

| 分布 | glm オプション |
|------------|--------------------------------|
| 正規分布 | <code>family(gaussian)</code> |
| 逆正規分布 | <code>family(igaussian)</code> |
| ベルヌーイ/二項分布 | <code>family(binomial)</code> |
| ポアソン分布 | <code>family(poisson)</code> |
| 負の二項分布 | <code>family(nbinomial)</code> |
| ガンマ分布 | <code>family(gamma)</code> |

評価版では割愛しています。

2. glm の用例

▷ Example 1

[R] `logistic` (*mwp-039*) の Example 1 において低体重児の出生に関わるリスク要因をモデル化すべくロジスティック回帰を実行しました。ここでは `glm` を用いて同様の結果が得られるかどうかを確認してみます。

```
. use http://www.stata-press.com/data/r15/lbw.dta *1
(Hosmer & Lemeshow data)
```

`glm` コマンドを `family(binomial), link(logit)` とする形で実行します。

- Statistics ▷ Generalized linear models ▷ Generalized linear models (GLM) と操作
- Model タブ: Dependent variable: `low`
Independent variables: `age lwt i.race smoke ptl ht ui`
Family and link choices: (Binomial, logit)

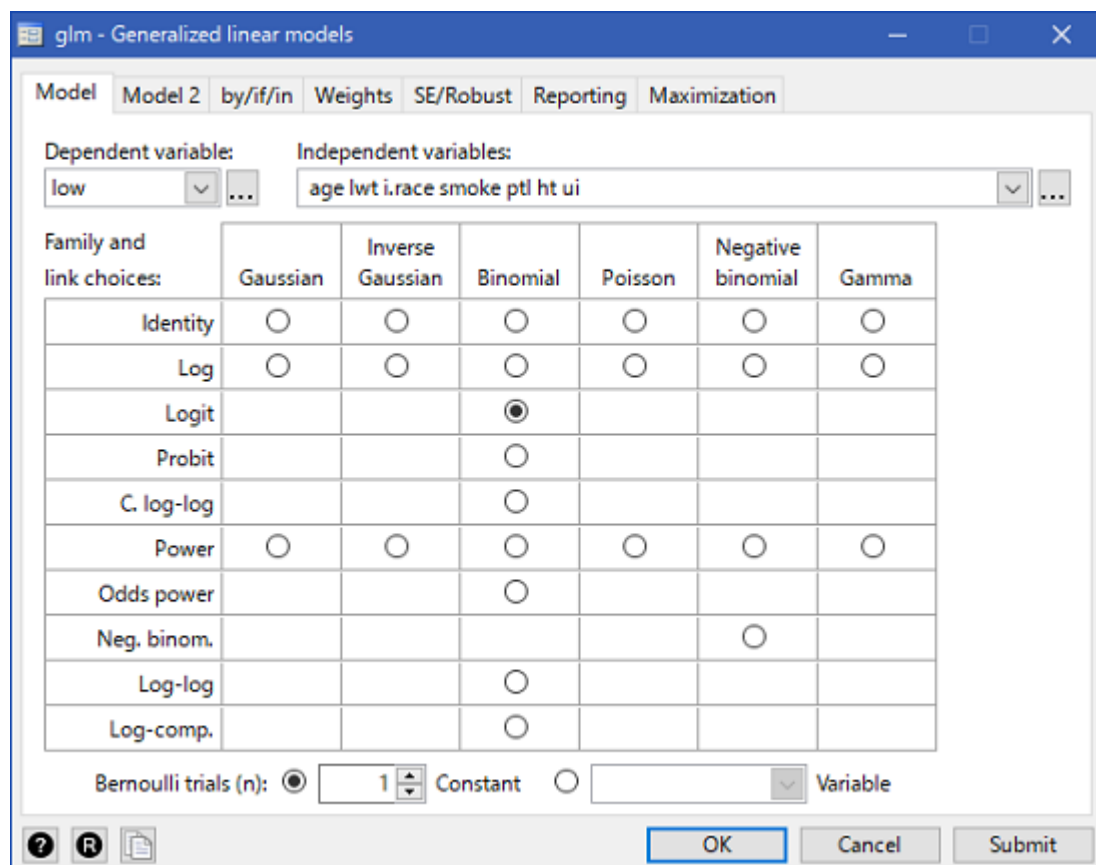


図 1 glm ダイアログ - Model タブ

*¹ メニュー操作 : File ▷ Example Datasets ▷ Stata 15 manual datasets と操作、Base Reference Manual [R] の `glm` の項よりダウンロードする。

```

. glm low age lwt i.race smoke ptl ht ui, family(binomial 1) link(logit)

Iteration 0:  log likelihood = -101.0213
Iteration 1:  log likelihood = -100.72519
Iteration 2:  log likelihood = -100.724
Iteration 3:  log likelihood = -100.724

Generalized linear models                               No. of obs      =       189
Optimization      : ML                                 Residual df    =       180
                                                         Scale parameter =         1
Deviance          = 201.4479911                       (1/df) Deviance = 1.119156
Pearson          = 182.0233425                         (1/df) Pearson  = 1.011241

Variance function: V(u) = u*(1-u)                    [Bernoulli]
Link function     : g(u) = ln(u/(1-u))                [Logit]

Log likelihood    = -100.7239956                       AIC           =       1.1611
                                                         BIC           =      -742.0665

```

| low | OIM | | | | |
|-------|-----------|-----------|-------|-------|----------------------|
| | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
| age | -.0271003 | .0364504 | -0.74 | 0.457 | -.0985418 .0443412 |
| lwt | -.0151508 | .0069259 | -2.19 | 0.029 | -.0287253 -.0015763 |
| race | | | | | |
| black | 1.262647 | .5264101 | 2.40 | 0.016 | .2309024 2.294392 |
| other | .8620792 | .4391532 | 1.96 | 0.050 | .0013548 1.722804 |
| smoke | .9233448 | .4008266 | 2.30 | 0.021 | .137739 1.708951 |
| ptl | .5418366 | .346249 | 1.56 | 0.118 | -.136799 1.220472 |
| ht | 1.832518 | .6916292 | 2.65 | 0.008 | .4769494 3.188086 |
| ui | .7585135 | .4593768 | 1.65 | 0.099 | -.1418484 1.658875 |
| _cons | .4612239 | 1.20459 | 0.38 | 0.702 | -1.899729 2.822176 |

推定された係数値、標準誤差共に、coef オプション（係数値表示）指定時の logistic 出力と一致していることが確認できます。

評価版では割愛しています。

▷ Example 2

評価版では割愛しています。

3. 分散推定法

評価版では割愛しています。

4. ユーザ定義関数

評価版では割愛しています。

