

## 一般化線形モデル

一般化線形モデル (GLM: generalized linear model) のモデル式は次のように表現されます。

$$g\{E(y_j)\} = \mathbf{x}_j\boldsymbol{\beta}, \quad y_j \sim F \quad (4.1)$$

ここに  $g()$  はリンク関数 (link function) と呼ばれ、 $F$  は指数型分布族を表します。GLM をサポートする Stata のコマンドは `glm` (`[R] glm` 参照) であり、推定に先立ちリンク関数と分布族の指定が必要となります。

GLM のフレームワークは極めて一般性の高いものなので、線形回帰のみならず、ポアソン回帰、指数回帰等、多様な回帰モデルがカバーされることとなります。しかし `regress`, `poisson`, `streg` 等の専用コマンドにもそれなりの長所があるため、Stata のコマンド体系の中では両者が併存する形となっています。

`glm` の推定法としては最尤法がデフォルトで用いられますが、IRLS (iterated reweighted least squares) 推定法を選択することもできます (`irls` オプション)。それぞれの分布族  $F$  ごとに正準リンク (canonical link) と呼ばれるリンク関数  $g()$  が対応して存在します。正準リンクを使用した場合には IRLS 推定の結果と最尤法推定の結果は同一のものとなります。ある分布族に対し正準リンク以外のリンク関数を組み合わせて使用した場合、回帰係数の標準誤差推定値は推定法によって異なった値となりますが、正当な推定値であることに変わりはありません。漸近的には最尤推定法と等価なものであるにせよ、小標本の場合には推定結果は多少異なったものとなります。

例えば二項分布族 (binomial family) に対する正準リンクは `logit` です。従ってこの組合せで `glm`, `irls` とコマンド投入した場合の推定結果は最尤法に基づく `logit` (または `logistic`) コマンドの実行結果と同一となります。一方、二項分布族と `probit` リンクを組み合わせるとプロビットモデルが構成されることとなりますが、`probit` はこの場合の正準リンクではありません。従ってこの組合せで `glm`, `irls` とコマンド投入した場合に得られる標準誤差推定値は、最尤法ベースの `probit` コマンドの結果と多少異なったものとなります。

`glm` とは別に専用のコマンドが用意されている場合には後者の使用を推奨します。`glm` の機能が有用であるのは、他のコマンドではカバーされない分布族とリンク関数の組合せを指向するような場合です。

`glm` は誤差が分散不均一 (heteroskedastic) であつたり自己相関を持つもの (autocorrelated) であつたりしても、一致性を維持した形で分散共分散行列を推定することができます。これらの VCE 推定法の中には専用コマンド側では提供されないものもあるので、その面での利用価値も考えられます。

■