mwp-024

streg - パラメトリックモデル　【 評価版 】

`streg` は最尤法を用いた**パラメトリック**な生存時間分析機能を提供します。

Stata14 で新たにサポートされた `xtstreg`, `mestreg` コマンドについてはそれぞれ `[XT]` **xtstreg** (*mwp-247*), `[ME]` **mestreg** (*mwp-249*) をご参照ください。

1. パラメトリックモデル

パラメトリックモデルの場合には、ハザード関数、もしくは誤差分布の形状について関数形を明示した上で推定を行うことになります。生存時間をパラメータ化する流儀にはいくつかありますが、`streg` コマンドでは PH **モデル** (proportional hazards model) と AFT **モデル** (accelerated failure-time model) の 2 種類に対応しています。

(1) PH モデル

この流儀ではハザード関数を

$$h(t) = h_0(t)g(\mathbf{x}) \tag{1}$$

のようにモデル化します。Cox 比例ハザードモデルの場合、$h_0(t)$ については何も規定せずとも推定が行えたわけですが、パラメトリックモデルの場合には $h_0(t)$ について関数形を規定します。`streg` では次の 3 種類の PH モデルがサポートされています。

■ **指数モデル**

モデル式 (1) において

$$h_0(t) = 1, \quad g(\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x}) \tag{2}$$

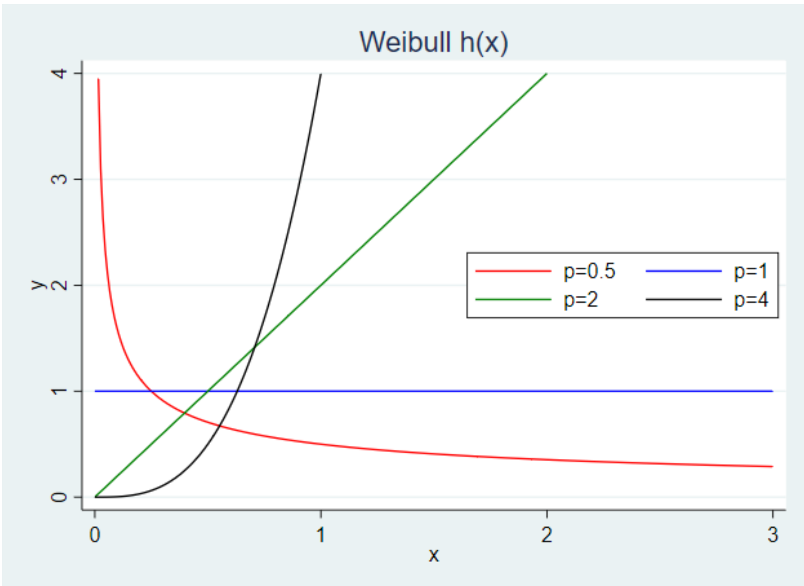
とモデル化するもので、Weibull モデルの $p = 1$ の場合に相当します。

■ **Weibull モデル**

モデル式 (1) において

$$h_0(t) = pt^{p-1}, \quad g(\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x}) \tag{3}$$

とモデル化するもので、補助パラメータ p も推定対象となります。ハザード関数の関数形状は次のようになります。

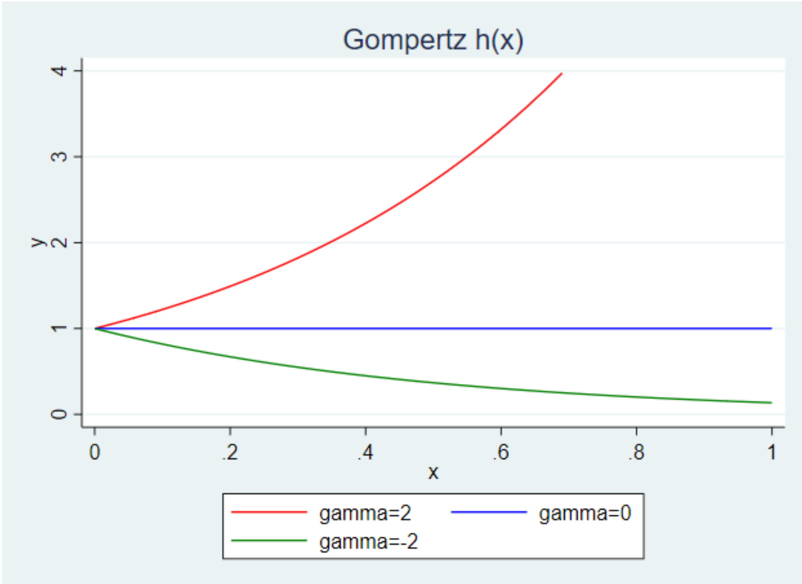


■ **Gompertz モデル**

モデル式 (1) において

$$h_0(t) = \exp(\gamma t), \quad g(\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x}) \tag{4}$$

とモデル化するもので、補助パラメータ γ も推定対象となります。ハザード関数の関数形状は次のようになります。



(2) AFT **モデル**

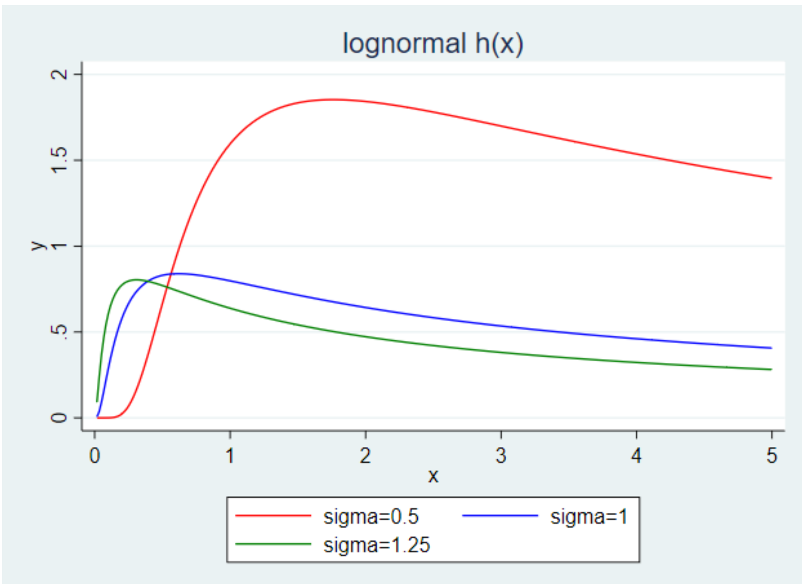
この流儀の場合には生存時間そのものがパラメータ化されます。ただし t 自体では $t \geq 0$ という制約が付いてしまうので、$\ln t$ を次のようにパラメータ化します。

$$\ln t = \boldsymbol{\beta}' \mathbf{x} + \epsilon \tag{5}$$

この場合、誤差 ϵ の分布関数 $f()$ をどう想定するかによってモデルが分かれることになります。

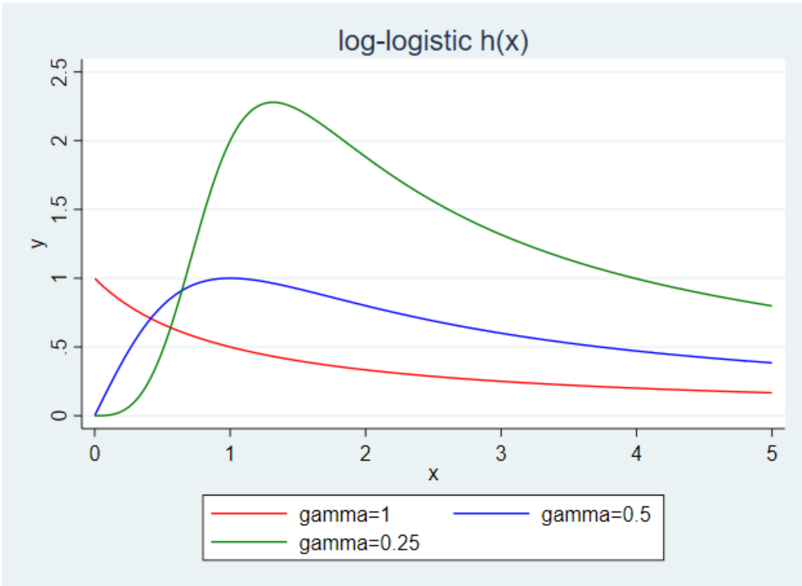
■ **対数正規モデル**

モデル式 (5) において ϵ の分布として正規分布を仮定した場合には対数正規モデルが構成されます。この場合のハザード関数の形状は途中に極大点を持つものとなります。



■ 対数ロジスティックモデル

モデル式 (5) において ϵ の分布としてロジスティック分布を仮定した場合には対数ロジスティックモデルが構成されます。この場合のハザード関数の形状は次のようなものとなります。



■ 一般化ガンマモデル

これについては表 1 を参照ください。

streg でサポートされている 6 種類のモデルを表にして示すと次のようになります。

表 1　パラメトリックモデル

分布	メトリック	生存関数	パラメータ化	補助パラメータ
Exponential	PH	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(\mathbf{x}_j \boldsymbol{\beta})$	
Exponential	AFT	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(-\mathbf{x}_j \boldsymbol{\beta})$	
Weibull	PH	$\exp(-\lambda_j t_j^p)$	$\lambda_j = \exp(\mathbf{x}_j \boldsymbol{\beta})$	p
Weibull	AFT	$\exp(-\lambda_j t_j^p)$	$\lambda_j = \exp(-p\mathbf{x}_j \boldsymbol{\beta})$	p
Gompertz	PH	$\exp\{-\lambda_j \gamma^{-1}(e^{\gamma t_j} - 1)\}$	$\lambda_j = \exp(\mathbf{x}_j \boldsymbol{\beta})$	γ
Lognormal	AFT	$1 - \Phi\left\{\frac{\log(t_j)-\mu_j}{\sigma}\right\}$	$\mu_j = \mathbf{x}_j \boldsymbol{\beta}$	σ
Loglogistic	AFT	$\{1 + (\lambda_j t_j)^{1/\gamma}\}^{-1}$	$\lambda_j = \exp(-\mathbf{x}_j \boldsymbol{\beta})$	γ
Generalized gamma				
if $\kappa > 0$	AFT	$1 - I(\gamma, u)$	$\mu_j = \mathbf{x}_j \boldsymbol{\beta}$	σ, κ
if $\kappa = 0$	AFT	$1 - \Phi(z)$	$\mu_j = \mathbf{x}_j \boldsymbol{\beta}$	σ, κ
if $\kappa < 0$	AFT	$I(\gamma, u)$	$\mu_j = \mathbf{x}_j \boldsymbol{\beta}$	σ, κ

なお、$\Phi(z)$ は標準正規分布の累積分布関数を意味します。また一般化ガンマモデルに関しては $\gamma = |\kappa|^{-2}$, $u = \gamma \exp(|\kappa| z)$, $I(a, x)$ は不完全ガンマ関数、$z = \text{sign}(\kappa)\{\log(t_j) - \mu_j\}/\sigma$ を意味します[*1]。

2. 基本的な用例

▷ Example 1: Weibull – PH **モデル**

`[ST]` **streg** の Example 1 には Example データセット `kva.dta` を用いた用例が紹介されています。

```
. use http://www.stata-press.com/data/r17/kva.dta
```
[*2]

```
(Generator experiment)
```

このデータセット中には発電機の耐久試験の結果が記録されています。

```
. list failtime load bearings
```
[*3]

	failtime	load	bearings
1.	100	15	0
2.	140	15	1
3.	97	20	0
4.	122	20	1
5.	84	25	0
6.	100	25	1
7.	54	30	0
8.	52	30	1
9.	40	35	0
10.	55	35	1
11.	22	40	0
12.	30	40	1

`load` は過負荷の量を、`bearings` は新しいタイプのベアリングを装着していたか否かを表す変数であり、`failtime` が故障が発生するまでの経過時間を意味しています。このデータセットは `failtime` を時間変数とする形で既に `stset` が済んでいます。

[*1] $\kappa = 1$ のときは Weibull 分布に、$\kappa = 1, \sigma = 1$ のときは指数分布に、$\kappa = 0$ のときは対数正規分布に等しくなります。
[*2] メニュー操作：File ▷ Example Datasets ▷ Stata 17 manual datasets と操作、Survival Analysis Reference Manual [ST] の streg の項よりダウンロードする。
[*3] メニュー操作： Data ▷ Describe data ▷ List data

```
. st
```

```
. st
-> stset failtime,

Survival-time data settings

         Failure event: (assumed to fail at time=failtime)
Observed time interval: (0, failtime]
     Exit on or before: failure
```

このデータセットに対して Weibull(PH) モデルを前提とした形で `streg` を実行してみます。

- Statistics ▷ Survival analysis ▷ Regression models ▷ Parametric survival models と操作
- Model タブ: Independent variables: `load bearings`
  Survival distribution and parameterization: Weibull PH

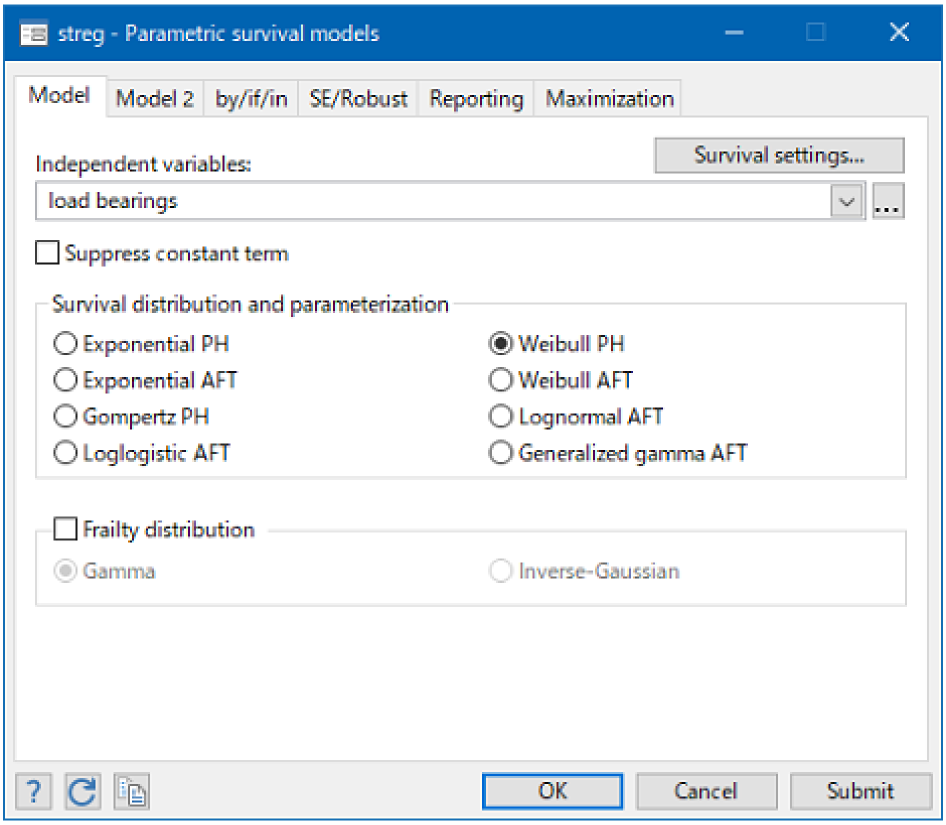


図 1 `streg` ダイアログ- Model タブ

```
. streg load bearings, distribution(weibull)

        Failure _d: 1 (meaning all fail)
  Analysis time _t: failtime

Fitting constant-only model:
Iteration 0:   log likelihood = -13.666193
Iteration 1:   log likelihood = -9.7427276
Iteration 2:   log likelihood = -9.4421169
Iteration 3:   log likelihood = -9.4408287
Iteration 4:   log likelihood = -9.4408286

Fitting full model:
Iteration 0:   log likelihood = -9.4408286
Iteration 1:   log likelihood =  -2.078323
Iteration 2:   log likelihood =  5.2226016
Iteration 3:   log likelihood =  5.6745808
Iteration 4:   log likelihood =  5.6934031
Iteration 5:   log likelihood =  5.6934189
Iteration 6:   log likelihood =  5.6934189

Weibull PH regression

No. of subjects =  12                                  Number of obs =     12
No. of failures =  12
Time at risk    = 896
                                                       LR chi2(2)    =  30.27
Log likelihood = 5.6934189                             Prob > chi2   = 0.0000

------------------------------------------------------------------------------
          _t | Haz. ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        load |   1.599315   .1883807     3.99   0.000     1.269616    2.014631
    bearings |   .1887995   .1312109    -2.40   0.016     .0483546    .7371644
       _cons |   2.51e-20   2.66e-19    -4.26   0.000     2.35e-29    2.68e-11
-------------+----------------------------------------------------------------
       /ln_p |   2.051552   .2317074     8.85   0.000     1.597414    2.505691
-------------+----------------------------------------------------------------
           p |   7.779969   1.802677                      4.940241    12.25202
         1/p |   .1285352   .0297826                      .0816192    .2024193
------------------------------------------------------------------------------
Note: _cons estimates baseline hazard.
```

[ST] **stcox** (*mwp-023*) の Example 1 では同じモデルを stcox によってフィットさせたわけですが、推定結果を対比させてみると次のようになります。

	ハザード比推定値	
	load	bearings
stcox	1.53	0.06
streg	1.60	0.19

streg の場合、ベースラインハザード関数の推定もフィットされており、Weibull 分布の形状パラメータが $\widehat{p} = 7.78$ と推定されています。この結果からすると、100 時間経過後のベアリングは 10 時間経過後のベアリングに比べ、$1,000,000$ 倍 ($(100/10)^{7.78-1}$) 以上こわれやすいことがわかります。 ◁

▷ Example 2: **係数値出力**

streg はデフォルトの場合、ハザード比の値を出力してきます。推定された係数値を $\widehat{\beta}$ としたとき、$\exp(\widehat{\beta})$ の値がハザード比推定値として表示されているわけです。直接係数の推定値 $\widehat{\beta}$ を出力させたい場合には次のようにコマンド入力します。

```
. streg, nohr
```

```
. streg, nohr

Weibull PH regression

No. of subjects =  12                                  Number of obs =     12
No. of failures =  12
Time at risk    = 896
                                                       LR chi2(2)    =  30.27
Log likelihood = 5.6934189                             Prob > chi2   = 0.0000

------------------------------------------------------------------------------
          _t | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        load |   .4695753   .1177884     3.99   0.000     .2387143    .7004363
    bearings |  -1.667069   .6949745    -2.40   0.016    -3.029194   -.3049443
       _cons |  -45.13191   10.60663    -4.26   0.000    -65.92053   -24.34329
-------------+----------------------------------------------------------------
       /ln_p |   2.051552   .2317074     8.85   0.000     1.597414    2.505691
-------------+----------------------------------------------------------------
           p |   7.779969   1.802677                      4.940241    12.25202
         1/p |   .1285352   .0297826                      .0816192    .2024193
------------------------------------------------------------------------------
```

この場合は引数を何も指定していないので表示の切替えのみで出力が得られます。Reporting タブ上の Do not report hazard ratios という項目を指定しても同じ結果が得られますが、その場合には推定が再実行される形になります。 ◁

▷ Example 3: Weibull – AFT **モデル**

評価版では割愛しています。

3. **複数** failure **事象**

評価版では割愛しています。

4. **モデルの選択**

評価版では割愛しています。

5. **補助パラメータのパラメータ化**

評価版では割愛しています。

6. **層化推定**

評価版では割愛しています。

7. **非共用** frailty **モデル**

評価版では割愛しています。

8. **共用** frailty **モデル**

評価版では割愛しています。

■