

stcox - Cox 比例ハザードモデル 【評価版】

stcox は最尤法を用いて比例ハザードモデルのフィットを行います。パラメトリックやノンパラメトリックな生存時間分析との対比において、セミパラメトリックな分析モデルと呼ばれることもあります。

1. Cox 比例ハザードモデル	
2. 基本的な用例	Example 1
	Example 2
3. Tied failures の扱い	
4. 時間変動型共変量	Example 3
	Example 4
	Example 5
5. ロバストな分散推定	Example 6
6. 複数 failure データ	Example 7
7. 層化推定	Example 8
8. ベースラインハザード関数	Example 9
9. 共用 frailty モデル	Example 10
補足 1	

1. Cox 比例ハザードモデル

Cox 比例ハザードモデル (Cox proportional hazards model) の基本的なモデル式は

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \cdots + \beta_k x_k) \quad (1)$$

のように表現され、ベースラインハザード関数 $h_0(t)$ と共変量依存部 $\exp(\mathbf{x}_j\beta)$ とが掛け合わされた形になっています。 $h_0(t)$ の関数形に対して何の仮定も置かずに推定が実行し得るという点に特徴があります。この点はノンパラメトリックな推定手法と性格を一にするものがありますが、 β についてはパラメトリックな推定が行われます。Cox 比例ハザードモデルがセミパラメトリックな推定手法と呼ばれる所以はこの点にあります。なお、stcox 自体からは $h_0(t)$ についての情報は得られませんが、postestimation 機能である predict コマンドを用いることによってベースライン生存関数 $S_0(t)$ や累積ハザード関数 $H_0(t)$ の推定値を求めることができます (図 1 参照)。

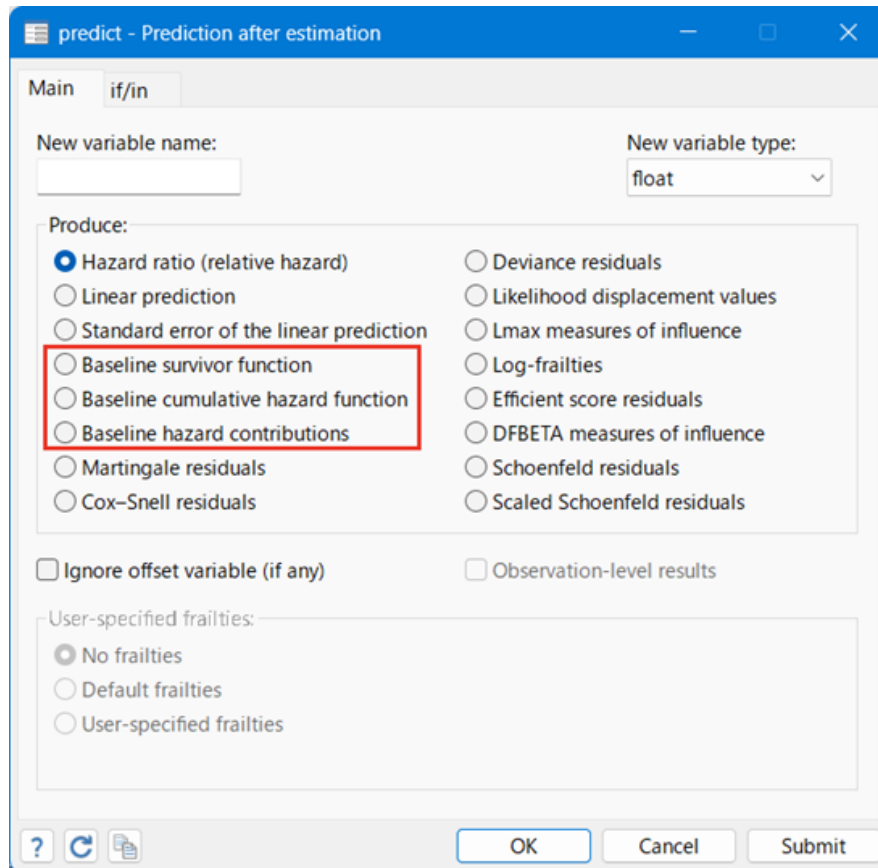


図 1 predict ダイアログ

これに対し `strata()` オプションを指定した場合のモデル式は

$$h_i(t) = h_{0i}(t) \exp(\beta_1 x_1 + \cdots + \beta_k x_k) \quad (2)$$

のようになります。この場合には年齢群等のグループ i ごとに別個のベースラインハザード関数を許容した形で推定が行われます（層化推定）。ただし $\exp(\mathbf{x}_j \beta)$ の部分はグループによらず共通である点に注意してください。

さらに `shared()` オプションの指定も可能ですが、その場合のモデル式は

$$h_{ij}(t) = h_0(t) \alpha_i \exp(\mathbf{x}_{ij} \beta) \quad (3)$$

のようになります。 $\nu_i = \log \alpha_i$ と書くことにすると (3) 式は

$$h_{ij}(t) = h_0(t) \exp(\mathbf{x}_{ij} \beta + \nu_i) \quad (4)$$

と書き直せることから、グループ固有の変量効果 (random effects) をモデル化したものと言え、共用 frailty (異質性) モデル (shared frailty model) と呼ばれます。グループ i に属するデータはすべて 1 つの異質性 ν_i を共用することから、グループ内相関をモデル化する際に利用されます。

2. 基本的な用例

▷ Example 1: 途中打ち切りなしのデータ

[ST] `stcox` 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 https://www.stata-press.com/data/r19/kva.dta *1
(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 *2
```

	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` が済んでいますが、一応確認しておきます。

```
. 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
```

*1 メニュー操作 : File ▷ Example Datasets ▷ Stata 19 manual datasets と操作、Survival Analysis Reference Manual [ST] の `stcox` の項よりダウンロードする。

*2 メニュー操作 : Data ▷ Describe data ▷ List data

このデータに対して `stcox` を実行してみます。モデル式 (1) において想定する共変量は $x_1 = \text{load}$ 、 $x_2 = \text{bearings}$ とするわけです。

```
. stset, noshow
```

- Statistics ▸ Survival analysis ▸ Regression models ▸ Cox proportional hazards (PH) model と操作
- Model タブ: Independent variables: load bearings

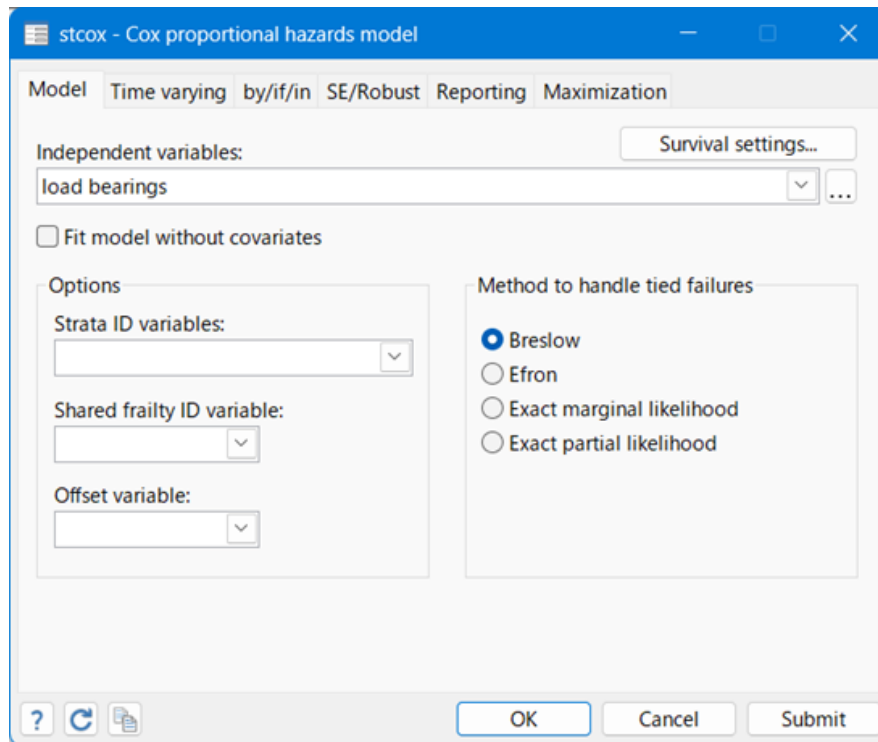


図 2 `stcox` ダイアログ- Model タブ

```
. stcox load bearings
```

```
Iteration 0:  Log likelihood = -20.274897
Iteration 1:  Log likelihood = -10.515114
Iteration 2:  Log likelihood = -8.8700259
Iteration 3:  Log likelihood = -8.5915211
Iteration 4:  Log likelihood = -8.5778991
Iteration 5:  Log likelihood = -8.577853
Refining estimates:
Iteration 0:  Log likelihood = -8.577853
```

Cox regression with Breslow method for ties

No. of subjects = 12

Number of obs = 12

No. of failures = 12

Time at risk = 896

LR chi2(2) = 23.39

Log likelihood = -8.577853

Prob > chi2 = 0.0000

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]	
load	1.52647	.2188172	2.95	0.003	1.152576	2.021653
bearings	.0636433	.0746609	-2.35	0.019	.0063855	.6343223

stcox はデフォルトの場合、ハザード比の値を出力してきます。負荷を固定させて考えた場合、bearings の値を 0 から 1 に変化させたときにハザード比が 0.064 に下がることが示されています。stcox の出力を係数表示に切り替えるには次のようにコマンド入力します。

```
. stcox, nohr
```

```
. stcox, nohr
```

Cox regression with Breslow method for ties

No. of subjects = 12

Number of obs = 12

No. of failures = 12

Time at risk = 896

LR chi2(2) = 23.39

Log likelihood = -8.577853

Prob > chi2 = 0.0000

_t	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
load	.4229578	.1433485	2.95	0.003	.1419999	.7039157
bearings	-2.754461	1.173115	-2.35	0.019	-5.053723	-.4551981

この場合には引数を何も指定していないので表示が切り替わるだけで stcox による推定プロセスが再実行されるわけではありません^{*3}。 $\hat{\beta}_2 = -2.754461$ と推定されているわけですが、ハザード比による推定値とは

$$\exp(\hat{\beta}_2) = 0.0636433$$

という関係がある点に注意してください。

◀

^{*3} nohr オプションは Reporting タブ上の Report coefficients, not hazard ratios という項目に対応します。

▷ Example 2: 途中打ち切りありのデータ

今度は Example データセット `drugtr.dta` を使用します。

```
. use https://www.stata-press.com/data/r19/drugtr.dta *4
(Patient survival in drug trial)
```

このデータセットにはある薬剤の効果に関する 48 件のデータが患者当り単一のレコード形式で記録されています。

```
. list studytime age drug died if _n <= 4 | _n >= (_N-3), abbreviate(9) separator(4)
```

	studytime	age	drug	died
1.	1	61	0	1
2.	1	65	0	1
3.	2	59	0	1
4.	3	52	0	1
45.	33	60	1	1
46.	34	62	1	0
47.	35	48	1	0
48.	39	52	1	0

この場合、`drug = 1` が薬剤を投与した患者を表し、0 はプラセボ (placebo) を意味します。また、`studytime` は生存月数を、`age` は調査開始時点における患者の年齢を表す変数です。Example 1 で用いた `kva.dta` の場合にはすべてのデータが故障という事象で終わっていたため、`stset` に際して特に `failure` 変数の指定はなかったわけですが、`drugtr.dta` の場合、観察終了時点で生存していた患者もいるため (途中打ち切り (censored)) 死亡したのか否かを示す `died` という変数が加わっています。念のため `stset` に際しての指定内容を確認しておきます。

```
. st
```

```
. st
-> stset studytime, failure(died)

Survival-time data settings

      Failure event: died!=0 & died<.
Observed time interval: (0, studytime]
Exit on or before: failure
```

*4 メニュー操作 : File ▷ Example Datasets ▷ Stata 19 manual datasets と操作、Survival Analysis Reference Manual [ST] の `stcox` の項よりダウンロードする。

時間変数としては `studytime` が、failure 変数としては `died` が指定されていたこと、及び死亡した患者数が 48 人中 31 人であったことが読み取れます。このデータに対して `drug` と `age` を共変量とした形で `stcox` を実行してみます。

- Statistics ▸ Survival analysis ▸ Regression models ▸ Cox proportional hazards (PH) model と操作
- Model タブ: Independent variables: `drug age`

```
. stcox drug age

      Failure _d: died
  Analysis time _t: studytime

Iteration 0:  Log likelihood = -99.911448
Iteration 1:  Log likelihood = -83.551879
Iteration 2:  Log likelihood = -83.324009
Iteration 3:  Log likelihood = -83.323546
Refining estimates:
Iteration 0:  Log likelihood = -83.323546

Cox regression with Breslow method for ties

No. of subjects = 48                Number of obs = 48
No. of failures = 31
Time at risk    = 744

Log likelihood = -83.323546          LR chi2(2)    = 33.18
                                      Prob > chi2    = 0.0000
```

	_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]
drug		.1048772	.0477017	-4.96	0.000	.0430057 .2557622
age		1.120325	.0417711	3.05	0.002	1.041375 1.20526

`drug` の値を 0 から 1 に変化させたときにハザード比が 0.105 に下がることが示されています。一方、変数 `age` については 1 年加齢するごとにハザード比が 1.120 増加するという結果になっていますが、5 年単位の変化を求めたい場合には、次のように `age` の内容を変換することで対応できます。

```
. replace age = age/5 *5
variable age was byte now float
(48 real changes made)
```

*5 メニュー操作：Data ▸ Create or change data ▸ Change contents of variable

```
. stcox drug age, nolog *6
```

```
. stcox drug age, nolog

      Failure _d: died
    Analysis time _t: studytime

Cox regression with Breslow method for ties

No. of subjects =  48                Number of obs =    48
No. of failures =  31
Time at risk    = 744

Log likelihood = -83.323544          LR chi2(2)    =  33.18
                                      Prob > chi2    =  0.0000
```

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]
drug	.1048772	.0477017	-4.96	0.000	.0430057 .2557622
age	1.764898	.3290196	3.05	0.002	1.224715 2.543338

◁

^{*6} `nolog` オプションの指定は Maximization タブ上で行えます。

3. Tied failures の扱い

評価版では割愛しています。

4. 時間変動型共変量

評価版では割愛しています。

5. ロバストな分散推定

評価版では割愛しています。

6. 複数 failure データ

評価版では割愛しています。

7. 層化推定

評価版では割愛しています。

8. ベースラインハザード関数

評価版では割愛しています。

9. 共用 frailty モデル

評価版では割愛しています。

補足 1 – グラフ作成コマンド操作

評価版では割愛しています。

