補論

平成 29 年 12 月 13 日更新

7.8 コックス回帰モデル

コックス (Cox) 回帰モデルでは, 説明変数 $X_{1i}, \ldots, X_{pi}$ を用いてハザード率関数が

$$h(t) = h_0(t) \exp(\beta_1 X_{1i} + \ldots + \beta_p X_{pi})$$

であると仮定します. $h_0(t)$ は基準ハザード率関数で特に関数形は仮定しないので, セミパラメトリックモデルとも呼ばれます [1]. 累積ハザード関数 $H_0(t)$ を

$$H_0(t) \quad = \quad \int_0^t h_0(s) ds$$

とおくと, T の確率密度関数 $f(t)$ は

$$f(t) \quad = \quad h_0(t) \exp\left\{\beta_1 X_{1i} + \ldots + \beta_p X_{pi} - H_0(t) \exp(\beta_1 X_{1i} + \ldots + \beta_p X_{pi})\right\},$$

となります. また δ_i を打ち切りのある時に 0, ないときに 1 とすると, 尤度関数 L は

$$\begin{aligned} L \quad = \quad & \prod_{i=1}^{n} \left[h_0(t_i) \exp\left\{\beta_1 X_{1i} + \ldots + \beta_p X_{pi} - H_0(t_i) \exp(\beta_1 X_{1i} + \ldots + \beta_p X_{pi})\right\}\right]^{\delta_i} \\ & \times \left[\exp\left\{-H_0(t_i) \exp(\beta_1 X_{1i} + \ldots + \beta_p X_{pi})\right\}\right]^{1-\delta_i} \end{aligned}$$

となります.

ここで実際に (打ち切りではなく) 継続時間の終了が観測された時点を $T_1, \ldots, T_m$ とし, また継続時間の終了は観測値ごとに異なる (**タイ**がない, といいます) とします. さらに R_i を T_i 時点の直前まで継続している観測値の集合 (**リスク集合**) とし, $X_{j(i)}$ を T_i 時点に継続時間が終了した観測値の説明変数とします. このとき,

$$L(\beta) = \prod_{i=1}^{m} \frac{\exp\left\{\beta_1 X_{1(i)} + \ldots + \beta_p X_{p(i)}\right\}}{\sum_{j \in R_i} \exp\left\{\beta_1 X_{1j} + \ldots + \beta_p X_{pj}\right\}}$$

を $\beta = (\beta_1, \ldots, \beta_p)$ の部分尤度 (partial likelihood) といいます. この部分尤度を最大にするような推定量 $\hat{\beta}$ は漸近的に正規分布にしたがうことが知られています. もし継続時間の終了が複数の観測値で同時に起こる (タイがある) 場合には調整が必要となりますが, その詳細は割愛します [2].

例 7.4 のように**データ 7.3** を用いて推定してみましょう. まず分析するデータを **Stata 7.9** のように継続時間として定義します. 続けてコックス回帰モデルをあてはめます.

[1] コックス比例ハザード (Cox proportional hazard) モデルとも呼ばれます. このモデルでは 2 つの個体のハザード $h_1(t)$ と $h_2(t)$ の比が

$$\frac{h_1(t)}{h_2(t)} = \frac{h_0(t) \exp(\beta_1 X_{11} + \ldots + \beta_p X_{p1})}{h_0(t) \exp(\beta_1 X_{12} + \ldots + \beta_p X_{p2})} = \frac{\exp(\beta_1 X_{11} + \ldots + \beta_p X_{p1})}{\exp(\beta_1 X_{12} + \ldots + \beta_p X_{p2})}$$

のように時間 t に依存せず一定となります.

[2] 例えば大橋・浜田 (1995) や Kleinbaum and Klein (2012) を参照してください.

Stata 7.15 コックス回帰モデル

統計 (S) ► 生存分析 ► 回帰モデル ►**Cox** 比例ハザードモデル

とし, 出てきた画面で独立変数に x を選びます. さらにレポートタブで, ハザード比でなく係数を表示するに ✓ をいれて **OK** をおします.

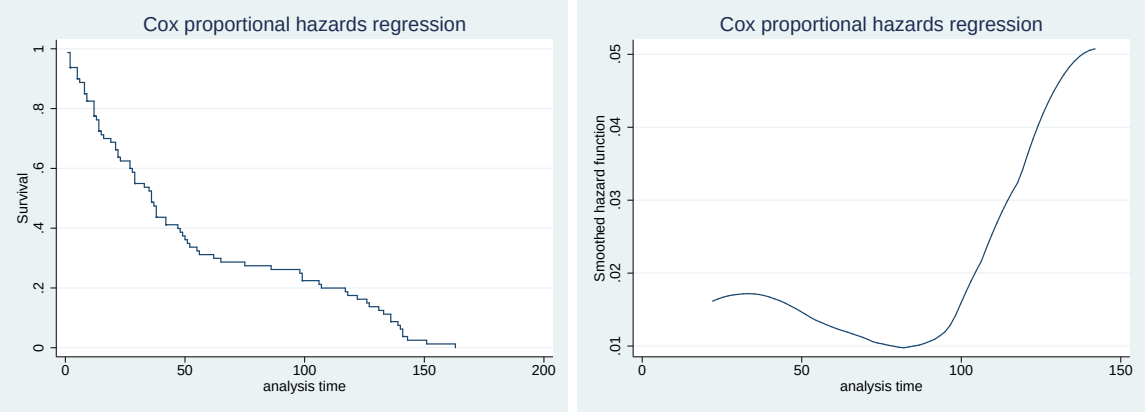
表 7.11: コックス回帰モデルの推定結果

```
. stcox x, nohr
         failure _d:  status == 1 ①
   analysis time _t:  y
Cox regression -- Breslow method for ties
No. of subjects =           80 ②      Number of obs    =          80
No. of failures =           80 ③
Time at risk    =         4283
                                      LR chi2(1)       =        0.06
Log likelihood  =   -274.65622        Prob > chi2      =      0.8007
------------------------------------------------------------------------------
_t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---+--------------------------------------------------------------------------
 x |   .6092947 ④ 2.424978     0.25   0.802 ⑤  -4.143574    5.362164
------------------------------------------------------------------------------
```

推定結果の表 7.11 において, status が 1 のときにストライキ継続時間の終了が観測されており (①), 観測対象の 80 社 (②) のなかで, 80 社すべてのストライキ継続時間の終了が観測された (③) ことを表しています. 回帰係数の推定値は $\hat{\beta}_1 = 0.609$ (④), p 値は $0.802 > 0.05$ であることから, 有意水準 5%で $H_0 : \beta_1 = 0$ を棄却することができません. つまりこの 80 社のデータでは, 説明変数 X (米国製造業における産業生産の対数値) はストライキの継続時間を説明するのには有用ではないということになります.

得られた推定値をもとに, **Stata 7.11** のように説明変数の平均値で評価した生存関数と, ハザード率関数を描いてみましょう.

図 7.15: 生存関数とハザード率関数



最後に Stata コマンドをにまとめておきます.

Stata 7.16 継続時間のコックス回帰モデルのプログラム

```
import delimited C:¥strike.csv
stset y, failure(status==1) scale(1)
stcox x, nohr
stcurve, survival
stcurve, hazard
```

参考図書

- 大橋靖雄・浜田知久馬 (1995)『生存時間解析—SAS による生物統計』東京大学出版会
- Kleinbaum, D. G. and Klein, M. (2012) *Survival Analysis.* Third edition. Springer.