113-1 Survival Analysis - Midterm Exam.

2014.11.05

1. The following data consists of the times to death of 12 bone marrow transplant patients.

$$6,6,6,7,10,13,6^+,9^+,10^+,11^+,17^+,19^+$$

(a) (10pts) Use the Kaplan-Meier estimator to estimate the survival function. Plot the estimated survival curve.

(b) (5pts) Estimate the survival rate $\hat{S}(8)$

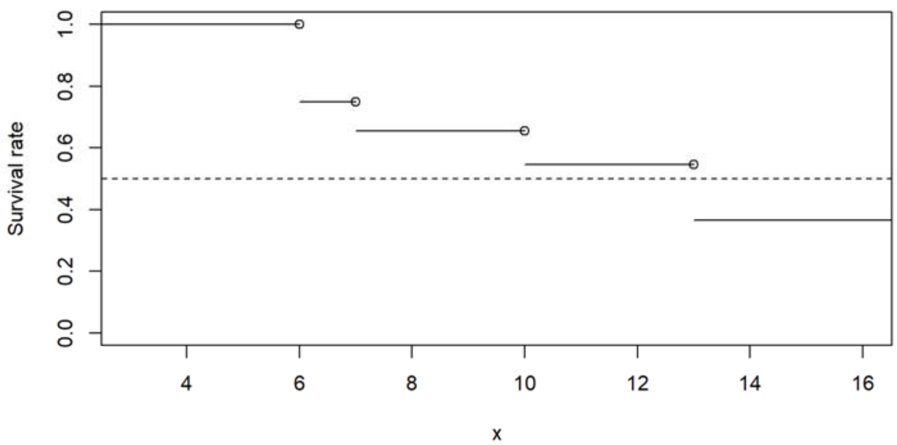
(c) (10pts) Estimate the variance of the survival rate $\hat{S}(8)$ by Greenwood's formula.

(d) (10pts) Estimate the cumulative hazard function by Nelson-Aalen estimator. Plot the estimated cumulative hazard function.

(e) (5pts) Find the median lifetime.

(a)

t_i	y_i	d_i	$\hat{S}(t_i)$
6	12	3	$1-\frac{3}{12}=0.75$
7	8	1	$0.75\cdot\left(1-\frac{1}{8}\right)=0.656$
10	6	1	$0.656\cdot\left(1-\frac{1}{6}\right)=0.547$
13	3	1	$0.547\cdot\left(1-\frac{1}{3}\right)=0.365$



(b) $\hat{S}(8)=\hat{S}(7)=0.656$

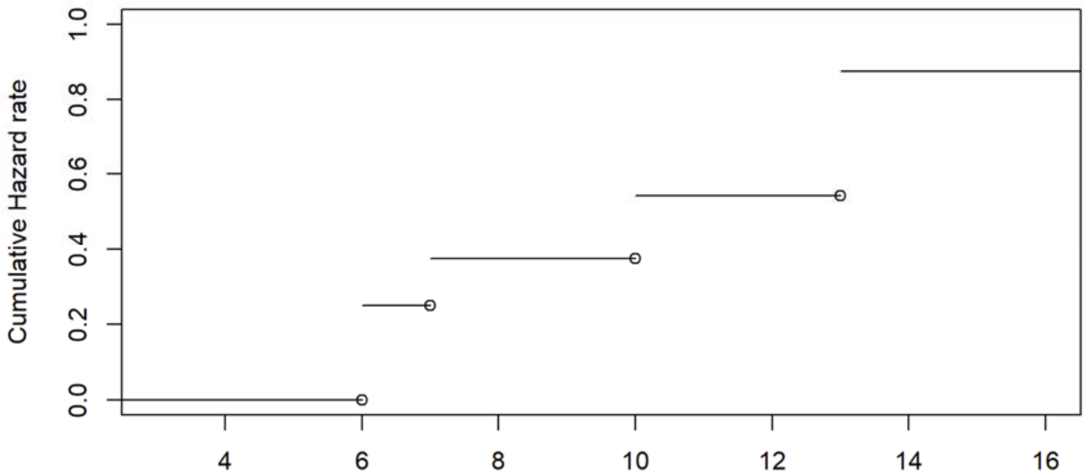
(c)

t_i	y_i	d_i	$\hat{S}(t_i)$	$\sum \frac{d_i}{y_i(y_i - d_i)}$	$\hat{V}\left(\hat{S}(t_i)\right)$
6	12	3	0.75	$\frac{3}{12(12-3)} = 0.028$	$0.75^2 \cdot 0.028$
7	8	1	0.656	$0.028 + \frac{1}{8(8-1)} = 0.046$	$0.656^2 \cdot 0.046 \approx 0.020$

$$\hat{V}\left(\hat{S}(8)\right) \approx 0.020$$

(d)

t_i	y_i	d_i	$\widehat{H}(t_i)$
6	12	3	$\frac{3}{12} = 0.25$
7	8	1	$0.25 + \frac{1}{8} = 0.375$
10	6	1	$0.375 + \frac{1}{6} = 0.542$
13	3	1	$0.542 + \frac{1}{3} = 0.875$



(e)

Median lifetime:13

2. Data was collected on bone marrow transplant patients for the non-Hodgkin's lymphoma(NHL, 非何杰金氏淋巴瘤) patients and the Hodgkin's lymphoma (HOD, 何杰金氏淋巴瘤) patients. Times to death or relapse are given for 23 NHL patients, 11 receiving an allogenic (Allo, 異體骨移植) transplant from an HLA-matched sibling donor and 12 patients receiving an autologous (Auto, 異體骨移植) transplant. Also, data on 20 HOD patients, 5 receiving an allogenic (Allo) transplant from an HLA-matched sibling donor and 15 patients receiving an autologous (Auto) transplant is given.

(a) (10pts)Treating NHL Allo as the baseline hazard function, state the appropriate coding which would allow the investigator to test for an interaction between type of transplant and disease type using main effects and interaction terms.

Hint: There are 2 main effects: disease type and transplant type. The main effects of disease type (NHL or HOD) would be the coefficients of $Z = I(HOD)$

(b) (10pts)Suppose that we have the following model for the hazard rates in the four groups:

$$h(t|NHL, Allo) = h_0(t)$$
$$h(t|HOD, Allo) = h_0(t)\exp(2)$$
$$h(t|NHL, Auto) = h_0(t)\exp(1.5)$$
$$h(t|HOD, Auto) = h_0(t)\exp(.5)$$

What are the risk coefficients in (a).

(a) Let $Z_1 = I(HOD), Z_2 = I(Auto), Z_3 = Z_1 \times Z_2 = I(Z_1 = 1, Z_2 = 1)$

$$h(t|Z_1, Z_2, Z_3) = h_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3)$$

(b)

$$\exp(2) = \frac{h(t|HOD, Allo)}{h(t|NHL, Allo)} = \frac{h(t|Z_1 = 1, Z_2 = 0, Z_3 = 0)}{h(t|Z_1 = 0, Z_2 = 0, Z_3 = 0)} = \exp(\beta_1)$$

$$\exp(1.5) = \frac{h(t|NHL, Auto)}{h(t|NHL, Allo)} = \frac{h(t|Z_1 = 0, Z_2 = 1, Z_3 = 0)}{h(t|Z_1 = 0, Z_2 = 0, Z_3 = 0)} = \exp(\beta_2)$$

$$\exp(0.5) = \frac{h(t|HOD, Auto)}{h(t|NHL, Allo)} = \frac{h(t|Z_1 = 1, Z_2 = 1, Z_3 = 1)}{h(t|Z_1 = 0, Z_2 = 0, Z_3 = 0)}$$

$$= \exp(\beta_1 + \beta_2 + \beta_3)$$

$$\beta_1 = 2, \qquad \beta_2 = 1.5, \qquad \beta_3 = 0.5 - 2 - 1.5 = -3$$

3. 考慮 244 筆臨床試驗資料，我們想要研究臨床試驗的發表時間和那些因素相關。資料變數定義如下:

   a. posres: 臨床試驗的結果是否是顯著的(正向的) 1=Yes, 0=No.
   b. clinend: Did the trial focus on a clinical endpoint? (臨床試驗是否著重在最後試驗結束時間時的結果?) 1=Yes, 0=No.
   c. budget: 臨床試驗預算, in millions of dollars.
   d. time: Time to publication (發表時間), in months.
   e. status: Whether or not the trial was published: 1=Published, 0=Not yet published.

   我們採用 Proportional Hazard Model (PHM)，R 軟體的分析結果如下:

```
> fit.pub<-coxph(Surv(time,status)~posres+clinend+budget,data=Publication)
> summary(fit.pub)
Call:
coxph(formula = Surv(time, status) ~ posres + clinend + budget,
    data = Publication)

  n= 244, number of events= 156

            coef exp(coef) se(coef)     z Pr(>|z|)
posres  0.572950  1.773492 0.178380 3.212  0.00132 **
clinend 1.886879  6.598739 0.218355 8.641  < 2e-16 ***
budget  0.003371  1.003376 0.001680 2.007  0.04477 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        exp(coef) exp(-coef) lower .95 upper .95
posres      1.773     0.5639     1.250     2.516
clinend     6.599     0.1515     4.301    10.123
budget      1.003     0.9966     1.000     1.007

Concordance= 0.719  (se = 0.022 )
Likelihood ratio test= 74.22  on 3 df,   p=5e-16
Wald test            = 88.7  on 3 df,   p=<2e-16
Score (logrank) test = 114.6  on 3 df,   p=<2e-16
```

   (a) (5pts)根據以上的結果，寫下 Proportional Hazard Model
   (b) (5pts)試驗結果 (posres) 分為正向或負向，哪一種會減少試驗的發表時間?
   (c) (5pts)根據 R 軟體的分析結果，哪一個變數對發表時間的影響最為顯著?並解釋此變數的係數。

(a) Let T be the time to publication, $Z_1 = posres, Z_2 = clinend, Z_3 = budget$

$$\hat{h}(t|z_1, z_2, z_3) = h_0(t)\exp\{0.573z_1 + 1.887z_2 + 0.003z_3\}$$

(b) 試驗結果正向 ($posres = 1$ or $Z_1 = 1$) 會減少試驗的發表時間。

(c) Clinend (Z_2) 的係數估計值 p-vlaue 最小，因此影響最為顯著。

解釋: 在其他變數不變的情況下，clinend =1 對 clinend =0 的 hazard rario (relative risk) 為 $\exp(1.887) = 6.599$。

4. Given the survival function $S(x) = \exp\left(-\theta x^{\beta}\right), \theta > 0, \beta > 0$. Derive the Survival function.
   (a) (10pts) Find the density function.
   (b) (5pts) Find the hazard function.

(a)

$$f(x) = -\frac{dS(x)}{dx} = \theta\beta x^{\beta-1}\exp\left(-\theta x^{\beta}\right)$$

(b)

$$h(x) = \frac{f(x)}{S(x)} = \theta\beta x^{\beta-1}$$

5. (10pts)Let T be the random variable associated with the survival time. The conditional hazard function follows a PHM:

$$h(t|z) = h_0(t)\exp(\beta z)$$

where $h_0(t)$ corresponds to its baseline hazard function. Considering baseline density function $f_0(t) = \alpha v t^{v-1}\exp(-\alpha t^{v})$, which is a Weibull density function with shape parameter v and the scale parameter $w = \alpha^{-1/v}$. Show that T given Z follows a Weibull distribution. Write down its corresponding shape and scale parameter.

Sol:

$$S_0(t) = \int_t^{\infty} \alpha v x^{v-1}\exp(-\alpha x^{v})\,dx = -\exp(-\alpha x^{v})|_t^{\infty} = \exp(-\alpha t^{v})$$

$$h_0(t) = \frac{f_0(t)}{S_0(t)} = \alpha v t^{v-1}$$

$$H_0(t) = \int_0^t \alpha v s^{v-1} ds = \alpha s^{v}|_0^t = \alpha t^{v}$$

$$h(t|z) = h_0(t)\exp(\beta z)$$

$$H(t|Z) = \exp(\beta z)\,H_0(t)$$

$$S(t|z) = \exp[-H(t|Z)] = \exp[-\exp(\beta z)\,H_0(t)] = \exp[-\exp(\beta z)\,\alpha t^{v}]$$
$$= \exp[-\alpha^{*} t^{v}],$$

where $\alpha^{*} = \alpha\exp(\beta z)$

Therefore, $T|Z$ follows a Weibull distribution with shape parameter v and scale parameter $(\alpha\exp(\beta z))^{-1/v}$