Survival Analysis Homework5

Zhuodiao Kuang

2023-10-20

Problem1

Construct 95% CI for the hazard ratio from a PH model shown in the follow table for the risk reduction between two treatment groups.

Analysis of Maximum Likelihood Estimates - discrete								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
trt	6-MP	1	-1.62822	0.43313	14.1316	0.0002	0.196	trt 6-MP

According to the table, the parameter estimate and the standard error are -1.628 and 0.433, respectively.

We've already learned that the 95% CI confidence interval for the coefficient estimates is as follows:

$$[\hat{\beta} - z_{0.975}se(\hat{\beta}), \hat{\beta} + z_{0.975}se(\hat{\beta})]$$

Then, from $\frac{h(t|Z=1)}{h(t|Z=0)} = e^{\beta}$, we can obtain the hazard ratio as 0.196, with its 95% CI:

$$[e^{\hat{\beta} - z_{0.975}se(\hat{\beta})}, e^{\hat{\beta} + z_{0.975}se(\hat{\beta})}]$$

```r
est <- -1.628
se <- 0.433
z <- qnorm(0.975)
lq <- exp(est-z*se)
uq <- exp(est+z*se)
lq;uq
```

```
## [1] 0.08402303
```

```
## [1] 0.4587107
```

So, the 95% CI for the hazard ratio is $[0.084, 0.459]$.

Problem2

The observed survival data $(T_i, \Delta_i, Z_i) i = 1, 2, 3, 4, 5, 6$ are (16,1,1), (20,0,1), (12,1,0), (14,0,0), (11,1,0), (9,1,1). Please construct the partial likelihood.

$$L_p(\beta) = \prod_{i=1}^{n} \{ \frac{e^{\beta z_i}}{\sum_{l \in R(t_i) e^{\beta' z_i}}} \}^{\Delta_i}$$

$$\because \Delta_1 = \Delta_3 = \Delta_5 = \Delta_6 = 1$$

$$\because \Delta_2 = \Delta_4 = 0, n = 6$$

$$\therefore L_p(\beta) = \frac{e^{\beta}}{e^{\beta} + e^{\beta}} \frac{1}{1 + 1 + e^{\beta} + e^{\beta}} \frac{1}{1 + 2 + 2e^{\beta}} \frac{e^{\beta}}{e^{\beta} + 3 + 2e^{\beta}}$$

$$= \frac{1}{2} \frac{1}{2 + 2e^{\beta}} \frac{1}{3 + 2e^{\beta}} \frac{e^{\beta}}{3 + 3e^{\beta}}$$

Problem3

Show that PH model score test is the same as the log-rank test for an indicator covariate when there is no ties.

In the log-rank test, $L = \sum_{i=1}^{k} (d_{0i} - e_{0i})$ When there is no ties, which means $d_i = 1$ for any $i = 1, 2, ..., n$.

So, we got $var(L) = \sum_{i=1}^{k} \frac{n_{0i} n_{1i}}{n_i^2}$

To get a statistic to test the hypothsis, we build $\frac{L}{\sqrt{var(L)}} \sim N(1, 0)$

In the score test, we denote d_j as Δ_j, so we get $U(0)$ and $I(0)$ as follows:

$$U(0) = \sum_{i=1}^{n} \Delta_i \{ z_i - \frac{\sum_{l \in R(t_i)} z_l}{\sum_{l \in R(t_i)} 1} \}$$

$$= \sum_{i=1}^{n} d_i \{ z_i - \frac{\sum_{l \in R(t_i)} z_l}{n_i} \}$$

$$= \sum_{j=1}^{J} \{ d_{1j} - d_j \frac{n_{1j}}{n_j} \}$$

$$= \sum_{j=1}^{J} \{ d_{1j} - e_{1j} \} = L$$

$$I(0) = \sum_{i=1}^{n} \Delta_i \{ \frac{\sum_{l \in R(t_i)} z_l^2}{\sum_{l \in R(t_i)} 1} - (\frac{\sum_{l \in R(t_i)} z_l}{\sum_{l \in R(t_i)} 1})^2 \}$$

$$= \sum_{j=1}^{J} d_j \{ \frac{n_{1j}}{n_j} - (\frac{n_{1j}}{n_j})^2 \}$$

$$= \sum_{j=1}^{J} \{ \frac{n_{1j}}{n_j} - (\frac{n_{1j}}{n_j})^2 \} = var(L)$$

Therefore, score test = log-rank test without ties

$$\frac{U(0)}{\sqrt{I(0)}} = \frac{L}{\sqrt{var(L)}} \sim N(0,1)$$