Statistical Science Master
Exam survival analysis

June 14, 2013, 10:00-13:00

Include all the `R` code that you used and attach any figures in `.pdf` format. It should be clear which code belongs to which problem.

Exercise 1 — Suppose that in a laboratory experiment 10 mice are exposed to carcinogens. The investigator decides to stop the study after 5 deaths. The survival times of the 10 mice are

```
>  4    5    8    9   10   10+ 10+ 10+ 10+ 10+
```

Assume that the death times follows and exponential distribution with parameter λ

$$f(x) = \lambda \exp(-\lambda x), \quad x > 0, \quad \lambda > 0 \ .$$

a. In case of right censored observation the likelihood is as follows

$$L(\lambda) = \prod_{i=1}^{n} [f(x_i)]^{\delta_i} [S(x_i)]^{1-\delta_i} \ .$$

By using pen and paper estimate the survival rate $\hat{\lambda}$ (note that all censored observations have the same censoring times).

b. Find the $100(1-\alpha)\%$ confidence interval for λ by applying

$$\frac{\hat{\lambda}\chi^2_{2r,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi_{2r,\alpha/2}}{2r}$$

since $2r\lambda/\hat{\lambda} \sim \chi_{2r}$; r is the number of deaths; use the `R` command `qchisq` to compute the quantile.

c. Estimate the mean survival time (hint we have an exponential distribution)

$$E(X) = \int_0^{\infty} S(t)dt \ .$$

d. Estimate the mean residual life time

$$mrl(x) = \frac{\int_x^\infty S(t)dt}{S(x)} \ .$$

e. Estimate the probability that a mouse exposed to the same carcinogen will survive longer than 8 weeks. You can use the `R` command `pexp`.

f. Plot the log-likelihood function $\ell(\lambda)$ for $0 < \lambda \leq 0.4$ and locate the maximum likelihood estimator $\hat{\lambda}$.

Exercise 2 — Given the small data set of censored survival data in the table below with 2 competing events estimate the following quantities:

Time	No. risk	Cause
0	10	0
2	10	1
4	9	0
5	8	0
6	7	2
7	6	1
8	5	0
10	4	1
13	3	2
16	2	0
18	1	1

a. Overall survival function estimated by Kaplan-Meier

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) \ .$$

b. Overall cumulative hazard

$$\widehat{H}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j} \ .$$

c. Failure rate cause $\widehat{\lambda}_k(t_j)$ for $k = 1, 2$ and the cause specific cumulative hazards $H_k(t)$

$$\widehat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j} \quad ; \quad \widehat{H}_k(t) = \sum_{j:t_j \leq t} \widehat{\lambda}_k(t_j) \ .$$

d. The probability of failing from cause k at t_j, $p_k(t_j) = P(T = t_j, D = k)$ estimated as

$$\widehat{p}_k(t_j) = \widehat{\lambda}_k(t_j)\widehat{S}(t_{j-1}) \ .$$

e. The cumulative incidence function F_k for $k = 1, 2$

$$\widehat{F}_k(t) = \sum_{j:t_j \leq t} \widehat{p}_k(t_j) \ .$$

You can solve the exercise with pen and paper or by writing `R` code. Check your result by using the function `cuminc` from the `mstate` package.

Exercise 3 — Centralization of surgery in cancer is a hot topic in the Netherlands and in the rest of Europe. In a recent paper in the British Journal of Surgery (Dikken et al. 2013), the relation between hospital volume (the number of resections of a certain type per year in a given hospital) and overall survival for oesophageal and gastric cancer was studied, in four countries in Europe.
The table below, taken from that paper, shows, in the column "2 year survival", the results of a Cox proportional hazards regression with hospital volume, sex, age, histological grade and TNM stage.

	Oesophagectomy	
	30-day mortality (%)*	2-year survival (%)†
	Odds ratio	Hazard ratio
Annual hospital volume		
1–10	1·00 (reference)	1·00 (reference)
11–20	0·82 (0·61, 1·11)	0·92 (0·78, 1·08)
21–30‡ or ≥ 21§	0·68 (0·50, 0·93)	0·84 (0·63, 1·11)
31–40	0·58 (0·39, 0·85)	0·77 (0·63, 0·94)
≥ 41	0·55 (0·42, 0·72)	0·79 (0·66, 0·96)
P for trend#	< 0·001	0·004
Sex		
M	1·00 (reference)	1·00 (reference)
F	0·77 (0·62, 0·95)	0·78 (0·69, 0·90)
Age (years)		
< 60	1·00 (reference)	1·00 (reference)
60–75	1·82 (1·45, 2·28)	1·40 (1·27, 1·55)
> 75	3·99 (3·06, 5·21)	1·87 (1·58, 2·23)
Histology		
Adenocarcinoma	1·00 (reference)	1·00 (reference)
SCC	1·44 (1·15, 1·79)	1·29 (1·15, 1·44)
Other carcinoma	1·28 (0·81, 2·04)	1·45 (1·03, 2·05)
TNM stage¶		
0		0·57 (0·29, 1·14)
I		1·00 (reference)
II		1·96 (1·46, 2·62)
III		3·71 (2·74, 5·04)
IV		8·13 (4·39, 15·08)
Unknown		1·77 (1·01, 3·11)

a. The follow-up was artificially censored at 2 years. What is this type of censoring called?

b. The primary interest is in hospital volume. Why did the researchers adjust for the other factors as well?

c. What is your conclusion with respect to the research question: "how does hospital volume influence overall survival in oesophageal cancer"? Please discuss also the direction of the effect of hospital volume on overall survival.

d. Suppose that the baseline cumulative hazard at 2 years (with all covariates taken at their reference values) was estimated to be 0.5. How much was the estimated 2 year survival probability for each of the hospital volume categories in the table (with the remaining covariate values still taken at their reference values)?

e. The authors also fitted a Cox model like that of the table above, but now with hospital volume as a continuous covariate.

i. Looking at the results of the table above, is it reasonable to assume that the effect of hospital volume as a continuous covariate is linear?

ii. How large would the hazard ratio of hospital volume, as a continuous covariate, be approximately? Choose between 0.80, 0.90 and 0.99, and motivate your choice.

f. How would the authors be able to test for differences of the effect of hospital volume on overall survival between the four countries involved?