# Statistical Science Master
# Exam survival analysis

**July 15, 2013, 10:00-13:00**

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

**Exercise 1** — The time to relapse, in months, for patients on two treatments for leukemia is compared using the following log normal regression model

$$Y = \log(X) = 2 + 0.5Z + 2W$$

where $W \sim N(0,1)$ and $z = 1$ if treatment $A$ and 0 if treatment $B$.

- Compute the survival probabilities of the two treatments at 12, 24, and 60 months by using the `R` command `pnorm`.

**Exercise 2** — 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+
```

Assume that the death times follows and exponential distribution with parameter $\lambda$

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0 \ .$$

In case of right censored observation the likelihood is as follows

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

where $\delta_i$ indicates whether the exact lifetime $X$ is observed ($\delta = 1$) or not ($\delta = 0$).

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

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

where $\hat{\lambda}$ denote the maximum likelihood estimate; $2r\lambda/\hat{\lambda} \sim \chi_{2r}$; $r$ is the number of deaths and $\chi_{2r}$ is a $\chi^2$ distribution with $2r$ degrees of freedom. Use the R command qchisq to compute the quantile.

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

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

c. Estimate the mean residual life time

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

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

**Exercise 3** — The following vector of times is generated from a Pareto distribution with shape and scale parameters respectively equal to 2 and 1

```
> yi <- c(7.532711,5.374554,3.491295,2.202139,9.916605,2.226205,
+    2.117130,3.026644,3.179074,32.369651)
```

a. Generate an independent sample of size $n = 10$ from the uniform distribution with parameters $a = 3$ and $b = 9$. Use set.seed(123) to make sure that we all get the same data. To generate data from an uniform distribution, use runif.

b. Use the values generated in (a) as **left censoring times** for the times ti given above. The data from a left censored sample can be represented by pairs of random variables $(T, \delta)$ where $\delta$ indicates whether the exact lifetime $X$ is observed ($\delta = 1$) or not ($\delta = 0$). Note that for left censoring the exact life time is known if and only if $T = \max(X, C_l)$ where $C_l$ stands for "left" censoring time.

c. Count how many events and how many censorings there are in this sample.

**Exercise 4** — A study on 36 patients with a malignant tumor in the kidney has been carried out. The patients have been treated with a combination of chemotherapy and immunotherapy, but additional a nephrectomy (surgical removed on the kidney) had been carried out on some patients. The age of a patient has been classified according to whether a patient is younger than 60, between 60 and 70 or older than 70. Of interest is whether the survival time of some patients depends on their age at the time of diagnosis and on whether or not they have received nephrectomy.
A Cox model with risk factor age has been fitted and the estimated log-hazard ratio and their standard errors are given in Table 1.

a. Compute the hazard ratio (HR) and the 95% confidence interval (95% CI) for patients in age categories $60-70$ and $>70$ and comments your results.

b. Estimate the hazard ratio and the 95% confidence interval for an individual older than 70 relative to someone aged between 60 and 70. $cov(\hat{\beta}_2, \hat{\beta}_3) = 0.0832$ where $\hat{\beta}_2$ and $\hat{\beta}_3$ are respectively the estimated log-hazard ratio for $age_{60-70}$ and $age_{>70}$. (Hint: $var(\hat{\beta}_3 - \hat{\beta}_2) = var(\hat{\beta}_2) + var(\hat{\beta}_3) - 2cov(\hat{\beta}_2, \hat{\beta}_3)$ ). What would be an easier way to estimate those values?

c. A Cox model has been fitted by considering as prognostic factors age and treatment. The estimated hazard function for the $i^{th}$ patient was found to be

$$\hat{h}_i(t|Z) = \hat{h}_0(t)exp(-1.411 \times Z_i + 0.013 \times Z_{1i} + 1.342 \times Z_{2i})$$

where $Z_i$ is 1 if the patient has had a nephrectomy and zero otherwise; $Z_{1i} = 1$ if a patient is aged between 60 and 70 and zero otherwise; $Z_{2i} = 1$ if a patient is older than 70 and zero otherwise. The estimated baseline survival at 12 months is $\hat{S}_0(12) = 0.486$.
Estimate the survival $\hat{S}(t)$ and the cumulative hazard $\hat{H}(t)$ at 12 months for an individual younger than 60 who has had a nephrectomy. (The cumulative hazard function is: $\hat{H}(t) = -\log \hat{S}(t)$)

| | B | SE | HR | 95% CI |
|---|---|---|---|---|
| $age_{60-70}$ | -0.065 | 0.498 | | |
| $age_{>70}$ | 1.824 | 0.682 | | |

Table 1: Cox model with age