# Computer Practical Day I
# Exercises

## 1 Examples of survival analysis

Discuss the following study protocols. Is survival analysis needed for their analysis? If yes, determine the outcome and relevant time scale, and discuss whether the assumption of independent censoring is reasonable.

1. In a randomized clinical trial, one group of patients receives a new treatment and the other group placebo. We want to know if the treatment prolongs survival.

2. The performance of several hospitals is compared. One of the indicators is the proportion of patients dying due to short-term surgery-related complications.

3. Survival after diagnosis of breast cancer is studied. Patients who have a distant metastasis are considered as censored.

4. A study is designed to investigate the relationship between the duration of breast-feeding and several explanatory variables.

5. A cohort of children is followed to study when they can identify at least 10 letters correctly. They are followed between their 5th and 7th birthday. During this period they are tested every 3 months. For practical reasons, the children are only followed as long as they attend one of the participating schools (all normal primary schools).

## ~~2 Parametric models~~

~~1. The lifetime of rats follows an exponential distribution with a hazard rate of 1/3 per week. Write down a formula for the survival function from 0 to 6 weeks, sketch it and determine the mean survival.~~

2. ~~The time in days to development of a tumor follows a Weibull distribution with $\gamma = 1.1$ and $\lambda = 0.05$. Calculate the hazard and survival functions at 1 day and 1 week.~~

# 3 Kaplan-Meier

The data listed below are from a study on the toxic effect of Respisonide. Twelve mice received a large dose of Respisonide and the time until death was recorded. The survival times (in hours) of the mice were: 30, 50, 50*, 51, 66*, 82, 92, 120*, 140, 150, 180, 190. A * indicates that an observation was censored.

1. Use these data to complete the following table:

| time | no. at risk | no. dead | no. censored | death risk | 'risk' of surviving | survival (dec) | se(survival) |
|---|---|---|---|---|---|---|---|
| 30 | 12 | 1 | 0 | 1/12 | 11/12 | 0.92 | 0.08 |

| $t_i$ | $R_i$ | $d_i$ | $c_i$ | $\hat{h}_i$ | $1-\hat{h}_i$ | $\hat{S}(t_i)$ | $se(\hat{S}(t_i))$ |
|---|---|---|---|---|---|---|---|

2. Make a sketch of the Kaplan-Meier curve and its 95 % pointwise confidence interval.

3. Check your result by using SPSS. The data are in the file 'mice.sav'. Open this file (start up SPSS and use the commands File, Open). You will see that there are two variables: time for the survival time and death, indicating if the mouse has died (death=1) or that its observation was censored (death =0). Use the commands Analyze; Survival; Kaplan-Meier to calculate the Kaplan-Meier curve. Move the variable indicating the survival time to the box Time and the censoring variable to the box Status. With Define Event you can indicate which value corresponds to a real event. Use the correct options such that SPSS makes

a plot of the survival curve. Compare the SPSS output and plot with your own calculations.

4. Report the probability that a mouse is still alive 5 days after receiving the drug.

5. Again go to the Kaplan-Meier Window. Choose option Save and save both Survival and standard error of the predicted survival. Compute the upper and lower limit of the 95 % confidence interval (estimated value ± 1.96 se) as new variables. Graph the points of the Kaplan-Meier curve as well as those of the upper and lower limit with time on the x-axis as an overlay plot (graphs, legacy dialogs, scatter/dot). Make sure the points are joined in a step function by choosing line type = Left-step (in screen add interpolation line).

# Estimate the survival and cumulative hazard functions

In Figure 1 the results of a clinical trial of a drug 6-mercaptopurine (6-MP) in 21 children with acute leukemia are shown. The trial was conducted at 11 American hospitals. Patients were selected who had a complete or partial remission of their leukemia induced by treatment with the drug prednisone. Data for two groups of patients by remission status (complete or partial) are reported in Table 1. Patients were followed until their leukemia returned (relapse) or until the end of the study (in months). In the table also time to relapse for placebo patients are reported but for the exercise we shall consider only time to relapse for 6-PM patients.

1. Download the file **drug6mp.sav**. This file contains the same as illustrated in Figure 1 (without column Time to relapse for placebo patients). Open this file (start up SPSS and use the commands File, Open). There are four variables in the file: patient ID, groups (partial remission, complete remission) time to relapse (timerelmp) and status, indicating if the patient had experienced relapse (relapse=1) or that his observation was censored (relapse =0). Determine the minimum and the maximum follow up.

2. Use the data in Figure 1 to estimate Kaplan-Meier curve by applying the product estimator (note that time to events is not ordered!).

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{r_i}\right), \; d_i : \text{number of events}; \; r_i : \text{number at risk};$$

3. Estimate the standard error of the survival curve by using the following formula
$$\text{se}(\hat{S}(t)) = \hat{S}(t)\sqrt{\sum_{t_i \le t} \frac{d_i}{r_i(r_i - d_i)}}$$

4. Write all calculations involved in the estimation process for the Kaplan-Meier curve and the standard error in a table

5. Estimate by means of the Nelson-Aalen estimator the cumulative hazard and its variance. Recall the definition
$$\hat{H}(t) = \sum_{t_i \le t} \frac{d_i}{r_i}; \quad \sigma^2(t) = \sum_{t_i \le t} \frac{d_i}{r_i^2} \ .$$

6. Check your results by using SPSS. Use the commands Analyze; Survival; Kaplan-Meier to calculate the Kaplan-Meier curve. Move the variable indicating the time to relapse to the box *Time* and the censoring variable to the box *Status*. With *Define Event* you can indicate which value corresponds to a real event. With the option *save* you can also estimate the cumulative hazard in order to check with your own computations. Use the correct options such that SPSS makes a plot of the survival curve and the cumulative hazard. Compare the SPSS output with your own calculations. Note that the estimation of the cumulative hazard $\hat{H}(t)$ with SPSS is not exactly as the estimator obtained with the Nelson-Aalen method. Can you explain the reason?

7. Compare SPSS computations with your results in question 2.

*Remission duration of 6-MP versus placebo in children with acute leukemia*

| *Pair* | *Remission Status at Randomization* | *Time to Relapse for Placebo Patients* | *Time to Relapse for 6-MP Patients* |
|---|---|---|---|
| 1 | Partial Remission | 1 | 10 |
| 2 | Complete Remission | 22 | 7 |
| 3 | Complete Remission | 3 | $32^+$ |
| 4 | Complete Remission | 12 | 23 |
| 5 | Complete Remission | 8 | 22 |
| 6 | Partial Remission | 17 | 6 |
| 7 | Complete Remission | 2 | 16 |
| 8 | Complete Remission | 11 | $34^+$ |
| 9 | Complete Remission | 8 | $32^+$ |
| 10 | Complete Remission | 12 | $25^+$ |
| 11 | Complete Remission | 2 | $11^+$ |
| 12 | Partial Remission | 5 | $20^+$ |
| 13 | Complete Remission | 4 | $19^+$ |
| 14 | Complete Remission | 15 | 6 |
| 15 | Complete Remission | 8 | $17^+$ |
| 16 | Partial Remission | 23 | $35^+$ |
| 17 | Partial Remission | 5 | 6 |
| 18 | Complete Remission | 11 | 13 |
| 19 | Complete Remission | 4 | $9^+$ |
| 20 | Complete Remission | 1 | $6^+$ |
| 21 | Complete Remission | 8 | $10^+$ |

$^+$Censored observation

Figure 1: Data remission duration in children with acute leukemia