Computer Practical Day II
Exercises

Log-rank test

You have two treatments A and B. The observed survival times are:
$\text{Treatment}_\text{A}$: 11; 15; 26^+; 30; 31; 35
$\text{Treatment}_\text{B}$: 12; 25; 29^+; 33; 36; 40^+.
Test whether there is any difference in survival between the two treatments. The hypothesis is as follows:

$$H_0 : S_A(t) = S_B(t) \text{ for all } t \text{ versus } H_A : S_A(t) \neq S_B(t) \text{for some } t.$$

Hint: begin by computing the expected number of deaths in group A

$$E_A = \frac{r_A \cdot (d_A + d_B)}{r_A + r_B}$$

and the contribution to the variance $E_A(1 - E_A)$. We compare the sum of deaths in group A ($\sum d_A$) to the sum of deaths in group B ($\sum E_B$). If there is no difference in survival, $\sum d_A$ won't differ much from $\sum E_A$. Under $H_0 : S_A(t) = S_B(t)$ for all t the quantity

$$Z = \frac{\sum_{\text{time points}}(d_A - E_A)}{\sqrt{\sum_{\text{time points}} E_A(1 - E_A)}}$$

follows a standard normal distribution...). Check your computations with SPSS (use file **logrank example.sav**).

Minimum, maximum and median follow-up

Download the file **drug6mp.sav**. You have used this file yesterday. 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).

1. Determine the minimum and the maximum follow up.
2. Estimate the median follow-up by using the two methods described in the lecture of this morning (by excluding patients who have died and by using Kaplan-Meier).
3. Estimate the median survival time.
4. Compare now the two groups: *partial remission* and *complete remission*. Test statistically whether the Kaplan-Meier curves of the partial remission and complete remission groups are equal (put in the window *factor* the variable *group*) then click compare factor and tick the box *log rank*. Plot the two curves together. What is your conclusion?

Life table method

In this exercise you will practice the use of the life table method. Use again file **drug6mp.sav** to perform the analysis.

1. From the menu choose: *Analyze, Survival, Life Tables,*
2. Select survival variable.
3. Select the *status variable* to define cases for which the terminal event has occurred.
4. Click the box *Define Event* to specify the value of the status variable that indicates that an event occurred.
5. Select time intervals in the window *Display Time Intervals*: type 36 as the maximum time interval to display and type 6 as the number of intervals to display *by* (you can also choose different values for the number of intervals to display *by* for example 2); plot the survival function. What do you notice?
6. For each interval i determine person years T_i and the number of events d_i. Estimate the hazard and the standard error in each interval by applying

$$\hat{h}_{(t_{i-1},t_i]} = \frac{d_i}{T_i} \ , \quad \mathrm{se}^2(\hat{h}_{(t_{i-1},t_i]}) = \frac{d_i}{T_i^2} \ .$$

Alternative observation schemes

Death times of elderly residents of a retirement community

1. We want to estimate the death rate between 1/1/2005 and 1/1/2009 in all retirement community in Amsterdam. The people living in those retirement community are followed for the entire period. The time scale is time since birth. During the follow up new people enter the data set and people die. Describe the censoring mechanisms present in this data.

2. We now want to estimate the death rate between 1/1/2005 and 1/1/2009 in all hospices in Den Haag. We set up a different experiment. The number of people living in the hospices are counted on 1/1/2005. Then on 1/1/2009 again the number of people resident in the hospices are counted. Which pieces of information we miss in this data?

Time to recurrence of a particular cancer

1. Consider a study in which interest centers on the time to recurrence of a particular cancer following surgical removal of the primary tumor. Three months after their operation, the patients are examined to determine if the cancer has recurred. At this time, some of the patients may be found to have a recurrence. Which type of censored data are these?

2. Consider again the example concerning the time to recurrence of a tumor. If a patient is observed to be free of the disease at three months, but is found to have had a recurrence when examined six months after surgery, the actual recurrence time of that patient is known to be between three months and six months. Which type of censoring mechanism is present in this data?