

# Applied Survival Analysis - January 2016

## Lab 3: Comparing survival curves between groups

In this lab, we are going to understand how to produce a Kaplan-Meier plot of survival estimates for more than one subgroup on the same graph, and compute a *Logrank* or *Wilcoxon* test in R.

- (a) Using the following artificial data from two treatment groups

Group0: 15, 18, 19, 19, 20    Group1: 16<sup>+</sup>, 18, 20<sup>+</sup>, 23, 24<sup>+</sup>,

compare the survival of the two groups by computing the *Logrank* test using the `survdif` function in R. What's your conclusion? **Optional:** Write an R program to construct tables like that on page 22 of today's notes.

- (b) We are going to work on data from a leukemia remission study (Garrett 1997). The data consist of 42 patients who are monitored over time to see how long (*weeks*) it takes them to go out of remission (*remiss*: 1 = yes, 0 = no). Half of the patients received a new experimental drug and the other half received a standard drug (*trt*: 1=6-MP, 0=Control). This dataset is called *leukem.csv*. Import the data in R. Use the `factor` function to encode the `trt` variable as a factor (please, see `?factor` for more details).

- (i) Use the `survfit` function to get the KM estimates in both groups. After using information from (a), guess what the syntax would be.
- (ii) Plot the survival estimates of the two treatment groups using the `plot` function. Which group seems to be doing better? Comment on the graph.
- (iii) Compare the survival between the two groups in terms of the *Logrank* and *Wilcoxon* test using the `survdif` function. What do you conclude from these tests? Why do you think that there is a difference in the p-values of the two tests? Explain your opinion.
- (iv) How could you informally check the proportional hazards assumption? Any ideas?