

## Homework 10 STAT 425/525

Due Monday November 21

*R code for this assignment will be provided. I will need to see the plots you are asked to look at below.*

The data are of survival times of patients with acute myelogenous leukemia (AML). The question at the time was whether the standard course of chemotherapy should be extended ('maintainance') for additional cycles. The data can be found in the data file `aml` in the `survival` package. We will look at Kaplan-Meier and Nelson-Aalen nonparametric survival curves and carry out the log-rank test to compare the patients in the two groups.

1. Print out the data set just so you have it in front of you. It is not that big: `time` is survival time, `status` is censoring status with 1 denoting death and 0 denoting censored and `x` is a binary categorical variable indicating membership in either the Maintained or Non-maintained groups.
2. The survival times for the Maintained group is

9    13    13+    18    23    28+    31    34    45+    48    161+

Fill in the table below by hand.

time	$\hat{q}_i$	$\hat{p}_i$	$\hat{P}_k$
------	-------------	-------------	-------------

Remember  $\hat{P}_k$  is actually the survival function estimate.

3. Give summary information including mean and median survival times, and Kaplan-Meier Curves for the two groups. Briefly discuss what these results seem to indicate about the two groups.
4. Compare the mean survival times for the two groups and the median survival times for the two groups. Which do you think is more important to use as a summary measure for comparison - the mean or the median? Justify your answer.
5. Carry out the log-rank test and give the results. What do they imply about the survival times of the two groups. Justify your answer.
6. Using the data in the table in Problem 2 calculate the Nelson-Aalen estimator of the survival curve.
7. Using the R code provided calculate the Nelson-Aalen estimator and plot it and the Kaplan-Meier Curves. You can use the NA output to confirm your calculations in the previous problem.
8. Fit a Cox Proportional Hazards model to these data and summarize the results giving me an estimate of the appropriate hazard ratio and an associated 95% CI. Interpret the results in terms of the problem.