

Math 150 - Methods in Biostatistics - Homework 9

your name here

Due: Wednesday, April 3, 2019, in class

```
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=3, fig.width=5,
                        fig.align = "center")
library(tidyverse)
library(broom)
library(tidylog)
library(survival)
```

Note: there are two places to check for hints on R code. One is the class notes (<http://st47s.com/Math150/Notes/>, see R Examples) and the other is the R manual associated with the textbook which is on Sakai.

1. Chp 9, A26

Provide a brief explanation of why the estimated variance of $\hat{S}_{KM}(0)$, and hence the standard error of $\hat{S}_{KM}(0)$, is equal to 0.

2. Chp 9, E4

Immediately after a heart transplant, patients are randomly assigned to two treatment therapies to improve recovery from the transplant, therapy 1 and therapy 2. The patients are then followed for up to 5 years after their surgery. Define the time-to-event random variable T as the time (in months) until recovery (the event) after a heart transplant. For each of the following study descriptions that involve T , sketch the graph of the survival curve (or curves) with as much detail as necessary. Please note that parts (a) through (d) are independent of each other.

- Therapy 1 is not very effective shortly after surgery, but everybody recovers before the study period is over.
- Therapy 2 is very effective shortly after surgery, but becomes less effective after 3 years. Not every patient fully recovers by the end of the study period.
- Two curves on the same plot: Therapy 1 is consistently more effective than Therapy 2 over time.
- Two curves on the same plot: Therapy 1 is more effective than Therapy 2 for the first 2.5 years, and then Therapy 2 is more effective than Therapy 1 for the remaining duration of the study.

3. Chp 9, E6

The Kaplan-Meier curve in Figure 9.17 displays hypothetical estimated survival probabilities of death due to brain cancer, where time (from diagnosis) until death is measured in months.

- Is the largest event time censored or complete? How do you know?
- Use the curve to estimate the mean time until death due to brain cancer.

4. Chp 9, E11 Male Fruit Fly Longevity

(Lots to read in the text about the dataset.)

- (a) Construct the Kaplan-Meier curve with a confidence interval for the **Fruitfly** data and describe the survival pattern for the fruitflies over time. Use **Longevity** as the time-to-event variable.
- (b) Construct the Kaplan-Meier curve for the lifetimes of the fruitflies by number of partners, using **Partners** as the grouping variable. Briefly comment on the observed relationship between survival and number of female partners.
- (c) Perform the log-rank and Wilcoxon tests. Report the test statistics and p-values for both tests. State the conclusions for both tests. If the tests yield different conclusions, briefly explain why.

5. Chp 9, E12 VA Lung Cancer Study

(Lots to read in the text about the dataset.)

- (a) Create a graph with both Kaplan-Meier curves to compare the survival time (use the variable 'time') for subjects with the standard and the test chemotherapy treatment. What do you observe about the survival probabilities for the groups of subjects?
- (b) Conduct the log-rank test and the Wilcoxon test to compare the survival curves of both treatment groups. Interpret the results.
- (c) It may be beneficial to incorporate health as a variable in the analysis. Patients with low Karnofsky scores are less healthy than patients with high Karnofsky scores. Create four groups with the **Veteran** data: **trt=1** and Karnofsky score low, **trt=1** and Karnofsky score high, **trt=2** and Karnofsky score low, and **trt=2** and Karnofsky score high. Recall that it is often best to keep sample sizes as equivalent as possible when you determine what is a low or high Karnofsky score. Create a Kaplan-Meier curve for each of the four groups. Conduct the log-rank test and the Wilcoxon test to compare the survival curves of the four groups. (While we have only discussed using these tests to compare two groups, they can easily be extended to more than two groups.) Did incorporating health into your analysis impact your conclusions?