

## BST 665 - Survival Analysis Homework 1

Assigned: January 28, 2019

Due: February 4, 2019, by end of class

### Instructions:

Review Chapters 1 and 2 of *Applied Survival Analysis*. Then complete the exercises below.

Responses should be typed or printed neatly (if you have multiple pages, please staple them).

Unless specified otherwise, you are encouraged to use software whenever possible to create plots and perform calculations. To receive full credit, you must show your work. It is not sufficient to circle a result on the SAS output or to simply write down a numerical answer without an explanation.

### Exercises:

1. A gynecologist has collected data on women using a particular intrauterine device (IUD). She is interested in estimating how long women will use the IUD before discontinuing, or removing, the device due to adverse events. The following data gives the number of weeks from beginning using the IUD until discontinuation. Times for women who were lost to follow-up or discontinued use for reasons other than an adverse event (i.e., censored times) are labelled with a + sign.

10, 13+, 18+, 19, 23+, 30, 36, 38+, 54+, 56+, 59, 75, 93, 97, 104+, 107, 107, 107+

Use these data to complete Parts A-C. Each of these problems should be done **by hand** (you can use SAS to check your answers if needed).

A. Compute and sketch the Kaplan-Meier estimator for these data.

B. Estimate the median time to discontinuation.

C. The gynecologist would like to know what proportion of women will use the IUD for at least a year (i.e., 52 weeks). Provide an answer to this question.

2. In survival analysis, using a fully parametric regression model means proposing a distribution for the survival time,  $T$ . While the Weibull and exponential distributions are the most common choices, any continuous distribution that restricts  $T$  to be positive can be used. Suppose we want to use the following as the cumulative distribution function for  $T$ :

$$F(t) = \frac{t^3}{1 + t^3}, t > 0$$

A. What is the survival function for  $T$ ?

B. What is the hazard function?

C. What is the median survival time?

3. A researcher is interested in the relationship between environmental stress and the development of tumors in rats. Ninety rats were divided into three groups and exposed to low-stress, medium-stress, and high-stress environments. The investigator injected tumor cells into the rats and observed them for 20 weeks. The outcome of interest was the tumor-free time (i.e., the number of weeks between injection and development of a tumor).

Download the dataset named “Rats.sas7bdat”. The variables in this dataset are:

- ID number for each rat (ID)
- The type of environment for each rat (Stress)
- The number of tumor-free weeks (Weeks)
- A censoring indicator (Tumor)
  - Tumor = 1 if the rat developed a tumor; Tumor = 0 otherwise

A. On the same graph, plot the Kaplan-Meier estimators for the three environments. Which group had the poorest survival? Which group had the best?

B. What was the median survival time for each of the three groups?

C. Is there evidence to conclude that environmental stress is associated with tumor development? Support your answer using the log-rank test. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

D. Write a short summary (1-2 paragraphs) of the results of this study. Be sure to include your interpretation of the results and to take the stated goal of the study into consideration.