Reading: • Kleinbaum *Survival Analysis* Chapters 1
           ○ article: Streiner (1995)
           ○ article: Dickson et al. (1989)

NOTE: Unless explicitly stated, direct computer output is **not** desired. Typically only part of the computer output is asked for (such as a confidence interval) and then proper interpretation of the statistics is requested.

DATA: The data for these exercises can be found on the class web page: `http://faculty.washington.edu/heagerty/Courses/VA-summer-epi/` in the *Homeworks* directory (ie. click on Homeworks from the main summer course page).

Survival Analysis

1. Streiner (1995) identifies several assumptions and issues associated with the use of survival analysis techniques (pp. 443-444). Consider the Mayo PBC data presented in Dickson et al. (1989) and answer the following questions about that study:

(a) In survival analysis we measure the time until an event occurs. This requires both a well defined starting time or origin ($t = 0$) in addition to an event (or events) that defines the "failure time". What defines the time origin in the PBC study?

(b) What is the definition of an event, or failure in the PBC study?

(c) What are the causes of censoring in the PBC study?

(d) What justifies that these mechanisms are unrelated to the outcome time?

(e) Figure 3 in Dickson et al. (1989) presents survival curves for 106 cross-validation patients. Based on the Kaplan-Meier curves, what fraction of each group dies within 5 years? (approximate)

Survival Analysis: 1-sample

2. The following data are from Avalos et al. (1993) and represent the time until death or relapse for 12 bone marrow transplant patients. These data are for a specific subset (autogeneic transplant, non-Hodgkin's lymphoma):

death / relapse times (days): 42, 53, 57, 63, 81, 140, 176, 252, 524

censored times (days): 210, 476, 1037

(a) Estimate by hand calculation survival as a function of time since transplantation using the Kaplan-Meier method. Complete the rows of the following table: (using our notation from the lecture notes: pg 255)

Time (days)	At Risk	Died	Censored	Survived	Conditional probability	K-M Estimate
t_i	R_i	d_i	l_i	S_i	$(1 - d_i/R_i)$	$\prod(1 - d_i/R_i)$
42	12	1	0	11	0.917	0.917
53	11	1	0	10	0.909	0.833
$\vdots$						$\vdots$

Survival Analysis: 2-samples

3. The data on the web page, `larynx.dat`, represent the death times of male laryngeal cancer patients that attended a Dutch hospital between 1970 and 1978. Times recorded are the interval (in years) between first treatment and either death, or the end of the study (January 1, 1983). Also recorded are the patient's age at diagnosis and the year of diagnosis.

(a) Summarize the distribution of Age at diagnosis, Year of diagnosis, and Stage of disease.

(b) Summarize the number of observed deaths and the number of subjects by disease stage (a simple 2×4 table).

(c) Summarize the year of diagnosis by disease stage (a simple 2×9 table, or mean year for each stage), and the age of diagnosis by disease stage (ie. mean, median, range for each stage).

(d) Calculate a Kaplan-Meier curve for the 90 total subjects. Interpret the point estimate at $t = 5$. Turn this plot in.

(e) Compute a new stage variable that groups the 3,4 and the 1,2 values. Calculate a pair of Kaplan-Meier curves (same plot if you can) and interpret what this suggests about the prognosis based on stage (3,4 versus 1,2). Turn this plot in. (optional: try and display standard error lines around the estimated survival function)

2

(f) Use the log-rank test to compare the survival course of stage=(3,4) versus stage=(1,2). State the null and alternative hypotheses and interpret the test result.

(g) Calculate a Kaplan-Meier curve for each level of stage (ie. 4 separate curves, same plot is preferred). Interpret what this plot suggests about the survival based on the stage of disease. Report the median survival time by stage of disease. Turn this plot in.

(h) How would you test if the stages have different survival functions? Execute this test and interpret the results.

(optional) Does the survival course seem to change in the later years of diagnosis? Consider the diagnosis times as pre-1975 and 1975+ and compare (graphically and with confirmatory procedures) whether there appears to be a change. Is the difference between stage=(3,4) and stage=(1,2) changing over the time of study entry?