Homework 6 - due 5/20/13
Please submit your assignment in Word or PDF format by 9:30 am (PDT) to the Biost 513 Assignment Dropbox at http://canvas.uw.edu/. You may go to "Assignments" at the left side menus on your Canvas page and then follow the instructions to upload your assignment.

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://courses.washington.edu/b513/ in the Homework directory.

Stata.help6: Key Stata commands that are useful for these exercises are described in the text file stata.help4, also available in the homework directory.

## PLEASE START YOUR RESPONSE TO EACH QUESTION ON A SEPARATE PAGE AND PUT YOUR NAME ON EACH PAGE.

## Survival Analysis

1) Consider the Mayo PBC data presented in Dickson et al. (Hepatology 10:1-7, 1989; the journal is in the HS library and the article on electronic reserve) 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 ( $\mathrm{t}=0$ ) 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) Do you believe these causes of censoring are unrelated to the outcome time? Why or why not?
(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
Estimate, by hand calculation, survival as a function of time since transplantation using the Kaplan-Meier method. Complete the rows of the following table: (see lecture notes: pp 301-302)

| Time (days) $\mathrm{t}_{\mathrm{i}}$ | At Risk <br> $\mathrm{R}_{\mathrm{i}}$ | Died <br> $\mathrm{d}_{\mathrm{i}}$ | Censored $l_{\mathrm{i}}$ | Failure Probability $\mathrm{d}_{\mathrm{i}} / \mathrm{R}_{\mathrm{i}}$ | Conditional Survival (1-d/R $\mathrm{R}_{\mathrm{i}}$ ) | $\begin{gathered} \text { K-M } \\ \text { estimate } \\ \prod\left(1-d_{i} / R_{i}\right) \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 42 | 12 | 1 | 0 | . 083 | . 917 | . 917 |
| 53 | 11 | 1 | 0 | . 091 | . 909 | . 833 |
| ! | : | : | : | : | : | : |

## Survival Analysis: 2-sample

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 $2 \times 4$ table), noting any obvious relationships.
(c) Summarize the year of diagnosis by disease stage (a $4 \times 9$ table, or mean year for each stage), and the age of diagnosis by disease stage (ie. mean, median, range for each stage), noting any obvious relationships.
(d) Calculate a Kaplan-Meier curve for the 90 total subjects. Interpret the point estimate at $\mathrm{t}=6$. 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)
(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 KM curve for each level of stage (i.e., 4 separate curves on the 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 (all 4 stages this time)? Execute this test and interpret the results.
