

HOMEWORK #2

(Due Thursday, January 24 in class)

Reading:

- I. H&L: §§ 2.4 and 2.5

Homework:

1. Read the article [Do Oscar Winners Live Longer than Less Successful Peers? A Reanalysis of the Evidence -- Sylvestre et al. 145 \(5\): 361 -- Annals of Internal Medicine](#) and be prepared to discuss during the Discussion Session on Tuesday, January 22. Answer the following questions in writing:
 - a. To what topic--page(s)-- in the lecture notes does this article correspond?
 - b. Are you convinced that “immortal time” bias is a real phenomenon? Why or why not?

2. The time (in months) from start of treatment to relapse (or end of follow-up) for 15 children with rhabdomyosarcoma treated with surgery and radiation but no chemotherapy was as follows:

Relapsed: 2, 3, 9, 10, 10, 15, 16, 30
 Disease free: 12, 15, 18, 24, 36, 40, 45

Estimate by hand calculation the disease-free survival as a function of time since treatment using the method of Kaplan and Meier (1958), by completing the rows in the table. You can check your answer using STATA if you wish.

Time Months t_i	Number At Risk n_i	Number Died d_i	Number Survived s_i	Conditional Probability s_i/n_i	K-M Estimate $\prod s_j/n_j$
2	15	1	14	0.933	0.933
:	:	:	:	:	:

3. Using STATA and the data in herpes.raw (documented in herpes.doc, both available on class web site) on time from the end of primary episode to first recurrence. Edit the data as suggested in class by treating all observations of a year or more duration as censored.
 - a. Plot Kaplan-Meier survival curves for each of the three HSV Type groups on one plot, showing the numbers remaining at risk in each group at 100, 200 and 300 days.
 - b. Estimate the median time from end of primary episode to first recurrence for each of the three HSV type groups in two ways: (1) by reading off the graph in part a.; and (2) by using the STATA command stsum.

- c. For each HSV Type, estimate and plot the life-table estimate of the hazard function at intervals of 50 days. Compare with the kernel smoothed hazards plotted in the lecture notes.
- d. Based on the values of $S(t)$ and its pointwise 95% confidence intervals, what is your guess about the statistical significance of each of the 3 pairwise comparisons of the three groups? Why?