Course EPIB-634: Survival Analysis & Related Topics [Winter 2007]

Assignment 4

`material in www.epi.mcgill.ca/hanley/c681/survival_analysis  unless otherwise specified`

`( username: c681 ; password: 8 letters, H***J*##  both case-sensitive )material in`

1 Refer to the information in "How long to get a PhD [in epidemiology]?" reachable from the page `www.epi.mcgill.ca/hanley/c681/survival_analysis (can link to it from 634 page)`

(i) Manually, [i.e., without using any survival analysis software, i.e. just using the data as listed in the .txt file, and taking advantage of the special way the data are ordered] compute the Kaplan-Meier estimate of the 'proportion still in the program' (i.e. 'still don't have a PhD') and of the complement (the cumulative graduation rate) for the first 4 years after entering the program.

(ii) Use the survival analysis software of your choice to calculate the full table and curve of the 'estimated proportion still in the program'. Check the first 3 entries against you hand calculations. Then extract from these curves/tables/outputs the (i) mean (ii) median duration to obtain a PhD (iii) the 5-year graduation rate and an associated 95%CI.

(iii) To assess whether the "time to PhD" is different among those who entered the program in the 1990's versus earlier, use a statistical test to formally compare (a) the entire 'proportion still in the program' *curves* for these two ('earlier' vs 'later') groups (b) the '*5-year-graduation rates*'* (**the term rate is used here in the 'risk' i.e., proportion, sense. To formally compare the 5-year graduation 'rates', refer the statistic* $(estimate_{earlier} - estimate_{later})/sqrt[SE^2_{earlier} + SE^2_{later}]$ *to the z-distribution; also note that the graduation rate is the one that <u>in</u>creases with time since entry i.e., is 'cumulative', whereas the proportion still in the initial stat is the one that <u>de</u>creases)*

2 Refer to Figure 1, and its legend.

(i) Compute (and comment on how close they are to each other) the Kaplan-Meier and Nelson-Aalen survival curves for the 5 sexually inactive fruitflies. (*If doing so manually, label the times* t_1 *to* t_5*; if using survival analysis software--where you need to supply times of the deaths--use some arbitrary time scale (deaths at 40, 55, 65, 75, and 90 days are close to reality)*

(ii) In order to compare the longevity of the 5 sexually active fruitflies with that of inactive ones, set up the computations for the log rank test and for a summary M-H odds ratio. *You don't need to complete the calculations manually, but lay out a few of the tables, along with the computations required, so that your research assistant could program the test and summary odds ratio in Excel or in R.*

How many *informative* 2 × 2 tables (risksets) are involved?

Refer to Figure 3, and its legend, where the objective is to compare 'like' active with 'like' inactive fruitflies, i.e., smaller with smaller, and larger with larger, but to still arrive at a single test and a single odds ratio. This version of the test is called a stratified Log-Rank test. The data layout and calculations for it and the summary ratio are the same as those used in Mantel and Haenszel's 1959 article where matching was on age and occupation. The log-rank test in (ii) just matches on time, whereas below we match on both time and thorax size.

(iii) Begin by setting up the 2 × 2 tables involving the 5 fruitflies with above average thorax lengths, represented by thicker lines (5 leftmost flies) of whom n=3 sexually active (shaded lines) and n = 2 sexually were inactive (black, reference group). Continue with those below average thorax lengths (5 righmost, with letter s for 'short'), and calculate the test and ratio over all informative tables. *Detailed calculations are not needed, but steps should be clearly indicated.*

In total, how many *informative* 2 × 2 tables (risksets) are involved in this stratified version?

[This type of stratified survival analysis will be used again when we cover Cox's regression model]

3 In the re-analysis by Sylvestre et al (Sylvestre MP, Huszti E, Hanley JA. Do OSCAR winners live longer than less successful peers? A reanalysis of the evidence Ann Intern Med. 2006 Sep 5;145(5):361-3; discussion 392.), the age at first Oscar nomination, age at first win, age in 2001 or at death, and vital status, for 20 of the performers were as follows

IDENTITY	AGE_NOM1	AGE_WIN1	AGE_LAST	DEAD
859	8	.	30	0
1290	31	.	39	0
1502	34	.	40	0
1457	36	.	50	1
690	52	.	53	1
628	44	52	58	1
535	34	.	59	0
493	35	40	60	1
786	38	.	60	1
1340	56	.	60	0
1487	23	.	62	0
177	61	.	65	0
308	53	.	66	0
291	56	.	68	0
1525	44	.	72	1
1009	39	.	73	1
1231	33	61	76	0
1052	46	46	77	1
1082	42	.	84	1
627	33	.	90	0

These 20 life-courses are re-arranged below, with the time (age) axis runnning from left to right.

```
         1         2         3         4         5         6         7         8         9
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
-------nnnnnnnnnnnnnnnnnnnnnnn------------------------------------------------------------
------------------------------nnnnnnnnn---------------------------------------------------
---------------------------------nnnnnnn--------------------------------------------------
-----------------------------------nnnnnnnnnnnnnnN----------------------------------------
---------------------------------------------------nN-------------------------------------
-------------------------------------------nnnnnnnnwwwwwwW--------------------------------
---------------------------------nnnnnnnnnnnnnnnnnnnnnnnnnn-------------------------------
----------------------------------nnnnnwwwwwwwwwwwwwwwwwwwwW-----------------------------
-------------------------------------nnnnnnnnnnnnnnnnnnnnnnN------------------------------
-------------------------------------------------------nnnnn------------------------------
----------------------nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn----------------------------
------------------------------------------------------------nnnnn-------------------------
----------------------------------------------------nnnnnnnnnnnnnn------------------------
-------------------------------------------------------nnnnnnnnnnnnn----------------------
-------------------------------------------nnnnnnnnnnnnnnnnnnnnnnnnnnnnN------------------
--------------------------------------nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnN-----------------
--------------------------------nnnnnnnnnnnnnnnnnnnnnnnnnnnnwwwwwwwwwwwwwwww--------------
---------------------------------------------wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwW-------------
-----------------------------------------nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnN------
--------------------------------nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
Age      1         2         3         4         5         6         7         8         9
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890

Highest Status by age indicated: n= nominee (N = death)     w= winner (W = death)
```

(i) Set up the risksets for a log-rank test and calculation of a summary odds ratio comparing the mortality experience in the person-time lived as a nominee and in the person time lived as a winner.

(ii) Calculate the Nelson-Aalen survival curve based on the mortality experience in the person time spend as a nominee and the corresponding one based on the mortality experience in the person time lived as a winner. (obviously, with little data, they are quite unstable).