Course EPIB-681: Data Analysis II [Winter 2004]

Assignment 4

including *Q 8 and Q 9 (follow-up studies*) on page 2, added Jan 21

`material in www.epi.mcgill.ca/hanley/c681/lifetables unless otherwise specified`
`( username: c607 ; password: 8 letters, H***J*## both case-sensitive )material in`

1 Refer to the information in "Population size, number of deaths, by age-sex, Canada 2001",
[extracted by JH from Statistics Canada websites. FYI, but not necessary for question: the links are http://www12.statcan.ca/english/census01/Products/standard/themes/ for population figures, based on 2001 census, and ___ for deaths; deaths in ages 1-15 were grouped in the publication, and were 'ungrouped' by JH in proportion to patterns seen in earlier years].

What was the (i) mean (ii) median age of the Canadian (a) male (b) female population in 2001? What was the (i) mean (ii) median age at death of those Canadian (a) males (b) females who died in 2001? What was the shape of the sex-specific distributions of the ages at death? [histograms help]

2 Construct "current" lifetables for each sex, based on the 2001 Canadian mortality rates. [if the L_x , T_x and e_x give you trouble, you can skip these 3 columns]. Plot the survivor functions. From the lifetables, calculate the (i) mean (ii) median age at death for each sex [also show histograms]. Why might your answers differ from those in Q1?

3 Which provinces and territories have the greatest/smallest "year-over-year" changes in life expectancy at birth for the years 1996 and 1997 [the ones available from the Canoe.ca website]. Explain why.

What do these variations have to do with the reliability measures (the cv's) given in the "Lifetables, Canada, Provinces and Territories 1995-1997" [link on /hanley/c681/lifetables page; cf Complete life tables] ? Why is the cv[e_0] for Ontario males 0.07% while for Quebec it is 0.09% and for Saskatchewan 0.25%?

These latter lifetables are based on 3 years of data; what would the cv's be if (as in Q2) you calculated them from just 1 year of data?

4 [2001 Canadian data] Calculate the (age-specific) mortality rate ratios contrasting mortality of males and females. Are the ratios constant across ages ? reasonably constant within certain age-decades?

On a single graph, plot the age-specific mortality rates against age for males and for females

Repeat the exercise , but with the mortality rates on a log scale. Are the male-female differences more 'constant' (invariant, homogeneous) on the rate difference scale or on the rate ratio scale?

What would be a [non-for-profit] premium for a 1-year 'term' insurance for $10,000 for (i) males (ii) females aged 25; 45; 65; 85 at last birthday? Explain your calculations.

5 Calculate and plot the (age-specific) mortality rate ratios (i) contrasting mortality of Swedish males in 1918 and Swedish males in 1917 (ii) contrasting the age-specific rates in the same two years for females. Comment on the patterns. Estimate how may excess deaths there were in Sweden in 1918, over 1917. [cf link from /hanley/c681/lifetables to Berkeley Mortality Database]

6 Examine the 1995-1997 Canadian provincial lifetables; link on /hanley/c681/lifetables page] What would be a [non-for-profit] premium for a 1-year 'term' insurance for $10,000 for (i) Quebec males (ii) Ontario males (iii) PEI males, all aged 55 at last birthday?

7 A 12-year old female passenger from Sweden survived the Titanic sinking of 1912 and died in 1981. (i) What proportion of her birth cohort were alive in 1912? What was the (i) mean (ii) median post-1912 longevity of this 'alive in 1912' comparison group? What percentage of her comparison group did she outlive? [the excerpt from the Swedish lifetable, found on /hanley/c681/lifetables takes a lot less time to download thanthe full1751-1910 version found in the Berkeley database]

Q8 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`

- Without using any survival analysis software, i.e. just using the data as listed in the .txt file, as well as a calculator or spreadsheet, calculate the estimated 'proportion still in the program' (i.e. still 'don't have a PhD') for the first 5 years after entering the program *(5 years is enough to show you understand the procedure)*
- Use the survival analysis software of your choice. to calculate full estimated 'proportion still in the program' curves, First, check your 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, doing so (a) overall and (b) separately for those who entered the program in the 1990's versus earlier (*).
- Are the 'proportion still in the program' *curves* significantly different for these two ('earlier' vs 'later') groups? Are the '*5-year-graduation rates*' significantly different for these two ('earlier' vs 'later') groups?
- When making these comparisons, why not drop the 'censored' observations? Answer first before actually doing so, i.e., predict what the effect of dropping them would be. Then, use the software to drop them anyway, and comment on what happens(*).

Hints

(*) create a new (0/1, "dummy") indicator variable that allows you to group the students into those who entered < 1990 versus ≥ 1990.

(**) run LIFETEST with a "WHERE" command i.e. , add a " WHERE (phd = 0); statement " to the PROC LIFETEST. (use a "if phd==0" phrase in the Stata 'sts' statements)

Q9 Refer to Figure 3, and its legend, in the second-last and last pages of "Survival Analysis / Follow-up Studies .. details" , in `www.epi.mcgill.ca/hanley/c681/survival_analysis` (pdf file is called survival_analysis_notes.pdf ; you have probably already printed it out).

- Restrict your attention to the 5 fruitflies with above average thorax lengths, represented by thicker lines (5 rightmost flies). Set up the computations for the log rank test comparing the longevity of the n=3 sexually active fruitflies (shaded lines) and n = 2 sexually inactive ones (black, reference group). *You don't need to complete the calculations, but lay out the steps so that your research assistant could complete the calculations by hand.*

  How many *informative* 2 × 2 tables are involved?

- Show how you could make one log-rank test that combined the information from the fruitflies with above average thorax lengths (rightmost) and those with below average thorax lengths (5 leftmost, déjà vu). *Again, detailed calculations are not needed, but steps should be clearly indicated. [approach is similar to what Mantel and Haenszel did to match not just on age (like time here), but simultaneously on occupation]*
- Say you had to do these log-rank calculations by computer, but could not extract the exact failure (death) times from Figure 3 (they are not to scale anyway). Is the result of the log-rank test sensitive to inaccuracies in your extraction of the death times? For example, for the analysis restricted to the analysis of the leftmost 5 flies, would you get a different answer if you entered their 'lifespans lived' as 30, 50, 53, 55 and 70 days, rather than as (say) 10, 20, 25, 45 and 50 days?

  In the analysis that pools results over the 2 size strata, would it matter if *you* extracted the 5 times for the "above average thorax size" flies, measured in days, and your *collaborator* extracted those for the 5 flies with below average thorax lengths but measured longevity in weeks? Why/why not?