Soc 756

Lifetable Estimates Using Sullivan's Method

In this problem set, you will be calculating gender differences in the expected duration of years lived in poverty for the United States in 2004. You will calculate variance estimates for these expectancies to determine if the estimates of differences between men and women are statistically meaningful, and at what ages.

1. Sullivan's method requires high quality lifetable data to construct total person years in each age group. Set up an account on the Human Mortality Database: https://www.mortality.org/ and get the nLx values for U.S. males and U.S. females in 2004 from the 5x1 abridged tables, e.g. for women: https://www.mortality.org/hmd/USA/STATS/fltper_5x1.txt

Install the R package HMDHFDplus and get the lifetable values directly: https://cran.r-project.org/web/packages/HMDHFDplus/HMDHFDplus.pdf The commands ask you to supply your user name and password, you'll still need to sign up at HMD first.

2. I have used General Social Survey data to estimate the poverty prevalence by age separately for men and women. I have defined the poverty prevalence as the proportion living at or below the poverty line at the time of the survey. You will find these values, as well as the size of the sample used to compute them in a table on the class webpage.

Because the GSS uses simple random sampling, we should be able to estimate standard errors using the simplified formula made available in Molla, Wagener, and Madans (2001), as opposed to correcting for the sampling scheme with a more complicated approach.**.

   (a) Calculate the expected number of years lived in poverty above age x separately for men and women ages 0-100 (using the given intervals).
   (b) Assess whether, at birth, the total expected number of years lived in poverty differs for men and women and whether this difference is statistically significant.
   (c) Assess these differences at each age interval.

You may use the approximation given in equation 13 of Molla, Wagener, and Madans (2001). A tip: pay careful attention to the notation on the components of each of these equations. For example, equation 11 uses the $_nL_x$ values, not the $_nL_x$ prime values.

***We must abstract away from reality in a few ways here: (1) the GSS does not include age values above 85. I have filled in the rest of the lifetable assuming that the poverty rates for the 85+ group could be applied to each of the sub-groups in that larger age intervals. I have also greatly increased the number of total respondents above age 85 by giving each subgroup above age 85 the same number of respondents. (2) The GSS does not interview children. I have pulled child poverty data from the National Center for Educational Statistics, assumed that the age-specific rates do not vary between boys and girls and have fabricated the sample sizes of children used to create these rates. Finally: note that I use the GSS here for illustrative purposes. How to measure poverty is nearly its own field and you are at the institution with people leading this field. Use the IRP as a resource if poverty measurement is important to your work: https://www.irp.wisc.edu/resources/how-is-poverty-measured/*