Feb 7, 2011

Computing for Research Homework Assignment

Stata 1

For this homework assignment, please turn in one document: the first portion should be the answers to the questions listed below. After the answers, at the end of the document, please cut and paste your do file so there is just ONE document submitted.

1. Open a new ‘do’ file in which to save all of your commands for this homework. Please use ‘comments’ to show which commands correspond to which question.
2. Breast cancer registry data:
   1. Read the dataset “Ohiosmall.dta” into stata. This is a very small subset of a larger dataset of the Ohio Breast Cancer Registry. It contains clinical information on women and their breast cancers, including cancer grade, stage, estrogen receptor status, progesterone receptor status, tumor size, age at diagnsosis, date of diagnosis and race.
   2. Use one or more stata commands to determine the number of observations and the number of variables in the dataset. How many of each are there?
   3. How many variables are coded as numeric and how many as ‘strings’? How did you find out?
   4. Generate a new variable which is equal to 1 if the patient has either ER or PR positive cancer and 0 if she is ER and PR negative. Determine an appropriate way to handle situations of missing in either ER or PR. Tabulate your new variable and explain how you handled the missings.
   5. Generate a new variable which is the month in which a woman was diagnosed and tabulate it. In this dataset, which month had the largest number of diagnoses?
   6. Drop 20% of the data by randomly selecting 20% of the observations to drop. Redo question e based on the new data
3. Case control study:
   1. A case control study was performed with 100 cases and 300 controls. Of the cases, 25 were exposed to a suspected pollutant. Of the controls, 39 were exposed to the suspected pollutant. Use an ‘immediate’ command in stata to determine the odds ratio, its 95% confidence interval and a p-value for testing that the odds ratio differs from 1.
   2. How would your answer to a change if I told you that the data arose from a cohort study design instead of a case control?
4. Close your do file!!