Methods in Epidemiologic Research

Sample Problems

Chapter 13 – Confounding

Preparation

We will carry on with the data that were used in the exercises for Chapter 6, but we will focus on the effects of being obese (bmi>30) or having coronary angioplasty on the risk of dying within 1 year of admission for the MI included in this dataset. All individuals in this dataset were followed for at least 1 year after admission, so there are no withdrawals (censored observations) to worry about.

The first thing you will have to do is generate four new variables.

- Create a 0/1 variable which indicates whether or not the individual died during the 1st year following admission (-died_1yr-).
- Create a variable indicating if the person was obese (bmi>30) or not (-obese-).
- Create 2 categorical variables for age:
  - a dichotomous (0/1) variable with age split at 75 years (-age_c2-),
  - a 5 category variable with splits at 60, 70, 80 and 90 years (-age_c5-).

Questions

1. Draw a causal diagram which you think represents the relationships among the following variables: death within 1 year of admission (-diead_1yr-), sex (-sex-), race (-white-), previous MI (-prmi-), cardiac angioplasty (-ptca-), obese (-obese-), and age (-age_c2- or -age_c5-).
   (a) Evaluate the correlations among all of these variables. (Note – correlations are not designed for use with dichotomous variables but they are a "quick and dirty" way of assessing the relationships among these variables. They certainly should not be used with categorical variables so you should leave -age_c5- out of this correlation matrix)
2. Compute and interpret the risk ratio (RR) for being obese.
3. Use matching to "control" for age when evaluating the effect of obese. Match on the 5-level version of age (-age_c5-). (Note – this takes quite a bit of programming skill, but (Stata) code is provided with the solution set).
   (a) Does controlling for age increase or reduce the apparent effect of being obese?
4. Now go back to the full dataset and analytically control (ie Mantel-Haenszel procedure) for the effect of -sex- on the effect -obese-.
   (a) Is -sex- an important confounder. If so, why? If not, why not?
5. Control for the effects of age in the analysis of -obese-.
   (a) First use the dichotomous version of age (-age_c2-). Is -age_c2- an important confounder. If so, why? If not, why not?
   (b) Next use the 5 category version of age (-age_c5-). Does controlling the effect of age with the 5-category version have a bigger or smaller effect than the 2-category version? Why?

6. We are now going to switch to evaluating the effects of angioplasty (-ptca-) on the risk of death.
   (a) First compute the crude RR for the effect of -ptca- on -died_1yr-. Interpret the result.
   (b) Now, control for age (-age_c2-). Is age a confounder? If so, why? If not, why not?
   (c) Next, control for whether or not the individual had had a previous MI (-prmi-). Is -prmi- a confounder? Is there evidence of interaction? What do these results tell you about the effect of -ptca-? If there is interaction, is there evidence of antagonism or synergism?
   (d) Evaluate interaction on the additive scale. (Note – to do this you will need to determine the risk difference (RD) separately for individuals with prmi=0 and prmi=1).
7. We will now switch to using propensity scores to help us evaluate the effects of -ptca-. Because angioplasty was not randomly assigned to individuals, it is very likely that those who had an angioplasty while in the hospital were different in some (or many) ways from those who did not get an angioplasty. Consequently, a simple comparison of ptca=0 vs ptca=1 is probably not appropriate. We would like to account for differences in the two groups in terms of: -sex- -age_c5- -white- -obese- -prmi-. The coding required to carry out the propensity analyses is a bit complicated, but the question we are addressing is very similar to the one used in Examples 13.10 – 13.12, so you can find Stata code for these analyses with the code for all of the analyses in Chapter 13.
   (a) First eliminate all records with any missing values for -ptca- or -sex- -age_c5- -white- -obese- -prmi-. (These will have to be ignored in the analysis anyway). Then compute propensity scores and determine if they meet the property of being balanced. When doing this, limit the computations to observations which fall in the region of "common support"
   (b) What is the range of propensity scores computed? Are there many observations outside the region of common support? Recompute the crude association between -died_1yr- and -ptca- using only the observations that fall in the region of common support (for comparison with later results).
   (c) Use the propensity scores to carry out nearest neighbour matching. Does this change the estimate of the effect of -ptca- by much?
   (d) Use the propensity scores to carry out a stratified analysis (based on the blocks created by the process of generating the propensity scores). Does this change the estimate of the effect of -ptca- by much.
   (e) Fit logistic models which control for the factors of interest by:

   Note – This question uses logistic regression models which are covered in Chapter 16. If you are not familiar with these types of models, you can skip this section.

      i. including propensity scores as a continuous predictor in the model,
      ii. including propensity scores as a categorical predictor (blocks) in the model, and
      iii. including the original predictors (-sex- -age_c5- -white- -obese- -prmi-) directly in the model instead of the propensity scores. Do these generally produce similar results? Are the results much different from what you get by ignoring this set of variables?