Lab #3

Biostatistics 210

NOT FOR HANDIN

0. BACKGROUND

This lab is meant to give you practice using propensity scores especially in the case of a continuous outcome.

1. DOWNLOAD

•Download the dataset lab3.dta. This is the HERS data which is used in Chapter 4 of the textbook (Vittinghoff, et al, 2nd edition). The selected variables are

• LDL (low-density lipoprotein in mg/dL)

• BMI (body mass index in kg/m^2)

• nonwhite (non-white race/ethnicity)

• smoking (1=smoker, 0 = non-smoker)

• age (in year)

• drinkany (1=current drinker of alcohol, 0= not current drinker)

The purpose of our analysis will be to estimate the causal effect of current drinking on mean LDL. The other variables above will be adjusted for as confounders.

2. CRUDE EFFECT

Perform a t-test to estimate the mean LDL among drinkers and non-drinkers

```
ttest LDL, by(drinkany)
```

3. PROPENSITY SCORES

Derive the propensity quintiles in the following steps

(a) Fit a simple propensity model for drinking

```
logistic drinkany smoking nonwhite age BMI
```

(b) Extract predicted probabilities of drinking for each person

```
predict out, p
```

(c) Create the variable “group” using “xtile” command which groups variables into equal sized. Here, we request 5 groups which creates categories based on quintiles

```
xtile group = out, nq(5)
```

4. OVERLAP

Overlap, requires that we have a good number of drinkers and non-drinkers at the levels of the propensity.

What does the table for overlap indicate?

```
tabulate drinkany group, col
```

Examine mean LDL value by drinking within quintile of propensity

```
table drinkany group, c(mean LDL)
```

5. QUINTILE ANALYSIS

```
reg LDL i.drinkany i.group
```

What is the estimated average causal effect of drinking on LDL using the quintile?

Use the margins command to estimate the causal mean LDL if the entire population was drinkers v. non-drinkers.

```
margins drinkany
```

Verify that the difference of these means corresponds to the coefficient in the above regression model.

```
margins, dydx(drinkany)
```

6. 20-LEVEL ANALYSIS

Create a 20 level spilt of the propensities to rule out the possibility of residual confounding in the quintile analysis

```
xtile group20 = out, nq(20)

reg LDL i.drinkany i.group20

margins drinkany
```

How do these results compare to the quintile analysis?

7. STANDARD ADJUSTED ANALYSIS

In a standard adjusted analysis we would do a linear regression on LDL with drinking adjusting for the potential confounders.

```
reg LDL i.drinkany smoking nonwhite age BMI
```

What is the estimated effect of drinking on LDL in the standard regression?

```
margins drinkany
```

What are the estimated mean LDL for drinkers and non-drinkers? How to these results compare to the propensity analysis?

8. HIGH DIMENSIONAL PROPENSITY MODEL

Fit a very rich propensity model which includes splines and interactions.

```
mkspline agesp = age, nknots(5) cubic
mkspline bmisp = BMI, nknots(5) cubic

logit drinkany smoking nonwhite smoking##c.agesp*  smoking##c.b-
misp* nonwhite##c.agesp* nonwhite##c.bmisp*
```
(the above needs to be written on one line)

```
predict outsp, p

xtile groupsp = out, nq(5)
```

Assess the overlap of the quintile groups “groupsp” based on the richer model and estimate the average causal effect based on it. Have your conclusions changed?