**Matching Design (Non-Equivalent Control Group Design)**

**(Due date: April 17)**

**Background**

The data for this module come from an experiment conducted at a Midwestern University (Shadish, Clark & Steiner, 2008). Here we use only a part of the data where 210 students could choose on their own whether they want to attend a short vocabulary or mathematics training. 79 students chose the math training while 131 selected into the vocabulary training. In order to statistically control for the systematic selection into the two trainings, a set of baseline covariates was measured (before students made their choice). The 23 covariates belong to five major domains: Proxy-pretests of the outcome, prior academic achievement, topic preference, psychological predisposition, and demographics. After the training, students were assessed in both mathematics and vocabulary.

**Data**

The data for the module have already missing data and errors recoded and cleaned. The datasets are the same as we used already in class (shadish&clark\_imputed\_QE\_nooutcome.sav and shadish&clark\_imputed\_QE.sav) includes the following variables:

|  |  |
| --- | --- |
| **Treatment variable** |  |
| vm | Treatment indicator for vocabulary (v) or mathematics (m) training | |
|  |  |
| **Outcomes** |  |
| mathall | Mathematics posttest (measured after the completion of the training) | |
| vocaball | Vocabulary posttest (measured after the completion of the training) | |
|  |  |
| **Baseline covariates by domain (measured before treatment selection)** | |
| *Proxy-Pretests* |  | | |
| vocabpre | Vocabulary Pretest (36-item Vocabulary Test II; Educational Testing Service, 1962, 1993)) | | |
| mathpre | Mathematics Pretest (15-item Arithmetic Aptitude Test; Educational Testing Service, 1962, 1993)) | | |
| *Prior Academic Achievement* | | | |
| actcomp | ACT Comprehensive Score | | |
| hsgpaar | High School GPA | | |
| collgpaa | College GPA | | |
| *Topic Preference (Preference for Math or Literature)* | | | |
| numbmath | Number of Prior Mathematics Courses | | |
| likemath | Like Mathematics | | |
| likelit | Like Literature | | |
| preflit | Preferring Literature over Mathematics | | |
| majormi | Math-Intensive Major | | |
| mars | Mathematics Anxiety Rating Scale (25 items; Faust, Ashcraft & Fleck, 1996) | | |
| *Psychological Predisposition* | | | |
| pextra | Extroversion (10 items; big 5 personality factors, Goldberg, 1999) | | |
| pagree | Agreeableness (10 items; big 5 personality factors; Goldberg, 1999) | | |
| pconsc | Conscientiousness (10 items; big 5 personality factors; Goldberg, 1999) | | |
| pemot | Emotionality (10 items; big 5 personality factors; Goldberg, 1999) | | |
| pintell | Intellect/Imagination (10 items; big 5 personality factors; Goldberg, 1999) | | |
| beck | Beck Depression Scale (13 items; Beck & Beck, 1972) | | |
| *Demographics* |  | | |
| cauc | Race/Ethnicity: Caucasian (dummy) | | |
| afram | Race/Ethnicity: African-American (dummy) | | |
| other | Race/Ethnicity: Other ethnicity (dummy) | | |
| age | Age | | |
| male | Sex: male (dummy) | | |
| married | Marital status: married (dummy) | | |
| momdegr | Mother’s Education (presumably measured with low reliability) | | |
| daddegr | Father’s Education (presumably measured with low reliability) | | |
| credit | College Credit Hours | | |

**Assignment:**

Estimate the *average treatment effect for the treated* (ATT) for the *vocabulary training*. vm is the group indicator with the vocabulary group representing the treatment group and the math group representing the control group. The outcome of interest is the *vocabulary achievement score* (vocaball).

**Step 1. Assess whether the set of observed covariates might meet the strong ignorability assumptions.**

Shadish, et al., (2008)’s dataset includes a large set of variables that cover the confounders. Most important determinants of the selection process and the outcome model are quite reliably

measured. We can be confident that most of them confounding bias can be removed.

**Step 2. Assess the initial imbalance in baseline covariates.**

Check how different the treatment and control groups are with respect to the baseline covariates. Use graphs and descriptive statistics to assess the initial imbalance. (This part is the same for all estimation techniques; and we already did it in class.)

**Step 3. For different PS and regression estimators, estimate the PS and the ATT**

**Part 1. Estimate the ATT using inverse propensity weighting  
Part 2. Estimate the ATT using PS stratification** *R* hint: In computing the expected frequencies for the marginal mean weights you may use  
 outer(apply(O, 1, sum), O['Vocabulary', ] / sum(O['Vocabulary', ])))

**Part 3. Estimate the ATT using matching** *R* hint: Use the pairmatch() function from the optmatch package if you want to do   
 a 1:1 matching;  
 Alternatively, you can do a full matching, but when computing the marginal ATT weights,   
 compute the expected frequencies according to the hint for the stratification approach.

**Part 4. Estimate the ATT using regression estimation**

For *two out of the four* estimation strategies do the following (i.e., you only need to do two parts of your choice):

1. **PS estimation:** Estimate the PS using a logistic regression model.
2. **Balance checks:** Check balance (using the same method as you use for estimating the treatment effect later). If the initial PS model does not establish sufficient balance in covariates, try to re-specify the PS model, and check balance again (and so on …; but at some point you need to stop, even if you do not get good balance).
3. **Effect estimation (without further covariate adjustments):** Estimate ATT and its standard error using the propensity score alone (weighting, stratification, matching, or regression).
4. **Effect estimation (with additional covariate adjustments):** Estimate ATT and its standard error using a doubly robust adjustment, i.e., combine the PS adjustment with an additional covariance adjustment in an outcome regression model.

**Report**

Write a short report regarding Steps 2-3 for *two* estimation strategies of your choice (i.e., choose two out of four parts in step 3). The write-up should contain the following:

(a) a plot showing the *initial imbalance* in baseline covariates and a short discussion thereof,

(b) the *propensity score model equation* (for each PS technique if they differ),

(c) a *balance plot* for each PS technique and assessment of the achieved balance,

(d) the *outcome model equations* (with and without covariates) used for estimating ATT,

(e) a *table* that summarizes all 4 ATT estimates (without and with additional covariate adjustment) and their standard errors, and

(f) a brief discussion of results (the size and confidence interval (significance) of effects, differences across estimation methods).

The write-up should not exceed 3 pages (plots can go into an appendix).