Homework for Chapter 14: Matching

1. You want to know whether practicing cursive improves your penmanship (on a 1-10 scale). You find that, among people who don’t practice cursive, average penmanship is 5, 10 people are left-handed, 2 are ambidextrous, and 88 are right-handed. Among people who do practice cursive, 6 are left-handed with average penmanship 7, 4 are ambidextrous with average penmanship 4, and 90 are right-handed with average penmanship 6.
   1. You want to create a set of weights that will *make the treated group match the control group on handedness*. Follow the process in section 4.2, paying attention to *why* certain numbers are going in certain positions. What weights will be given to the left, ambidextrous, and right-handed people *in the control group*?

Give a weight of 1 to everyone who is not treated.

* 1. What weights will be given to the left, ambidextrous, and right-handed people *in the treated group*?

Weight given to the left-handed people in the treated group: 10/6≈1.667

Weight given to the ambidextrous people in the treated group: 2/4=0.5

Weight given to the right-handed people in the treated group: 88/90≈0.978

* 1. Use the weights from part b to calculate the *proportion of left-handed people in the treated group*, as well as the proportion of ambidextrous people and the proportion of right-handed people. If you don’t get 10%, 2%, and 88% (or very close with some rounding error), your weights are wrong, try again.

Proportion of left-handed people in the treated group= 1\*1.667\*6/(1.667\*6+0.5\*4+0.978\*90) ≈0.01

Proportion of ambidextrous people in the treated group =1\*0.5\*4/(1.667\*6+0.5\*4+0.978\*90) ≈0.02

Proportion of right-handed people in the treated group = 1\*0.978\*90/(1.667\*6+0.5\*4+0.978\*90) ≈0.88

* 1. What is the weighted average penmanship score in the treated group?

Weighted average penmanship score in the treated group =

(1.667\*7+0.5\*4+0.978\*6)/(1.667+0.5+0.978) = 6.212083

* 1. What is the effect of practicing cursive that we would estimate using this data?

Weighted average penmanship score in the untreated group = 5\*1=5

Average treatment on the untreated = 6.212083-5=1.212083

1. For each of the following descriptions of matching on the variable , determine whether this is describing *one-to-one distance matching, k-nearest-neighbor distance matching, kernel matching,* or *propensity score matching* (hint: it’s one of each)*.*
   1. The treated observation has . For each control observation, is calculated, with the result run through a weighting function. The resulting weight is applied to that observation.

*kernel matching*

* 1. The treated observation has . Among the control observations, the nearest values are and . The observations with and are chosen as a control, since they’re the two closest.

*k-nearest-neighbor distance matching*

* 1. The treated observation has . You estimate a model that suggests that observations with have a .6 chance of being treated. You similarly calculate the chance of treatment for each control observation, and use those calculated probabilities to create a weight for each observation.

*propensity score matching*

* 1. The treated observation has . Among the control observations, the observation with is closest to that, and so is selected as a control.

*one-to-one distance matching*

1. For each of the following decisions to be made in the process of matching, determine which option produces *more bias* (in each case, the other option will produce *more variance*)
   1. (A) selecting one control match for each treatment vs. (B) selecting multiple control matches for each treatment

(B)

More but worse matches produce estimates with more bias but more precision.

* 1. (A) using a relatively wide bandwidth vs. (B) using a narrower bandwidth

(A)

When we make the bandwidth wider, we’re allowing in more bad matches, which makes the quality of match worse and the idea that we’re closing back doors less plausible. This brings bias back into the estimation.

* 1. (A) selecting matches with replacement vs. (B) selecting matches without replacement

(B)

Matching with replacement ensures that each treated observation gets to use its best (or k best, or all acceptable) matches. This reduces bias because this approach lets us pick the best matches. However, this approach means that we’re using the same control observations over and over - each control observation has more influence on the mean, and so sampling variation will be larger.

* 1. (A) selecting one control match for each treatment vs. (B) applying a weight that accepts many controls but decays with distance

(B)

In general, when making any sort of decision about matching, including this one, the choice is often between fewer but better matches that produce estimates with less bias but less precision, or more but worse matches that produce estimates with more bias but more precision.

1. Why should exact matching (or coarsened exact matching) generally be reserved for very large samples or situations where a very small number of matching variables is appropriate?

Because exact matching or coarsened exact matching, if applied to not-large-enough-sized samples or any size sample with too many matching variables, can lead to lots of treated observations being dropped due to not finding any matches. Lots of treatment observations ended up getting dropped will make the treatment effect estimates much noisier, and can also lead the result to be a poor representation of the average treatment effect if certain kinds of treated observations are more likely to find matches than others.

1. You are looking at the effect of participating in high school sports on high school grades. You compare students who did and did not participate in sports, using one-to-one matching with a Mahalanobis distance, with replacement and a caliper of .3, to match on high school athleticism, parental income, gender, race, and middle school grades. You find that sports participation reduces grades, but by only .1 grade points. As clearly and precisely as possible, outline the steps that were taken in performing this analysis.
2. First, take each matching variable and divide its value of the observation by its standard deviation. Now, each matching variable has a standard deviation of 1. This makes sure that no variable ends up being weighted more heavily just because it’s on a bigger scale. Then we can calculate distance. For a given treated observation A and a given control observation B, the Mahalanobis distance is the sum of the squares of all the differences between A and B. After we’ve taken the sum, we take the square root. In other words, it’s the sum of squared residuals we’d get if trying to predict the matching variables of A using the values of B, if the standard deviation of each matching variable was 1. That’s the Mahalanobis distance. Or, we can put the differences in matching variables between an untreated and the treated observation next to the covariance matrix and get a Mahalanobis distance.
3. Using one-to-one matching, we select untreated observations with lowest Mahalanobis distances to match the treated observations. An untreated observation can be selected as the best matches for multiple treated observations. If even the closest Mahalanobis distance is beyond the caliper of 0.3, we drop the treated observation. We perform this by using R function Match() in the Match package.
4. Use summary() function to calculate the post-matching average treatment on the treated effect of high school sports participation on high school grades.
5. Which of the following is a downside of propensity score matching compared to other methods of matching? d (\*b is a fact but seems not a downside)
   1. It can’t be combined with exact matching in cases where one variable must be exactly matched
   2. It focuses on matching adjustment on differences that close back doors, rather than all differences
   3. It requires the selection of matches instead of the use of weights, which increases variance.
   4. It requires that the model used to estimate the propensity score is properly specified.
6. You are planning to evaluate the effect of a tax-rebate plan for small businesses. Some businesses were eligible based on their tax returns and others weren’t. You would like to match on industry and number of employees. A table showing the number of businesses for each combination of industry and number of employees for the treated and untreated groups are in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Number of small businesses by industry, number of employees, and treatment status | | | | | |
|  | Treated |  |  | Untreated |  |
| N. Employees | Retail | Service |  | Retail | Service |
| 1-5 | 3 | 4 |  | 0 | 4 |
| 6-10 | 3 | 2 |  | 4 | 3 |
| 11-20 | 0 | 5 |  | 5 | 1 |

* 1. For what group of treated businesses would we say that the common-support assumption definitely fails?

The Retail group with 1-5 employees.

* 1. There are no treated retail businesses with 11-20 employees. Is this a concern for the common support assumption if we are trying to estimate an average treatment on the treated?

No. If we are trying to estimate an average treatment on the treated, we are making the untreated group match the treated group. Since there is no treated observation of Retail business with 11-20 employees, it means the group is more likely to be untreated and there is no need to match. That’s not a problem.

* 1. What concern might we have about there only being one untreated Service business with 11-20 employees?

We cannot find enough comparable control observations matching on the 5 observations of the treated Service business with 11-20 employees from only one untreated Service business with 11-20 employees or at best we would have that one observation match all the treated observations - the sample mean for the controls would just be that one observation’s outcome value and would have a standard error that’s just equal to the standard deviation of the outcome. Dividing by √N to get the standard error doesn’t do much if N = 1.

* 1. If we resolved the common support problem for the group from problem (a) by dropping members of that group from the data, what problem would that create for our analysis?

By dropping members of treated Retail group with 11-20 employees from the data, the comparison won’t be all that representative of the actual groups with Retail business with 11-20 employees and the result would be a poor representation of the average treatment effect.

1. You perform a matching analysis on a schooling reform to create a set of matching weights, matching on the per-capita income and expenditures of the school. You then produce the below weighted balance table comparing the weighted means for treatment and control.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Treated | |  | Untreated |  |  |
|  | N | Mean | SD | N | Mean | SD | Test |
| Expenditure | 29 | 389 | 106.677 | 21 | 351.524 | 71.529 | F=1.951 |
| Income | 30 | 7749.7 | 1127.359 | 21 | 7406.381 | 888.136 | F=1.356 |
| Matching weights applied. Statistical significance markers: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01 | | | | | | |  |

* 1. This particular balance table reports F-statistics of differences in means, with statistical significance markers. Are there statistically significant differences in either of the variables between the treated and untreated group at the 95% level?

No.

* 1. You don’t have enough information to actually evaluate this, but make a list of two things you’d think about when deciding whether it looks like there’s a balance problem based on the difference in means regardless of whether the difference is statistically significant. As an example, answer while thinking of the difference of 7749.7 – 7406.4 = 342.3 between treated and untreated in Income.

A standardized difference in means, where the difference between treated and control groups is divided by their shared standard deviation, or standardized bias, where the difference in means is divided by a shared variance term, which we get by adding up the variance in the treatment group with the variance in the control group, multiplying the result by .5, and then taking the squared root. Constructing the bal.plot to see the distributional balance for distance before and after matching, or the love.plot to see the absolute standardized mean differences.

* 1. Imagine you *did* find lots of significant differences here after constructing matching weights using propensity score matching, even though these variables were included as matching variables. What would your next step be?

Going back to tweak the matching method, e.g., adding more variables, trying different functional form for a propensity score model, using a tighter caliper, etc. Do iterative matching/balance checking until the balance looks good: no meaningfully large differences between means of the treated and control groups in matching variables.

1. Explain why selecting untreated observations to match the treated observations produces an average treatment effect on the treated (ATT), while selecting treated observations to match the untreated observations produces an average treatment effect on the untreated (ATUT).

By selecting untreated observations to match the treated observations, we want to make the untreated group comparable to the treated group according to a set of matching variables, in which case we adjust for the matching variables to close all the back doors between the treatment and the outcome. In doing so, we are indeed comparing our estimate of what the treated group would have gotten without treatment against what the treated group actually got. This is the average treatment effect on the treated - the average treatment effect among those who actually received the treatment in the study.

By selecting treated observations to match the untreated observations, we want to make the treated group like the untreated group according to a set of matching variables. We adjust for the matching variables to close all the back doors to make the untreated group and the treated group comparable to compare what we actually got from no treatment against what we think that group would have gotten if treated. This is the average treatment effect on the untreated - the average treatment effect among those who did not actually receive the treatment in the study.

1. Coding