

Datastat5

Himangshu Raj Bhantana

11/14/2021

```
## cobalt (Version 4.3.1, Build Date: 2021-03-30 09:50:18 UTC)

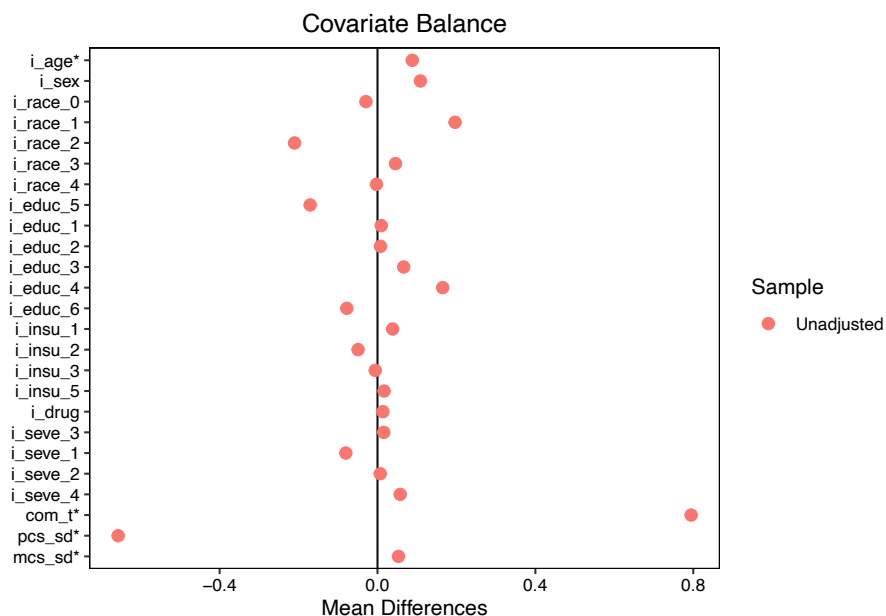
##
## Attaching package: 'cobalt'

## The following object is masked from 'package:MatchIt':
##
## lalonde
```

Part 1. Asthma Patients in California

Are the covariates in this data balanced between the two groups? If no, which covariates are not? How did you assess balance?

Covariates isn't balanced between two groups, there were 8 unbalanced covariate `i_sex`, `i_race_1`, `i_race_2`, `i_educ_4`, `i_educ_5`, `com_t`, and `pcs_sd` which also can be seen below in our plot. We use the ASD (absolute standardized difference) to assess the balance, where the absolute standardized difference value is greater than 0.1.

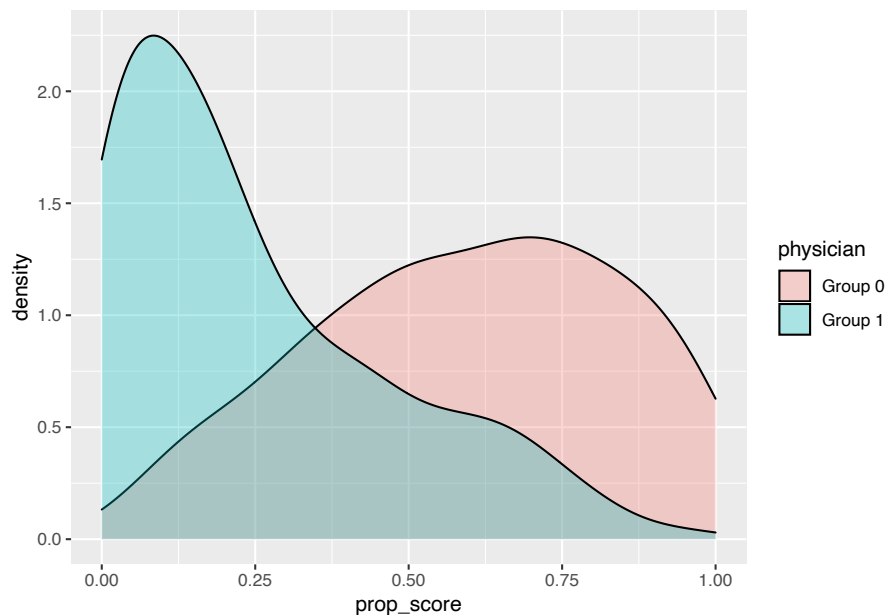


Part2

Question 1

Are there any observations with an estimated propensity score e that is out of the range of e in the other group? If there are only a few such outliers (less than 5), keep them; If many, discard them and report the number of the discarded observations.

There are observations with an estimated propensity score e that is outside the range of e in the other group; there were more than 5, therefore we chose to discard them; the number of observations discarded was 48, with one group having a number of 40 and another having a number of 8.



Question 2

Using one-to-one, nearest neighbor matching on the estimated propensity scores, check balance again. Are the covariates balanced now? If no, which ones are not?

Using one-to-one matching yielded 97 matches, but balancing left the covariates uneven. The unbalanced ones are `i age`, `i race2`, `i race4`, `i educ6`, `i insul`, and `mcs sd`.

Question 3

Estimate the average causal effect Q “directly” using the matched sample obtained above. Also, report a standard error for your estimate (use the formula for computing standard error for difference in proportions; if you are not familiar with this, check page 280 of the third edition of the OIS book we used for the online summer review). Construct a 95% confidence interval and interpret your findings.

We obtained an average casual/treatment impact of 0.16, rounded from 0.155. In addition, I calculated the standard error for my estimate, which came out to be 0.065. Creating a 95% confidence interval, which ends out to be 0.03 (real 0.028) and 0.282. This 95% confidence interval can be read as follows: because 0 appears to be in our interval, casual/treatment affect is substantial.

Question 4

Fit a logistic regression to the response variable using the main effects of all pre-treatment variables on the matched data. Also include the treatment variable and the propensity score e as predictors. Report the estimated causal odds ratio. If it is significant, interpret the effect in context of the problem. Note that this estimated effect is not an estimate of $Q = p_2 - p_1$ but intuitively, it still makes sense to look at it.

Fitting logistic regression to the response variable using the main effects of all pre-treatment variables on the matched data, as well as including the treatment variable and the propensity score e as predictors, we can report that the estimated casual odds ratio is 2.51, with the treatment variable also appearing to be significant, indicating that the treatment group (group 0) has higher satisfaction than the control group (group 1)..

Question 5

Repeat parts (b) to (d) using one-to-many (five) nearest neighbor matching with replacement, instead of one-to-one nearest neighbor matching. How do your results compare to what you had before?

We still have unbalanced co-variates when we use one-to-many (five) matching instead of one-to-one closest neighbor matching, but we do observe a shift in the average treatment value compared to the previous one, which was 0.16 rounded from 0.155 and is now 0.17 rounded from 0.166. The confidence interval likewise turns out to be 0.04 and 0.3, not zero, while the standard error for the estimate comes out to be 0.064. We may also report that our treatment variable is still significant, with an estimated casual odds ratio of 2.41 (rounded). Overall, treatment group 0 has a 17 percent higher level of satisfaction than control group 1 and also receives superior care.

Question 6

Which of the methods do you consider most reliable (or feel most comfortable with) for estimating the causal effect? Why?

One to many matching, in my opinion, is quite reliable for estimating the causal effect because it appears that more than one match has additional options to explore and provide proper results, as opposed to one to one matching, which is much more conservative and only looks at one nearest neighbor where there is a low probability of not being matched. From my perspective, having a one-to-many option is superior to one-to-one.