Lab 6: Causal Inference

Srishti Saha (ss1078), Altamash Rafiq (ar327) and Ronald Nhondova (rtn6)

18 November, 2019

Question 1

Using Absolute Standardized Difference (ASD) with an absolute value threshold of 0.1 as our metric for evaluating covariate balance, we found that many of the covariates are not balanced (have absolute standardized difference > 0.1). The covariates that are imbalanced are as follows:

	Type	Diff.Un
cat1_MOSF w/Sepsis	Binary	0.1721039
resp_Yes	Binary	-0.1276883
card_Yes	Binary	0.1394948
neuro_Yes	Binary	-0.1078969
aps1	Contin.	0.4837124
scoma1	Contin.	-0.1160363
wtkilo1	Contin.	0.2639761
meanbp1	Contin.	-0.4868502
resp1	Contin.	-0.1641331
hrt1	Contin.	0.1460095
pafi1	Contin.	-0.4566445
paco21	Contin.	-0.2880367
ph1	Contin.	-0.1162506
hema1	Contin.	-0.2954146
crea1	Contin.	0.2678085
bili1	Contin.	0.1329124
alb1	Contin.	-0.2009522

Question 2

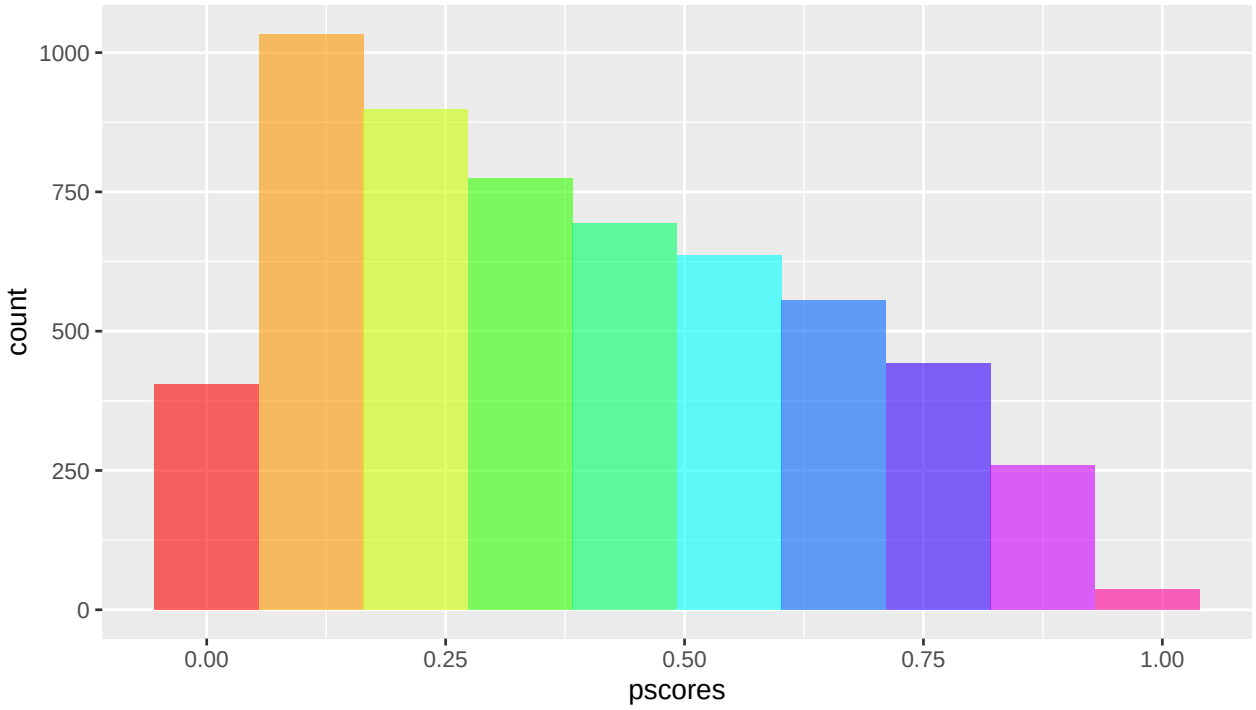
The estimated ATT of the patients with RHC = TRUE is 0.0755. This was obtained by subtracting the means of the predicted odds of death of the patients who originally received the treatment versus when the treatment flag was turned to FALSE for them (p0-p1).

Question 3

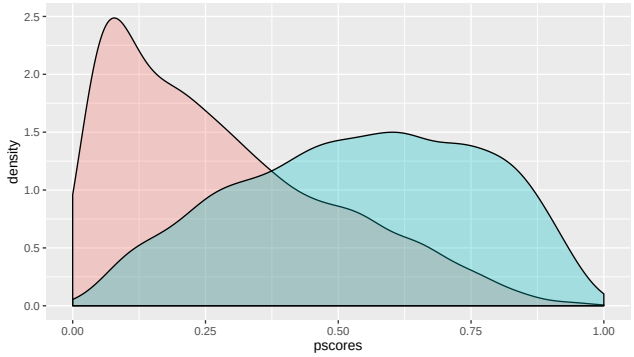
Based on our estimated effect, the treated patients might be better off with RHC. The decrease in chances of survival (or increase in chances of death) is 7.55%. We do not trust the conclusions because of the imbalance in data which is likely to skew the results. The confidence interval for the ATT is 0.048 to 0.099 i.e. (4.8% to 9.9%).

Question 4

As per our histogram of the estimate propensity scores below, not many of the scores are near 1. As such, we are not going to remove any outliers based on this histogram.

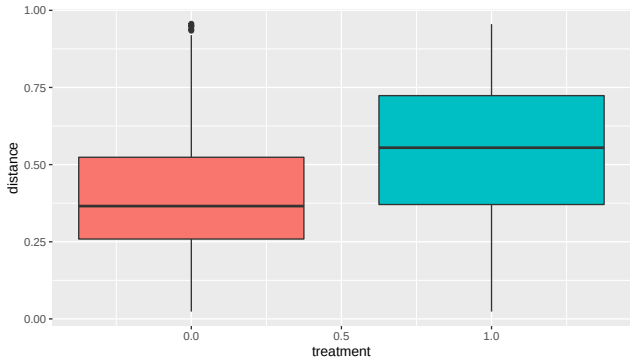


Next, we made the following density plots to assess overlap and found that 103 observations are violating overlap. We decided to exclude these observations and discarded them.



Question 5

After conducting one-to-one matching we checked for covariance balance and found that the covariances are now balanced.



The covariates that are still not balanced are :-

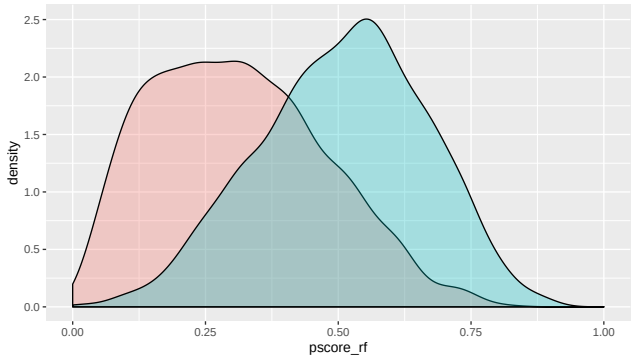
	Mean Diff.
raceother	-15.717242
cat1MOSF w/Malignancy	-242.757994
hemaYes	-5.539349
caYes	-4.600729
temp1	-69.124492

Question 6

Treated patients are worse off with RHC as the effect is still high (significantly different from 0). The treatment effect is now reported to be 7.91% i.e. the odds of death at 30 days is now 7.91% higher given the treatment. The 95% confidence interval of this is 5.10% to 10.71%.

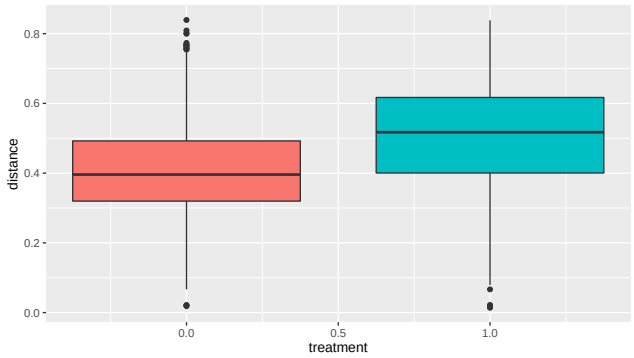
Question 7

After estimating propensity score using random forest and assessed overlap (see the density plot below). We also determined that 13 observations violated the overlap and these observations were therefore dropped.



Question 8

After doing one-to-one matching using the new propensity scores, we found that the covariates are more balanced now.



The following covariates remain unbalanced :-

	Mean Diff.
racewhite	-276.5124555
ninsclasMedicare & Medicaid	-11.2488568
cat1MOSF w/Malignancy	-259.9882318
cat2Colon Cancer	-157.6000000
cat2Lung Cancer	-0.7824726
cat2MOSF w/Malignancy	-15.8350543
metaYes	-178.0091417
hemaYes	-108.1965582
orthoYes	-84.7919656
caNo	-27.9441718
caYes	-27.8200463
temp1	-91.8232875
resp1	-10.0477561
pot1	-193.7783823
renalhxTRUE	-61.7074702
malighxTRUE	-27.3875625

Question 9

The 95% confidence interval now is 1.89% to 7.59%. The new ATT is **4.7%**, this means that the treatment increases the odds of death by 4.7% as compared to patients who are not treated.

Question 10

The logistic regression model coefficient for Treatment (level= Yes) is 0.3568. This means that given all other factors are constant and controlled for, a person who has received the treatment has 42.9% higher odds of death than the person who did not receive the treatment. This is in line with the observations in Q9 where we see that the effect of treatment on the odds of death is incremental.