Online Test 5

Instructions: The test is worth up to 5 points. Each correct response is worth 1 point. There is a bonus question. Download the files "DataAnalyticsFunctions.R", "CreditRatingtoPost.csv" and "OnlineTest5.R" to your working directory for this assignment. Open and execute the code "OnlineTest5.R" to answer the following questions. The data set for this assignment relates to the data we discussed in class.

The Rubin model is currently the preferred way to conceptualize causality questions. It allows you to clearly formulate the potential outcomes (i.e. counterfactuals) which in turn allow you to speak properly about causal effects. Consider the OK Cupid discussion in class. Let Y be the (binary) target variable indicating if a message led to a response (Y=1) or if the initial message was not responded (Y=0). We discussed the binary "treatment" of using the word "beautiful" in the message (d=1) or not using the word "beautiful" (d=0). it provides you a way to understand potential bias and the estimands we are looking for.

1. Which of the following relations led to OK Cupid's conclusion about the use of the word "beautiful"?

a) $E[\ Y \mid d=1] > E[\ Y \mid d=0\ ]$
b) $E[\ Y(1) \mid d=1] > E[\ Y(0) \mid d= 0]$
c) $E[\ Y(0) \mid d=1] > E[\ Y(0) \mid d=0]$
d) ✓$E[\ Y \mid d=1] < E[\ Y \mid d=0\ ]$

OK Cupids' conclusion was based on predictive arguments using d=1 if used the word "beautiful", d=0 otherwise. In that case they computed $E[\ Y \mid d=1]$ and $E[\ Y \mid d=0]$. It follows that $E[\ Y \mid d=1] < E[\ Y \mid d=0]$ (the response rate for using "beautiful" was below than the baseline response rate for all messages, in particular below to not using "beautiful"). Then option d) is correct.
Note that option b) and c) is based on the potential outcomes and option a) would give the reverse conclusion from the reading.

2. For those who chose to use the word "beautiful", which of the following measures the actual average impact on response rates of using the word "beautiful" in a message?

a) $E[\ Y(1) \mid d=1] - E[\ Y(0) \mid d=0\ ]$
b) ✓$E[\ Y(1) \mid d=1] - E[\ Y(0) \mid d=1]$
c) $E[\ Y(0) \mid d=1] - E[\ Y(1) \mid d=0]$
d) $E[\ Y(0) \mid d=0] - E[\ Y(1) \mid d=0]$

This requires the use of the potential outcomes. We are asking for the average treatment effect on the treated (ATT). The average impact of using "beautiful" among those who choose to do it (d=1) is

E[Y(1)-y(0) | d=1] = E[Y(1) | d=1] - E[Y(0) | d=1]

which is option b). Option a) is what OKCupid did. Option c) and d) involve people who did not use "beautiful" (there is a conditioning d=0).

3. Considering the reasoning discussed in class to rationalize why the use of the word "beautiful" led to a lower response rate. That is, people who received a message with "beautiful" are more likely to attract more messages and his response rate goes down simply by lack of time to respond to all. If including the word "beautiful" in the message actually increase the response rate, it would imply that:

a) E[Y(0) | d=0] < E[Y(0) | d=1]
b) E[Y(1) | d=0] < E[Y(0) | d=1]
c) **✓E[Y(0) | d=1] < E[Y(1) | d=1]**
d) E[Y(0) | d=0] < E[Y(1) | d=1]

*We use the potential outcomes framework to be able to write this. If including "beautiful" actually increases the response rate among those who receive the treatment (d=1) it means that the average of the potential outcome Y(1) given d=1, **E[Y(1) | d=1],** is larger than the average of the potential outcome Y(0) given d=1, **E[Y(0) | d=1]**. This corresponds to option c).*

Note that all the other options compare across people that received the treatment (d=1) and did not receive the treatment (d=0).

4. Consider the example of measuring the impact of obtaining credit ratings on the leverage of the firm discussed in class. Consider the calculations based on the data obtained via an observational study

```
> aggregate(Y ~ d, FUN = mean, data= CR)
  d         Y
1 0 0.2868934
2 1 0.4248909
> resNaive <- aggregate(Y ~ d, FUN = mean, data= CR)
> resNaive$Y[2] - resNaive$Y[1]
[1] 0.1379975

> res <- with(CR, Match( Y=Y, Tr=d, X=profit+finance))
> summary(res)

Estimate...  0.055603
AI SE......  0.0027764
T-stat.....  20.027
p.val......  < 2.22e-16

Original number of observations..............  585
Original number of treated obs...............  458
Matched number of observations...............  458
Matched number of observations  (unweighted).  2202
```

How much is the selection bias (see class slides)?

a) ✓**`0.0823945`**
b) `0.1379975`
c) `0.2868934`
d) `0.3248909`

In the class slides we discussed the selection bias and we have that
E[Y|d=1] - E[Y|d=0] = ATT + selection bias
where ATT is the average treatment effect on the treated (which is computed by the "Match" function). The left hand side is the easy part to compute using the data (e.g. via the "aggregate" function).

Based on the output above we have that
ATT=0.055603 and
E[Y|d=1] - E[Y|d=0] = 0.1379

Thus the selection bias is given by 0.1379 - 0.055603 = 0.0823945 which is option a).

5. Consider the example of measuring the impact of obtaining credit ratings on the leverage of the firm discussed in class. Consider the result for the t-test in the controlled experiment

```
> t.test( Yexp ~ dexp, data= CR )

        Welch Two Sample t-test

data:  Yexp by dexp
t = -4.0788, df = 582.979, p-value = 5.157e-05
alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
 -0.08457684 -0.02959857

sample estimates:
mean in group 0 mean in group 1
      0.3515260       0.4086137
```

Is this result consistent with the result obtained via matching based on the propensity score (using the Match function) in the data from the observational study? Why?

a) No, the estimators are very different (negative confidence interval in the controlled experiment while a positive confidence interval in the observational study), therefore they are not consistent.
b) No, the estimator for the ATT in the experiment is `0.4086137, while 0.055603 in the other case.`
c) ✓**Yes, they are consistent as their difference, `0.0570877-0.055603` is less than 1 SE.**
d) Yes, they are consistent because both are small.

We need to measure discrepancies accounting for sampling uncertainties. Indeed the estimator obtained via the "Match" function is 0.055603 and the estimator obtained via an experiment is 0.4086137-0.3515260 = 0.0570877 (the difference between the groups). Note that their difference 0.0570877-0.055603 is less than 1SE which is intrinsic fluctuation from the random sampling. There are no discrepancies (option c is correct).

6. In class we discussed using Lasso to select variables for a linear regression model. However, in our causal modeling discussion, we learned that if we want to use the p-value to see test the null hypothesis associated with the coefficient of a specific variable d, model selection needs to account for the selection mechanism. In particular, we saw the "double selection" technique to construct confidence interval. Which of these statements is false?

a) After running the double selection only one p-value is valid.
b) The traditional Lasso or post-Lasso to have a better out of sample R^2 performance than the double selection is somewhat expected.
c) If d is assigned independent of the other variables as in a controlled experiment, we expect the Lasso and double selection to select the same variables.
d) ✓Although Lasso does not have p-values, if we run a regression on the selected variables (i.e. Post-Lasso) we will have valid p-values.

Option d) is incorrect. Post-Lasso will not be reliable to construct confidence intervals for the coefficient of the treatment. Indeed, it ignores the "assignment of the treatment" (e.g. propensity score) in the selection of the variables. The idea of the double selection is to precisely correct that. Such Post-Lasso approach was essentially the approach used in the Freakonomics book which missed important variables leading to omitted variable biases (which led to their conclusions).