Manzano Romulo - Problem Set 3

October 17, 2017

1 Problem Set #1

1.1 Experiments and Causality

Romulo Manzano (15Oct2017)

```
In [1]: # load packages
        library(data.table)
        library(foreign)
        library(lmtest)
        library(sandwich)
        library(multiwayvcov)
        library(data.table)
        library(stargazer)
```

```
Warning message:
: package 'lmtest' was built under R version 3.3.2Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric

Warning message:
: package 'sandwich' was built under R version 3.3.2Warning message:
: package 'multiwayvcov' was built under R version 3.3.3Warning message:
: package 'stargazer' was built under R version 3.3.2
Please cite as:

 Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
 R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

2   1 Replicate Results

Skim Broockman and Green's paper on the effects of Facebook ads and download an anonymized version of the data for Facebook users only.

```
In [2]: d <- read.csv("./data/broockman_green_anon_pooled_fb_users_only.csv")
        head(d)
```

studyno	treat_ad	cluster	name_recall	positive_impression
2	0	Study 2, Cluster Number 1	0	0
2	0	Study 2, Cluster Number 2	1	0
2	0	Study 2, Cluster Number 3	0	0
2	0	Study 2, Cluster Number 4	1	0
2	1	Study 2, Cluster Number 7	1	1
2	1	Study 2, Cluster Number 7	0	0

```
In [3]: nrow(d)
```

2706

a. Using regression without clustered standard errors (that is, ignoring the clustered assignment), compute a confidence interval for the effect of the ad on candidate name recognition in Study 1 only (the dependent variable is "name_recall").

- **Note**: Ignore the blocking the article mentions throughout this problem.
- **Note**: You will estimate something different than is reported in the study.

```
In [4]: study.1 <- d[d$studyno ==1,]
        study.2 <- d[d$studyno ==2,]

        head(study.1)
```

	studyno	treat_ad	cluster	name_recall	positive_impression
1343	1	1	Study 1, Cluster Number 1	0	0
1344	1	1	Study 1, Cluster Number 1	0	0
1345	1	1	Study 1, Cluster Number 3	0	0
1346	1	1	Study 1, Cluster Number 4	0	0
1347	1	1	Study 1, Cluster Number 5	0	0
1348	1	1	Study 1, Cluster Number 9	0	0

```
In [5]: m1 <- lm( name_recall ~ treat_ad , data = study.1)
        coeftest(m1)
```

```
t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.1824687  0.0161422 11.3038   <2e-16 ***
treat_ad    -0.0097979  0.0210122 -0.4663   0.6411
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Calculating confidence intervals

```
In [6]: confint(m1,'treat_ad')
```

	2.5 %	97.5 %
treat_ad	-0.05101765	0.03142188

    b. What are the clusters in Broockman and Green's study? Why might taking clustering into account increase the standard errors?

In the study, clusters are defined based on age, town and gender. Clustering allows for correation between observations, that is it accounts for phenomena that affect groups of observations uniformly (within cluster) but differently across clusters (variance isn't constant cross section). Robust standard errors correct for this issue

    c. Now repeat part (a), but taking clustering into account. That is, compute a confidence interval for the effect of the ad on candidate name recognition in Study 1, but now correctly accounting for the clustered nature of the treatment assignment. If you're not familiar with how to calculate these clustered and robust estimates, there is a demo worksheet that is available in our course repository: `./code/week5clusterAndRobust.Rmd`.

```
In [7]: coeftest(m1,cluster.vcov(m1,~ cluster))


t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.1824687  0.0184915  9.8677    <2e-16 ***
treat_ad    -0.0097979  0.0237536 -0.4125    0.6801
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can derive the confidence intervals from the cluster robust standard errors. For a 95% confidence interval we use the product of 1.96 and the cluster-robust std. error.

```
In [8]: -0.0097979 + (1.96 * 0.0237536)
```

    0.036759156

```
In [9]: -0.0097979 - (1.96 * 0.0237536)
```

    -0.056354956
    Cluster-robust 95% confidence interval is (-0.056354956,0.036759156)

    d. Repeat part (c), but now for Study 2 only.

```
In [10]: m2 <- lm( name_recall ~ treat_ad , data = study.2)
         coeftest(m2)
```

```
t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.6057884  0.0154543 39.1987   <2e-16 ***
treat_ad    -0.0028033  0.0308740 -0.0908   0.9277
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In [11]: coeftest(m2,cluster.vcov(m2,~ cluster))

```
t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.6057884  0.0181889  33.305   <2e-16 ***
treat_ad    -0.0028033  0.0355033  -0.079   0.9371
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can derive the confidence intervals from the cluster robust standard errors. For a 95% confidence interval we use the product of 1.96 and the cluster-robust std. error.

In [12]: -0.0028033-(1.96*0.0355033)

   -0.072389768

In [13]: -0.0028033+(1.96*0.0355033)

   0.066783168
   Cluster-robust 95% confidence interval is (-0.072389768,0.066783168)

e. Repeat part (c), but using the entire sample from both studies. Do not take into account which study the data is from (more on this in a moment), but just pool the data and run one omnibus regression. What is the treatment effect estimate and associated p-value?

In [14]: m3 <- lm(name_recall ~ treat_ad , data = d)
         coeftest(m3)

```
t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.454196   0.012189 37.2622 < 2.2e-16 ***
treat_ad    -0.155073   0.018762 -8.2652 < 2.2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Treatment effect is estimated to be -0.155 and seems to be very significant p-value (2.2e-16)

f. Now, repeat part (e) but include a dummy variable (a 0/1 binary variable) for whether the data are from Study 1 or Study 2. What is the treatment effect estimate and associated p-value?

```
In [15]: d.1f <- cbind(d)
         d.1f$dummy.study <- (d.1f$study-1)
         head(d.1f)
```

studyno	treat_ad	cluster	name_recall	positive_impression	dummy.study
2	0	Study 2, Cluster Number 1	0	0	1
2	0	Study 2, Cluster Number 2	1	0	1
2	0	Study 2, Cluster Number 3	0	0	1
2	0	Study 2, Cluster Number 4	1	0	1
2	1	Study 2, Cluster Number 7	1	1	1
2	1	Study 2, Cluster Number 7	0	0	1

```
In [16]: m4 <- lm( name_recall ~ treat_ad + dummy.study, data = d.1f)
         coeftest(m4)
```

```
t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.1806848  0.0159938 11.2971    <2e-16 ***
treat_ad    -0.0067752  0.0181767 -0.3727    0.7094
dummy.study  0.4260988  0.0179555 23.7309    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

g. Why did the results from parts (e) and (f) differ? Which result is biased, and why? (Hint: see pages 75-76 of Gerber and Green, with more detailed discussion optionally available on pages 116-121.)

If we think about the dummy variable as being a 'block' indicator, we can observe that the probability of being assigned to a treatment group is much higher in study 1 (dummy variable = 0). This means that if blocks happen to have different potential outcomes on average (even of the absense of a true treatment effect) regressing across blocks might signal an average treatment effect (effectely the difference across blocks is being captured by the treatment variable on such a simple model).

As a result, the first regression (pooling the data across studies) is biased.

h. Skim this Facebook case study and consider two claims they make reprinted below. Why might their results differ from Broockman and Green's? Please be specific and provide examples.

- "There was a 19 percent difference in the way people voted in areas where Facebook Ads ran versus areas where the ads did not run."

Answer

Effectively, there was no randomization across what Brookckman and Greens would have defined as clusters (counties). This study instead chose to assign treatments only to residents of the two most populous counties. This claim on 19% vote difference fails to take into consideration inherent differences across counties in Florida to begin with. Perhaps less densely populated counties had a tendency to vote in favor (on average) of the proposition. That would attribute the findings to simply biased estimates due to covariates not controlled for due to the lack of proper randomization.

- "In the areas where the ads ran, people with the most online ad exposure were 17 percent more likely to vote against the proposition than those with the least."

Answer

This claim is tied to a post-treatment covariate (relative exposure to ad) which could be in turn correlated to inherit differences between the treated populations. For example, they explain that the different treatments (ads) *targeted people who listed terms like "teacher," "pta" "math teacher" to reach educators. Because both the polling and Facebook research indicated that the issue carries special resonance with parents of school children, it even included interests like "I love my son" and "I love my daughter" (and layered them with demographic targeting.)*

This is problematic as perhaps the different demographics could have influenced both the success on target campaings reaching them (inherit differences in using facebook for example) as well as their potential outcomes despite any treatment. These two examples would render the article's claim biased.

3   2 Peruvian Recycling

Look at this article about encouraging recycling in Peru. The paper contains two experiments, a "participation study" and a "participation intensity study." In this problem, we will focus on the latter study, whose results are contained in Table 4 in this problem. You will need to read the relevant section of the paper (starting on page 20 of the manuscript) in order to understand the experimental design and variables. (*Note that "indicator variable" is a synonym for "dummy variable," in case you haven't seen this language before.*)

a. In Column 3 of Table 4A, what is the estimated ATE of providing a recycling bin on the average weight of recyclables turned in per household per week, during the six-week treatment period? Provide a 95% confidence interval.

`In [17]:` `0.024-1.96*0.039`

-0.05244
ATE is 0.187 Kg, with (0.12428,0.24972) being the confidence interval at 95%

b. In Column 3 of Table 4A, what is the estimated ATE of sending a text message reminder on the average weight of recyclables turned in per household per week? Provide a 95% confidence interval.

-0.024 kg, with (-0.05244,0.10044) confidence interval at 95%

c. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of providing a recycling bin?

- Percentage of visits turned in a bag
- Avg. no. of bins turned in per week
- Avg. weight (kg) of recyclables turned in per week
- Avg. market value of recyclables given per week

d. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of sending text messages?

No outcome measure show statistically significant effects of sending text messages.

e. Suppose that, during the two weeks before treatment, household A turns in 2kg per week more recyclables than household B does, and suppose that both households are otherwise identical (including being in the same treatment group). From the model, how much more recycling do we predict household A to have than household B, per week, during the six weeks of treatment? Provide only a point estimate, as the confidence interval would be a bit complicated. This question is designed to test your understanding of slope coefficients in regression.

We predict household A to turn in 2.562kg more recyclables than houshold B. (Coefficient on baseline wgt. is 0.281).

f. Suppose that the variable "percentage of visits turned in bag, baseline" had been left out of the regression reported in Column 1. What would you expect to happen to the results on providing a recycling bin? Would you expect an increase or decrease in the estimated ATE? Would you expect an increase or decrease in the standard error? Explain your reasoning.

We would expect an increase in the treatment effect as variation that would be otherwise explained by the baseline variable would transfer into the error term. Similarly, we would expect the ATE to also increase as more pronounced effects will be observed in the treatment groups without being absorved by the baseline covariate.

g. In column 1 of Table 4A, would you say the variable "has cell phone" is a bad control? Explain your reasoning.

I would say it is not a bad covariate. There is of course correlation between the eligibility for SMS treatment and ownership of a cell phone but the only wait to properly measure the effects of SMS treatment is by controling for 'Has a cell phone' == 1. That's not to say that the design around the SMS experiment is flawless, however given the treatment design, the covariate proves useful.

h. If we were to remove the "has cell phone" variable from the regression, what would you expect to happen to the coefficient on "Any SMS message"? Would it go up or down? Explain your reasoning.

It would go up. In reality the 'Any SMS message' would be capturing the interaction effect between cell phone ownership and being subject to the SMS treatment, so inevitably the covariate will be picking up some of the variability attributed to cell phone ownership.

4   3 Multifactor Experiments

Staying with the same experiment, now lets think about multifactor experiments.

a. What is the full experimental design for this experiment? Tell us the dimensions, such as 2x2x3. (Hint: the full results appear in Panel 4B.)

This experiment is a 3x3 factorial design capturing the effects of two potential factors and all of its variations.

1) Providing a bin (two potential treatments, plus one control)
2) Providing a bin (two potential treatments, plus one control)

One important component is the blocking done on cell phone ownership covariate due to its interaction with one of the treatments (SMS)

b. In the results of Table 4B, describe the baseline category. That is, in English, how would you describe the attributes of the group of people for whom all dummy variables are equal to zero?

The baseline category refers to the expected potential outcomes observed in the control group, and the variability on recycling measures that can be attributed to non-treatment related covariates.

c. In column (1) of Table 4B, interpret the magnitude of the coefficient on "bin without sticker." What does it mean?

It means that the standalone effect of providing a recycling bin to households (without stickers) is an increase of 3.5% on the number of visits in which the household had the opportunity to turn in a bag for reclycling, and actually did so.

d. In column (1) of Table 4B, which seems to have a stronger treatment effect, the recycling bin with message sticker, or the recycling bin without sticker? How large is the magnitude of the estimated difference?

'Bin with a sticker' seems to have a stronger treatment effect (2%). In terms of confidence intervals for each estimate, the magnitude of the difference ranges between (-3.88,7.88%) at the 95% confidence level.
Further notes
Confidence interval of 'Bin with sticker' is 5.5% +/- 2.94%, or (2.56%,8.44%)
Confidence interval of 'Bin without sticker' is 3.5% +/- 2.94%, or (0.56%,6.44%)

e. Is this difference you just described statistically significant? Explain which piece of information in the table allows you to answer this question.

It is not statistically significant. The F-test p-value (0.31) helps address that question

f. Notice that Table 4C is described as results from "fully saturated" models. What does this mean? Looking at the list of variables in the table, explain in what sense the model is "saturated."

Table 4C captures treatment effects across all combinations of factorial treatments in addition to the blocking variable (has phone) and the other covariates. The model is saturated as all potential combinations of the treatments are captured, taking into account the variability attributed to covariates and the inherit differences across blocks

5   4 Now! Do it with data

Download the data set for the recycling study in the previous problem, obtained from the authors. We'll be focusing on the outcome variable Y="number of bins turned in per week" (avg_bins_treat).

```
In [18]: d4 <- read.dta("./data/karlan_data_subset_for_class.dta")
         head(d4)
```

street	havecell	avg_bins_treat	base_avg_bins_treat	bin	sms	bin_s	bin_g	sms_p	sms_g
7	1	1.0416666	0.750	1	1	1	0	0	1
7	1	0.0000000	0.000	0	1	0	0	1	0
7	1	0.7500000	0.500	0	0	0	0	0	0
7	1	0.5416667	0.500	0	0	0	0	0	0
6	1	0.9583333	0.375	1	0	0	1	0	0
8	0	0.2083333	0.000	1	0	0	1	0	0

Do some quick exploratory data analysis with this data. There are some values in this data that seem a bit strange. Determine what these are, and figure out what you would like to do with them. Also, notice what happens with your estimates vis-a-vis the estimates that are produced by the authors when you do something sensible with this strange values.

a. For simplicity, let's start by measuring the effect of providing a recycling bin, ignoring the SMS message treatment (and ignoring whether there was a sticker on the bin or not). Run a regression of Y on only the bin treatment dummy, so you estimate a simple difference in means. Provide a 95% confidence interval for the treatment effect.

```
In [19]: m4.1 <- lm( avg_bins_treat ~ bin , data = d4)
         coeftest(m4.1)


t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 0.635350   0.011793 53.8740 < 2.2e-16 ***
bin         0.135380   0.020291  6.6721 3.356e-11 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In [20]: confint(m4.1,'bin')

	2.5 %	97.5 %
bin	0.09558421	0.1751758

    b. Now add the pre-treatment value of Y as a covariate. Provide a 95% confidence interval for the treatment effect. Explain how and why this confidence interval differs from the previous one.

In [21]: m4.2 <- lm(avg_bins_treat ~ bin + base_avg_bins_treat , data = d4)
        coeftest(m4.2)

```
t test of coefficients:

                      Estimate Std. Error t value  Pr(>|t|)
(Intercept)           0.349603   0.013731 25.4604 < 2.2e-16 ***
bin                   0.124693   0.016667  7.4814 1.149e-13 ***
base_avg_bins_treat   0.392965   0.013386 29.3560 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In [22]: confint(m4.2,'bin')

	2.5 %	97.5 %
bin	0.09200378	0.1573822

The 95% confidence interval for the treatment variable is narrower once the pre-treatment covariate is introduced. This is due to the fact that some of the variability has now moved from the treatment effect to the pre-treatment covariate.

    c. Now add the street fixed effects. (You'll need to use the R command factor().) Provide a 95% confidence interval for the treatment effect.

In [23]: m4.3 <- lm(avg_bins_treat ~ bin + base_avg_bins_treat + factor(street), data = d4)
        confint(m4.3,'bin')

	2.5 %	97.5 %
bin	0.08042873	0.1473448

    d. Recall that the authors described their experiment as "stratified at the street level," which is a synonym for blocking by street. Explain why the confidence interval with fixed effects does not differ much from the previous one.

Due to the large number of blocks (~240) the stratification fails to capture any meaningful heterogeneity in the different sub-groups, or at least it doesn't do a materially better job than ignoring the differences across streets. The reason why the confidence interval does not differ much is because the prior model (ignoring blocks) didn't remain unbiased due to the lage variability on ATE and and low density of the blocks.

e. Perhaps having a cell phone helps explain the level of recycling behavior. Instead of "has cell phone," we find it easier to interpret the coefficient if we define the variable " no cell phone." Give the R command to define this new variable, which equals one minus the "has cell phone" variable in the authors' data set. Use "no cell phone" instead of "has cell phone" in subsequent regressions with this dataset.

```
In [24]: d.4e <- cbind(d4)
         d.4e$no.cell.phone <- (1- d.4e$havecell)
         head(d.4e)
```

street	havecell	avg_bins_treat	base_avg_bins_treat	bin	sms	bin_s	bin_g	sms_p	sms_g	no.
7	1	1.0416666	0.750	1	1	1	0	0	1	0
7	1	0.0000000	0.000	0	1	0	0	1	0	0
7	1	0.7500000	0.500	0	0	0	0	0	0	0
7	1	0.5416667	0.500	0	0	0	0	0	0	0
6	1	0.9583333	0.375	1	0	0	1	0	0	0
8	0	0.2083333	0.000	1	0	0	1	0	0	1

f. Now add "no cell phone" as a covariate to the previous regression. Provide a 95% confidence interval for the treatment effect. Explain why this confidence interval does not differ much from the previous one.

```
In [25]: m4.4 <- lm( avg_bins_treat ~ bin + base_avg_bins_treat + factor(street) + no.cell.phone
```

Cell phone ownership doesn't seem bear any effect on the treatment coefficient. One hypothesis is that some of the cues captured by cell phone ownership might already be absorbed by the street (given it could be a proxy to demographic information correlated to cell phone ownership) or pre-treatment covariates.

```
In [26]: confint(m4.4,'bin')
```

	2.5 %	97.5 %
bin	0.08166792	0.1485336

However the confidence interval does become wider as a result, which could be a result of the correlation between cell phone ownership and street (block) coefficients

g. Now let's add in the SMS treatment. Re-run the previous regression with "any SMS" included. You should get the same results as in Table 4A. Provide a 95% confidence interval for the treatment effect of the recycling bin. Explain why this confidence interval does not differ much from the previous one.

```
In [27]: m4.5 <- lm( avg_bins_treat ~ bin + base_avg_bins_treat + sms + no.cell.phone + factor(s
         summary(m4.5)$coefficients[2:5,]
```

	Estimate	Std. Error	t value	Pr(>|t|)
bin	0.115053649	0.01705105	6.747600	2.095786e-11
base_avg_bins_treat	0.373482860	0.01430947	26.100390	2.285610e-125
sms	0.005124375	0.02085563	0.245707	8.059407e-01
no.cell.phone	-0.046702054	0.02037507	-2.292118	2.202841e-02

In [28]: `confint(m4.5,'bin')`

	2.5 %	97.5 %
bin	0.08160886	0.1484984

Given that the treatments (sms and bin) were assigned randomly in an independent fashion, there is no correlation between the two as the presence of SMS should be balanced across bin treatments and control groups. As a result, there is virtually no change on the standard errors associated with the treatment (bin) in consideration

h. Now reproduce the results of column 2 in Table 4B, estimating separate treatment effects for the two types of SMS treatments and the two types of recycling-bin treatments. Provide a 95% confidence interval for the effect of the unadorned recycling bin. Explain how your answer differs from that in part (g), and explain why you think it differs.

In [29]: `m4.6 <- lm( avg_bins_treat ~ bin_s + bin_g +sms_p + sms_g + no.cell.phone +base_avg_bin`
`        summary(m4.6)$coefficients[2:7,]`

	Estimate	Std. Error	t value	Pr(>|t|)
bin_s	0.127812892	0.02223628	5.7479441	1.080308e-08
bin_g	0.103190216	0.02188886	4.7142810	2.636923e-06
sms_p	-0.008041152	0.02503693	-0.3211716	7.481224e-01
sms_g	0.019707117	0.02519764	0.7821016	4.342709e-01
no.cell.phone	-0.046383459	0.02037843	-2.2761053	2.297176e-02
base_avg_bins_treat	0.373852178	0.01431366	26.1185536	1.733949e-125

In [30]: `confint(m4.6,'bin_g')`

	2.5 %	97.5 %
bin_g	0.06025627	0.1461242

The effect of providing an unadorned recycling bin on the average no. of bins turned in per week is 0.103 with a 95% confidence interval of (0.060,0.146). The answer in this case differs from question g. as for the prior model the treatment variable used captured the combined effect of providing a bin (whether unadorned or adorned). By isolating the effect of only the unadorned bin we observe a less significant, more variable average treatment on recycling.

Given that the unadorned bin addresses one of the two major problems cited on the experiment design (inconvenience) and not both (the other being lack of knowledge - which the sticker addresses) the lower treatment effect seems reasonable. The increased variability could also be attributed to the same phenomena.

6   5 A Final Practice Problem

Now for a fictional scenario. An emergency two-week randomized controlled trial of the experimental drug ZMapp is conducted to treat Ebola. (The control represents the usual standard of

care for patients identified with Ebola, while the treatment is the usual standard of care plus the drug.)

Here are the (fake) data.

d <- read.csv("./data/ebola_rct2.csv") head(d)

```
In [31]: d5 <- read.csv("./data/ebola_rct2.csv")
         head(d5)
```

temperature_day0	vomiting_day0	treat_zmapp	temperature_day14	vomiting_day14	male
99.53168	1	0	98.62634	1	0
97.37372	0	0	98.03251	1	0
97.00747	0	1	97.93340	0	1
99.74761	1	0	98.40457	1	0
99.57559	1	1	99.31678	1	0
98.28889	1	1	99.82623	1	1

You are asked to analyze it. Patients' temperature and whether they are vomiting is recorded on day 0 of the experiment, then ZMapp is administered to patients in the treatment group on day 1. Vomiting and temperature is again recorded on day 14.

a. Without using any covariates, answer this question with regression: What is the estimated effect of ZMapp (with standard error in parentheses) on whether someone was vomiting on day 14? What is the p-value associated with this estimate?

```
In [33]: m5 <- lm(vomiting_day14 ~ treat_zmapp , data = d5)
         summary(m5)


Call:
lm(formula = vomiting_day14 ~ treat_zmapp, data = d5)

Residuals:
    Min      1Q   Median      3Q      Max
-0.84746 -0.03803  0.15254  0.21197  0.39024

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.84746    0.05483  15.456   <2e-16 ***
treat_zmapp -0.23770    0.08563  -2.776   0.0066 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4212 on 98 degrees of freedom
Multiple R-squared:  0.0729,	Adjusted R-squared:  0.06343
F-statistic: 7.705 on 1 and 98 DF,  p-value: 0.006595
```

Answer

Estimated effect of Zmapp is ~ 23% reduction (standard error of 8.5%) on the likelihood someone would be vomitting after 14 days of the treatment. The p-value of this coefficient is 0.0066

13

b. Add covariates for vomiting on day 0 and patient temperature on day 0 to the regression from part (a) and report the ATE (with standard error). Also report the p-value.

```
In [38]: m5.1 <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 , data = d5)
         summary(m5.1)


Call:
lm(formula = vomiting_day14 ~ treat_zmapp + vomiting_day0, data = d5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.96208 -0.06578  0.03792  0.22530  0.62310

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.56428    0.07562   7.462 3.67e-11 ***
treat_zmapp   -0.18738    0.07762  -2.414   0.0177 *
vomiting_day0  0.39780    0.08059   4.936 3.31e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3785 on 97 degrees of freedom
Multiple R-squared:  0.259,Adjusted R-squared:  0.2437
F-statistic: 16.95 on 2 and 97 DF,  p-value: 4.85e-07
```

Answer
Estimated effect of Zmapp is ~ 18.7% reduction (standard error of 7.7%) on the likelihood someone would be vomitting after 14 days of the treatment. The p-value of this coefficient is 0.0177

c. Do you prefer the estimate of the ATE reported in part (a) or part (b)? Why?

Answer
My preference would be the estimate produced on part (b) due to the fact it gives a more accurate picture on the treatment effect than part (a). In essence, adding vomiting_day0 effectively helps blocking the population by a pre-treatment covariate, which helps us better assess the true effect of the treatment.

d. The regression from part (b) suggests that temperature is highly predictive of vomiting. Also include temperature on day 14 as a covariate in the regression from part (b) and report the ATE, the standard error, and the p-value.

```
In [37]: m5.2 <- lm(vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day14, data = d5)
         summary(m5.2)
```

```
Call:
lm(formula = vomiting_day14 ~ treat_zmapp + vomiting_day0 + temperature_day14,
    data = d5)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0323 -0.1978  0.1093  0.2254  0.6208

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -6.61356    2.91521  -2.269 0.025530 *
treat_zmapp       -0.12876    0.07933  -1.623 0.107835
vomiting_day0      0.31878    0.08486   3.756 0.000296 ***
temperature_day14  0.07269    0.02951   2.463 0.015560 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3689 on 96 degrees of freedom
Multiple R-squared:  0.3031,Adjusted R-squared:  0.2813
F-statistic: 13.91 on 3 and 96 DF,  p-value: 1.32e-07
```

Answer

Estimated effect of Zmapp is ~ 12.7% reduction (standard error of 7.9%) on the likelihood someone would be vomitting after 14 days of the treatment. The p-value of this coefficient is 0.10783

e. Do you prefer the estimate of the ATE reported in part (b) or part (d)? Why?

Answer

My preference would be the estimate produced on part (b). Given temperature_day14 is essentially a post-treatment variable and the fact it could be correlated with the outcome will essentially produce biased results.

f. Now let's switch from the outcome of vomiting to the outcome of temperature, and use the same regression covariates as in part (b). Test the hypothesis that ZMapp is especially likely to reduce men's temperatures, as compared to women's, and describe how you did so. What do the results suggest?

First step would be to create an additional feature that helps us track the difference in temperature between days 0 and 14, as well as another one to account for any interaction between gender and treatment.

To generate the interaction feature we can simply leverage the existing dummy variables (treatment and male). By using their product, we can interpret value == 1 as being a male and receiving the treatment, and values = 0 otherwise (i.e. not being a male subject to the treatment).

```
In [44]: d5.f <- cbind(d5)
         d5.f$temperature_diff <- (d5.f$temperature_day14 - d5.f$temperature_day0)
         d5.f$gender_interaction <- (d5.f$male * d5.f$treat_zmapp)

         head(d5.f)
```

temperature_day0	vomiting_day0	treat_zmapp	temperature_day14	vomiting_day14	male	temp...
99.53168	1	0	98.62634	1	0	-0.905
97.37372	0	0	98.03251	1	0	0.658?
97.00747	0	1	97.93340	0	1	0.925?
99.74761	1	0	98.40457	1	0	-1.343
99.57559	1	1	99.31678	1	0	-0.258
98.28889	1	1	99.82623	1	1	1.537?

To measure the effect on temperature changes we can use the newly created feature as our target variable (temperature_diff) and regress on treatment and block on gender and gender interaction.

```
In [52]: m5.3 <- lm(temperature_diff ~ treat_zmapp + vomiting_day0 + male + gender_interaction ,
         summary(m5.3)


Call:
lm(formula = temperature_diff ~ treat_zmapp + vomiting_day0 +
    male + gender_interaction, data = d5.f)

Residuals:
     Min       1Q   Median       3Q      Max
-1.04361 -0.34325 -0.07118  0.34953  1.08533

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.4546     0.1092   4.162 6.93e-05 ***
treat_zmapp         -0.2327     0.1340  -1.736   0.0858 .
vomiting_day0       -0.7626     0.1091  -6.989 3.80e-10 ***
male                 3.0647     0.1427  21.477  < 2e-16 ***
gender_interaction  -1.9474     0.2146  -9.076 1.56e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5102 on 95 degrees of freedom
Multiple R-squared:  0.8572, Adjusted R-squared:  0.8512
F-statistic: 142.6 on 4 and 95 DF,  p-value: < 2.2e-16
```

The above estimate suggests that there treatment has a more pronounced effect for men. Overall, the -1.94 coefficient suggests that the temperature drop is 1.94 higher in men when the treatment is applied. The p-value for this coefficient is also highly significant.

g. Suppose that you had not run the regression in part (f). Instead, you speak with a colleague to learn about heterogenous treatment effects. This colleague has access to a non-anonymized version of the same dataset and reports that he had looked at heterogenous effects of the ZMapp treatment by each of 10,000 different covariates to examine whether each predicted the effectiveness of ZMapp on each of 2,000 different indicators of health, for 20,000,000 different regressions in total. Across these 20,000,000 regressions your colleague ran, the treatment's interaction with gender on the outcome of temperature is the only heterogenous treatment effect that he found to be statistically significant. He reasons that this shows the importance of gender for understanding the effectiveness of the drug, because nothing else seemed to indicate why it worked. Bolstering his confidence, after looking at the data, he also returned to his medical textbooks and built a theory about why ZMapp interacts with processes only present in men to cure. Another doctor, unfamiliar with the data, hears his theory and finds it plausible. How likely do you think it is ZMapp works especially well for curing Ebola in men, and why? (This question is conceptual can be answered without performing any computation.)

I would not trust the rationale employed by my colleague as there was no test on whether the estimated interaction effect could have occurred by chance. One approach could have been to apply randomize inference to derive p-values analytically.

Perhaps more importantly, there was no prior hypothesis formulated prior building the model, rather my colleague just found whatever signal the data suggested and did selective research to find literature that align to his case, that would in my opinion render the analysis invalid.

h. Now, imagine that what described in part (g) did not happen, but that you had tested this heterogeneous treatment effect, and only this heterogeneous treatment effect, of your own accord. Would you be more or less inclined to believe that the heterogeneous treatment effect really exists? Why?

I would be more inclined to believe there was heterogeneous effect given the hypothesis was advanced a priori.

i. Another colleague proposes that being of African descent causes one to be more likely to get Ebola. He asks you what ideal experiment would answer this question. What would you tell him? (*Hint: refer to Chapter 1 of Mostly Harmless Econometrics.*)

The ideal experiment proposed would be take into account the additional factors that could be correlated with the proposed covariate (i.e. being of African descent). Apart from randomization of course, you would want to block instead by contextual variables that are as little correlated as possible to being of african descent. For example, using income, level of education, assessment on the conditions of public health services on subject's area of residence would prove helpful to really isolate the true effects of confounding factors.