Problem Set 3

Alex, Micah, Scott, and David

12/01/2021

Contents

1 Peruvian Recycling

1.1 Recycling bin ATE

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. Provide a short narrative using inline R code, such as r inline_reference.

```
# Use this code chunk to show your code work (if needed)
```

Answer: The ATE of providing a recycling bin on the average weight of recyclables turned in per house per week, during the six-week treatment period is 0.187kg with a standard error of 0.032kg. The 95% confidence interval is between 0.123kg to 0.251kg. ' This means that on average 0.187kg more material is recycled by households receiving a bin. Additionally, we are 95% confidence that receiving a bin increases the average material in the range of 0.123kg to 0.251kg. The result is statistically significant

1.2 SMS ATE

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 and provide a short narrative using inline R code.

```
# Use this code chunk to show your code work (if needed)
```

Answer: the estimated ATE of sending a text message reminder on the average weight of recyclables turned in per household per week is -0.024kg. The confidence interval is [-0.102, 0.054]kg. This means that 0.024kg less material is recycled by households that receive a text message. Additionally, we are 95% confident that the average amount of materials recycled for households that receive text message lies between -0.102kg to 0.054kg. The result is statistically insignificant.

1.3 What outcomes does a recycling bin affect?

Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of providing a recycling bin? How are you dealing with the issue that there are several different tests that have been run, and that you are reading? How, if at all, do the authors deal with this?

Answer: The following outcome measures are statistically significant: - Percentage of visits turned in bag - Average number of bins turned in per week - Average weight(kg) turned in per week - Average market value per week

In a situation like this, I would avoid overstating statistical significance by Bonferroni correction. I don't believe the authors are accounting for this. However, since we have multiple statistically significant p-values in 4/5 hypothesis, it is likely that we are not overstating the statistical significance.

1.4 What outcomes does a SMS affect?

Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of sending text messages? Now that you have read across two different treatments, and many outcomes, what, if anything do the p-values mean to you? Does this feel like p-hacking or doing careful investigation?

Answer: No statistically significant effects for sending text messages. This seems like p-hacking to me because suppose we had another column and we got a statistically significant value there. If we reported that as though that was a legit result then we would have committed p-hacking.

1.5 Marginal effects

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.

```
# Use this code chunk to show your code work (if needed)
```

Answer: From Column 3 of table 4A, we can see that household baseline average weight of recyclables turned in per week is 0.281kg. Since household A has twice as much recycling compared to household B, the result would be (2 x 0.281 = 0.562)

1.6 Covariates or confounders?

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.

```
# Use this code chunk to show your code work (if needed)
```

Answer: If we leave the "percentage of visits turned in bag, baseline" from the regression, we would expect to see the estimated ATE increase on providing the recycling the bin because the effects of omitted variable will be added to the existing variables in the model. Some of the variance related to the baseline will show up in the "providing a recycle bin" variable and the standard error will increase as well.

1.7 Bad control or useful subset?

In column 1 of Table 4A, would you say the variable "has cell phone" is a bad control? Explain your reasoning, and engage both with the definition of a bad control, and also the implications of including a bad control in a model.

Answer: A bad control happens when the covariates are affected by the treatment. I would not assume that the "has cell phone" covariate is a bad control because there is no hint that the treatment affect subjects to obtain new phones are get rid of their phones.

1.8 What happens if you remove "has cell phone"?

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.

```
# Use this code chunk to show your code work (if needed)
```

Answer: "has cell phone" has a positive coefficient across all columns. If we remove it, the effect has to be observed by other variables. In our case, the coefficient of "Any SMS message" will likely increase absorbing the effects of "has cell phone."

2 Multifactor Experiments

2.1 Experiment design?

What is the full experimental design for this experiment? Tell us the dimensions, such as 2x2x3. The full results appear in Panel 4B. We'll note that the dimensions of an experiment are defined in terms of the *treatments that the experiment assigns*, not in terms of other features about the data.

Answer: First treatment includes a bin with sticker, a bin without a sticker, or no bin. Second treatment include having a personal message, having a generic message, or no message. Therefore, it is 3x3 dimensions.

2.2 Baseline for interpretation

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?

Answer: The group for whom all dummy variables are zero means that they received no bins, received no SMS message and have no cell phone

2.3 Bin without sticker effect

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

Answer: A subject who received a bin without a sticker on average had an increase of 0.035 in "percentage of visit turned in bags." It is statistically signficant and we are 95% confident that it ranges between 0.005 to 0.035 (0.5% to 3.5%)

2.4 With or without a sticker?

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?

Answer: The recycling bin with the sticker has the the larger treatment effect of 0.055. The magnitude of the estimated difference is (0.055-0.035 = 0.020). Both coefficients have the same SDE.

2.5 Statistical significantly different with or without a sticker?

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

Answer: The difference is not statistically significant because the value of F-test (0.31) is not significant. In other words, we cannot reject the null hypothesis that the effects of bin with stickers and bins without stickers is the same.

2.6 Fully saturated?

Notice that Table 4C is described as results from "fully saturated" models. What does this mean? What does David Reiley propose this definition means to him in the async lecture? What do the authors seem to think it means to them? Looking at the list of variables in the table, explain in what sense the model is "saturated."

Answer: Saturated means that we can account for all the combinatiosn in the treatment. This model is saturated because it includes all the independent variables and have coefficients for the 3 bins and 3 the SMS situations.

3 Now! Do it with data

Minimum street value of -999 does not seem to be accurate. Let's visualize this.

let's convert all streets that have values of -999 to NAs.

3.1 Treatment only model

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, using **of course** robust standard errors (use these throughout) and provide a brief narrative using R inline statements.

```
mod_basic <- d[, lm(avg_bins_treat ~ bin)]
mod_basic$vcovHC_ <- vcovHC(mod_basic)
coeftest(mod_basic, vcov. = mod_basic$vcovHC_)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  0.63515    0.01194 53.1954 < 2.2e-16 ***
## bin          0.13470    0.02133  6.3147 3.467e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Narrative: p-value is statistically significant therefore we can reject the null that providing a bin does not affect the average__bins__treated outcome. Those who were provided bins had an average of 0.13470 increase in average__bins__treated. We are 95% confident that this value ranges between 0.0929 to 0.1765

3.2 Treatment and pre-treatment values

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.

```
mod_pretreat <- d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat)]
mod_pretreat$vcovHC_ <- vcovHC(mod_pretreat)
coeftest(mod_pretreat, vcov. = mod_pretreat$vcovHC_)
```

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         0.354406   0.022059 16.0661 < 2.2e-16 ***
## bin                 0.127106   0.017745  7.1628 1.184e-12 ***
## base_avg_bins_treat 0.385347   0.031514 12.2279 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: The ATE of bin got reduced a bit because some of the effect got absorbed by the base__avg__bins__treat variable. Furthermore, the confidence interval became a bit narrower because the pre-treatment value of Y is predictive of Y which reduces noise. In other words, some variance was removed from the bin ATE coefficient because it was accounted by baseline base__avg__bins__treat variable.

3.3 Add state fixed effects

Now add the street fixed effects. (You'll need to use the R command `factor()` You can do this either within the `lm` call, or you can move this factoring up in the data pipeline so that it persists through the rest of your analysis. The only thing we would recommend that you *not* do is to engineer a new, persistent feature at this point.) Provide a 95% confidence interval for the treatment effect and provide a brief narrative using r inline statements.

Narrative: The ATE of bin has reduced a bit more because some of the effect has been absorbed by street level variables. Furthermore, the confidence interval has become narrower since a lot variance has been obtained by each street variable.

3.4 Test for block fixed effects

Recall that the authors described their experiment as "stratified at the street level," which is a synonym for blocking by street. Does including these block fixed effects change the standard errors of the estimates *very much*? Conduct the appropriate test for the inclusion of these block fixed effects, and interpret them in the context of the other variables in the regression.

Answer: Conducting blocking generally increases the precision with which the treatment effect is estimated given that the block help predict the outcome. In our case, it seems like blocking had very little impact on the ATE

3.5 Feature (no) cell phone

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.

```
d$no_cell_phone <- 1-d$havecell

head(d)
```

```
##    street havecell avg_bins_treat base_avg_bins_treat bin sms bin_s bin_g sms_p
## 1:      7        1      1.0416666               0.750   1   1     1     0     0
## 2:      7        1      0.0000000               0.000   0   1     0     0     1
## 3:      7        1      0.7500000               0.500   0   0     0     0     0
## 4:      7        1      0.5416667               0.500   0   0     0     0     0
## 5:      6        1      0.9583333               0.375   1   0     0     1     0
## 6:      8        0      0.2083333               0.000   1   0     0     1     0
##    sms_g no_cell_phone
## 1:     1             0
## 2:     0             0
## 3:     0             0
## 4:     0             0
```

```
## 5:     0             0
## 6:     0             1
```

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.

```r
mod_cellphone <- d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat + no_cell_phone + as.factor(street) )

mod_cellphone$vcovHC_ <- vcovHC(mod_cellphone)
#first rows values
coeftest(mod_cellphone, vcov. = mod_cellphone$vcovHC_)[1:7,]
```

```
##                          Estimate Std. Error     t value     Pr(>|t|)
## (Intercept)          0.2876250041 0.05441443  5.28582203 1.438810e-07
## bin                  0.1171694452 0.01963637  5.96696176 3.019939e-09
## base_avg_bins_treat  0.3668399995 0.03145357 11.66290469 3.928908e-30
## no_cell_phone       -0.0429672280 0.01864142 -2.30493374 2.130819e-02
## as.factor(street)3   0.0473169981 0.11456510  0.41301408 6.796562e-01
## as.factor(street)4  -0.0292088270 0.08113261 -0.36001342 7.188885e-01
## as.factor(street)5  -0.0009819708 0.11098914 -0.00884745 9.929420e-01
```

Answer:The confidence interval did not seem to change much at all. Because it does not appear like we are gaining anything of significance out of the treatment because the having cell phone or not does not seem to have any predictive power.

3.6 Add the sms treatment

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.

```r
mod_sms <- d[, lm(avg_bins_treat ~ bin + base_avg_bins_treat + no_cell_phone + sms +as.factor(street) )]

mod_sms$vcovHC_ <- vcovHC(mod_sms)
#first rows values
coeftest(mod_sms, vcov. = mod_sms$vcovHC_)[1:7,]
```

```
##                        Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)          0.27862128 0.05577292  4.9956374 6.561925e-07
## bin                  0.11696781 0.01964595  5.9537864 3.267811e-09
## base_avg_bins_treat  0.36708159 0.03141555 11.6847094 3.112828e-30
## no_cell_phone       -0.03342530 0.02398533 -1.3935730 1.636561e-01
## sms                  0.01740018 0.02454833  0.7088132 4.785522e-01
## as.factor(street)3   0.04562982 0.11508827  0.3964767 6.918106e-01
## as.factor(street)4  -0.03135202 0.08116736 -0.3862639 6.993568e-01
```

Answer: The confidence interval does not change since sms does not seem to explain any of the variance/noise

3.7 Reproduce Table 4B, Column (2)

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.

Answer: The confidence interval is has a narrower range. This makes sense because the extra precision comes from breaking bins into to bin_g and bin_s. Each explaining some of the variance.

Note mod_full won't knit when I try to get a pdf version so I am adding the code below as a proof that I can do it.

mod_full <- d[, lm(avg_bins_treat ~ bin_g + bin_s + sms_g + sms_p + base_avg_bins_treat + no_cell_phone+ as.factor(street))]

mod_full$vcovHC_ <- vcovHC(mod_full)

coeftest(mod_full, vcov. = mod_full$vcovHC_)[1:7,]

#confidence interval below

mod_full_ci <- coefci(mod_full, vcov. = vcovHC(mod_full))

mod_full_ci[1:7,]

4 A Final Practice Problem

4.1 Simple treatment effect of Zmapp

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 dehydrated on day 14? What is the p-value associated with this estimate?

```
zmapp_1 <- d[, lm(dehydrated_day14 ~ treat_zmapp)]
summary(zmapp_1)
```

```
##
## Call:
## lm(formula = dehydrated_day14 ~ treat_zmapp)
##
## 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:The ATE of ZMapp is -0.238 with an standard error of (0.08563). The p-value is 0.006595 which is less than 0.05. Therefore, it is statistically significant.

4.2 Add baseline covariates

Add covariates for dehydration 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.

```
zmapp_2 <- d[, lm(dehydrated_day14 ~ treat_zmapp + dehydrated_day0 + temperature_day0)]
summary(zmapp_2)
```

```
##
## Call:
## lm(formula = dehydrated_day14 ~ treat_zmapp + dehydrated_day0 +
##     temperature_day0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79643 -0.18106  0.04654  0.23122  0.68413
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -19.46966    7.44095  -2.617  0.01032 *
## treat_zmapp       -0.16554    0.07567  -2.188  0.03113 *
## dehydrated_day0    0.06456    0.14635   0.441  0.66013
## temperature_day0   0.20555    0.07634   2.693  0.00837 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3668 on 96 degrees of freedom
## Multiple R-squared:  0.311,  Adjusted R-squared:  0.2895
## F-statistic: 14.45 on 3 and 96 DF,  p-value: 7.684e-08
```

Answer:The ATE of ZMapp is -0.16554 with an standard error of (0.07567). The p-value is 0.006595 which is less than 7.684e-08. Therefore, it is statistically significant.

4.3 Interpret estimates

Do you prefer the estimate of the ATE reported in the chunk called `dehydration model` or `add pre-treatment measures`? Why? Report the results of the F-test that you used to form this opinion.

```
zmapp_test_object <- anova(zmapp_1, zmapp_2, test = "F")
zmapp_test_object
```

```
## Analysis of Variance Table
##
## Model 1: dehydrated_day14 ~ treat_zmapp
## Model 2: dehydrated_day14 ~ treat_zmapp + dehydrated_day0 + temperature_day0
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     98 17.383
## 2     96 12.918  2    4.4653 16.592 6.472e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: The second model with added pre-treatment measures is preferable. F-test also validates this to us and rejects the null hypothesis and adding the pre-treatment does not make our model better. This makes sense adding baseline temperature and hydration gives us a better estimate for hydration on day 14 clinically.

4.4 Add day fourteen temperature

The regression from part `add pre-treatment measures` suggests that temperature is highly predictive of dehydration. Add, temperature on day 14 as a covariate and report the ATE, the standard error, and the p-value.

```
zmapp_3 <- d[, lm(dehydrated_day14 ~ treat_zmapp + dehydrated_day0 + temperature_day0 + temperature_day
summary(zmapp_3)
```

```
##
## Call:
## lm(formula = dehydrated_day14 ~ treat_zmapp + dehydrated_day0 +
```

```
##     temperature_day0 + temperature_day14)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87745 -0.27436  0.04701  0.24801  0.66445
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -22.59159    7.47727  -3.021  0.00323 **
## treat_zmapp        -0.12010    0.07768  -1.546  0.12541
## dehydrated_day0     0.04604    0.14426   0.319  0.75033
## temperature_day0    0.17664    0.07642   2.312  0.02296 *
## temperature_day14   0.06015    0.02937   2.048  0.04335 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3609 on 95 degrees of freedom
## Multiple R-squared:  0.3402, Adjusted R-squared:  0.3124
## F-statistic: 12.24 on 4 and 95 DF,  p-value: 4.545e-08
```

Answer: The ATE of ZMapp is -0.12010 with an standard error of (0.07768). The p-value is not statistically significant.

4.5 Interpret estimates

Do you prefer the estimate of the ATE reported in part `add pre-treatment measures` or `add day 14 temperature`? What is this preference based on?

Answer:I prefer the estimate in part 'add pre-treatment measures' because 'add day 14 temperature' might be a bad control. And the reason that it is a bad control is that it is measured after treatment is given

4.6 Look at temperature

Now let's switch from the outcome of dehydration to the outcome of temperature, and use the same regression covariates as in the chunk titled `add pre-treatment measures`. Test the hypothesis that ZMapp is especially likely to reduce mens' temperatures, as compared to womens', and describe how you did so. What do the results suggest?

```
zmapp_4 <- d[, lm(temperature_day14 ~ treat_zmapp + dehydrated_day0 + temperature_day0 + male + treat_zm

stargazer(zmapp_4, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                          temperature_day14
## -----------------------------------------------
## treat_zmapp                   -0.231*
##                               (0.119)
##
```

```
## dehydrated_day0                0.041
##                               (0.182)
##
## temperature_day0              0.505***
##                               (0.095)
##
## male                          3.085***
##                               (0.126)
##
## treat_zmapp:male              -2.077***
##                               (0.192)
##
## Constant                      48.713***
##                               (9.266)
##
## -----------------------------------------------
## Observations                     100
## R2                              0.906
## Adjusted R2                     0.901
## Residual Std. Error        0.452 (df = 94)
## F Statistic             180.953*** (df = 5; 94)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

Answer: I created the model as described above. I also added a "male" variable and an indicator variable "male" with treat_zmapp. Therefore, the coefficient of treat_zmapp*male would give us the effects of being male and in treatment. As we can see, the estimated ATE for males who are in treatment is the highest in our model with a value of -2.077(0.192).

4.7 Compare health outcomes

Which group – those that are coded as `male == 0` or `male == 1` have better health outcomes (temperature) in control? What about in treatment? How does this help to contextualize whatever heterogeneous treatment effect you might have estimated?

```
zmapp_female <- d[male == 0, lm(temperature_day14 ~ treat_zmapp + dehydrated_day0 + temperature_day0)]
zmapp_male <- d[male == 1, lm(temperature_day14 ~ treat_zmapp + dehydrated_day0 + temperature_day0)]


stargazer(zmapp_female, zmapp_male, zmapp_4, type = "text")
```

```
##
## ==========================================================================================
##                                              Dependent variable:
##                      ---------------------------------------------------------------------
##                                               temperature_day14
##                               (1)                    (2)                    (3)
## ------------------------------------------------------------------------------------------
## treat_zmapp                 -0.229*              -2.239***               -0.231*
##                             (0.119)               (0.154)                (0.119)
##
## dehydrated_day0              0.265                -0.379                  0.041
```

```
##                                (0.227)               (0.297)               (0.182)
##
## temperature_day0               0.374***              0.767***              0.505***
##                                (0.116)               (0.161)               (0.095)
##
## male                                                                       3.085***
##                                                                            (0.126)
##
## treat_zmapp:male                                                          -2.077***
##                                                                            (0.192)
##
## Constant                      61.463***              26.232               48.713***
##                               (11.341)              (15.698)               (9.266)
##
## -----------------------------------------------------------------------------------------
## Observations                     63                    37                    100
## R2                              0.554                 0.918                 0.906
## Adjusted R2                     0.532                 0.910                 0.901
## Residual Std. Error    0.452 (df = 59)        0.440 (df = 33)       0.452 (df = 94)
## F Statistic         24.460*** (df = 3; 59) 122.877*** (df = 3; 33) 180.953*** (df = 5; 94)
## =========================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```

Answer: As we can see above, ATE for average treated females -0.229(0.119). ATE for treated males is -2.239(0.154). ATE for untreated males is 26.232(15.698)

4.8 Collaborating with others, Part (1)

Suppose you speak with a colleague to learn about heterogeneous treatment effects.

This colleague has access to a non-anonymized version of the same dataset and reports that they looked at heterogeneous effects of the ZMapp treatment by each of 80 different covariates to examine whether each predicted the effectiveness of ZMapp on each of 20 different indicators of health.

Across these regressions your colleague ran, the treatment's interaction with sex on the outcome of temperature is the only heterogeneous treatment effect that he found to be statistically significant. They reason that this shows the importance of sex for understanding the effectiveness of the drug, because nothing else seemed to indicate why it worked. Bolstering your colleague's confidence, after looking at the data, they also returned to their medical textbooks see the whispers of a theory about why ZMapp interacts with processes only present in men to cure.

Another doctor, unfamiliar with the data, hears your colleague's 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.)

Answer: This seems to be a fishing expedition. Since my colleague looked at 80 different covariates on 20 different indicators of health, he is bound to find a heterogeneous effect by chance.

4.9 Collaborating with others, Part (2)

Suppose that your colleague conducted their research looking at the interaction of 80 covariates with ZMapp, but that you on your own tested this and only this HTE, and discovered a positive result. How, if at all, does your colleague's behavior change the interpretation of your test? Does this seem fair or reasonable?

Answer: If this was the first test I performed, I would be confident in rejecting the null since the p-value is quite small. We would have over 95% confidence interval. It seems reasonable because I would not have gone fishing expedition to eventually be able to reject the null. However, if I knew that he had already gone fishing and found this result to be the only statistically result he found, then that would invalidate my result as well.

4.10 Collaborating with others, Part (3)

Now, imagine that your colleague had not conducted the 80 different regressions. Instead, they tested this heterogeneous treatment effect, and only this heterogeneous treatment effect, of their own accord. Would you be more or less inclined to believe that the heterogeneous treatment effect really exists? Why? Is there a general principle that is guiding your reasoning?

Answer:I would be more inclined to believe his result if he does not conduct the 80 different regressions because he would not be doing a fishing expedition. The general principle would be to run statistical tests with interaction terms given that we don't go fishing for HTE effect.