Problem Set 3

Code ▾

Hide

```
#install.packages("dplyr")
library(data.table)
library(sandwich)
library(lmtest)
library(ggplot2)
library(patchwork)
library(foreign)
library(dplyr)
```

1. Peruvian Recycling

Look at this article (./readings/recycling_peru.pdf) 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.*)

1. 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.

> The estimated ATE of providing a recycling bin is an increase between 0.123 and 0.251 kg per week

2. 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.

> The estimated ATE of sending a text message reminder is between -.102 kg and 0.054 kg per week

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

> Percentage of visits turned in bag, Avg. no. of bins turned in per week, Avg. weight (in kg) of recyclables turned in per week, and Avg. market value of recyclables turned in per week all show significant effects at the 5% level

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

> None of the outcome measures in TAble 4A show statistically significant effects at the 5% level of sending text messages.

5. 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.

> The model provides a baseline weight of recyclables in kg for all households that is equal to 0.281, but does not provide a relationship between baseline values and treatment outcomes. For this reason, two houses distinguishable only by the 2kg difference in baseline recycling weight would expect to differ by 2kg of recycling weight per week throughout the study period, assuming that they underwent the same treatment and were in fact otherwise identical.

6. 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 our reasoning.

> I would expect very small changes to the ATE as compared to the magnitude of the ATE, assuming that random selection was performed correctly. Given that both coefficients are positive, I would expect this change to be a slight increase to the ATE. I would expect an increase in the standard error, because the uncertainty associated with the baseline value, and accounted for in the baseline value that is reported, would instead be included in the uncertainty of the treatment, causing that value to increase.

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

> A bad control, generally speaking, is something that affects or is affected by the treatment. The variable 'has cell phone', although it always accompanies the SMS treatment, is not itself affecting the treatment or an outcome of the treatment itself.

8. 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.

We would not expect removing 'has cell phone' from the regression to have a large impact on the 'Any SMS coefficient', as 'has cell phone' is not affecting nor directly correlating with the treatment assignment. However, given that the coefficients for both 'has cell phone' and 'any SMS message' are both positive and not statistically significant, we might expect to see a slight increase in the 'Any SMS message' coefficient.

2. Multifactor Experiments

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

1. 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.

3x3 representing 3 bin choices (none, generic, sticker) and 3 SMS choices (none, generic, personal)

2. 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 is someone with no bin who did not receive any SMS messages.

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

The estimated ATE of someone receiving a bin without a sticker is 0.035% of visits turning in a bag greater compared to someone without no bin or text receipts at all.

4. 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?

The recycling bin with the sticker seems to have a stronger treatment effect. The magnitude of the estimated difference is 0.02% of visits turning in a bag.

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

The difference between bins with a sticker and bins without a sticker is not statistically significant. The standard error values show this because they are greater than half of the distance between these two point values.

6. 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."

The model is fully saturated because it includes indicators for each unique combination of treatments, thus accounting for all of the variance in the model by providing the term for each combination of covariates.

3. 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).

Hide

```
d <- foreign::read.dta("karlan_data_subset_for_class.dta")
d <- data.table(d)
d <-na.omit(d)
head(d)
```

street <dbl>	havecell <dbl>	avg_bins_treat <dbl>	base_avg_bins_treat <dbl>	bin <dbl>	s... <dbl>	bin_s <dbl>	bin_g <dbl>	sms_p <dbl>	sms_g <dbl>
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

6 rows

Hide

```
NA
```

1. 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).

Hide

```
mod_1 <- lm(d$avg_bins_treat ~ d$bin)
mod_1.vcovHC <- vcovHC(mod_1)
mod_1_robust_ci <- coefci(mod_1, vcov.=mod_1.vcovHC, level=0.95)
mod_1_robust_ci
```

```
                 2.5 %    97.5 %
(Intercept) 0.61338412 0.6583912
d$bin       0.09244614 0.1736952
```

2. 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.

Hide

```
mod_2 <- lm(d$avg_bins_treat ~ d$bin + d$base_avg_bins_treat)
mod_2.vcovHC <- vcovHC(mod_2)
mod_2_robust_ci <- coefci(mod_2, vcov.=mod_2.vcovHC, level=0.95)
mod_2_robust_ci
```

```
                           2.5 %    97.5 %
(Intercept)           0.31027926 0.3938329
d$bin                 0.09052669 0.1579526
d$base_avg_bins_treat 0.33011333 0.4498869
```

> This confidence interval differs from the previous by controlling for the potential HTE resulting from pre-treatment recycling volume.

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

Hide

```
mod_3 <- lm(d$avg_bins_treat ~ d$bin + d$base_avg_bins_treat + factor(d$street))
mod_3.vcovHC <- vcovHC(mod_3)
mod_3_robust_ci <- coefci(mod_3, vcov.=mod_3.vcovHC, level=0.95)
mod_3_robust_ci
```

```
                           2.5 %       97.5 %
(Intercept)            0.299099913  0.436527247
d$bin                  0.076608669  0.150809466
d$base_avg_bins_treat  0.314302811  0.433038422
factor(d$street)2     -0.210974037  0.021751354
factor(d$street)3     -0.253260855  0.163877777
factor(d$street)4     -0.263764473 -0.006356714
factor(d$street)5     -0.297876294  0.109438329
factor(d$street)6     -0.273637920  0.172181817
factor(d$street)7     -0.360198352  0.250851459
factor(d$street)8     -0.164701629  0.056641290
factor(d$street)9     -0.286441393  0.112725845
factor(d$street)10    -0.186833361  0.131739863
factor(d$street)11    -0.296741313  0.154855114
factor(d$street)15    -0.280793544  0.240847840
factor(d$street)17    -0.151493081  0.230721283
factor(d$street)20    -0.387461565 -0.084486861
factor(d$street)21    -0.250713510  0.098831832
factor(d$street)22    -0.256918855  0.006378983
factor(d$street)23    -0.267807886 -0.052558407
factor(d$street)26    -0.202549760  0.147136768
factor(d$street)32    -0.341229774  0.653568920
factor(d$street)37    -0.233019480  0.579203374
factor(d$street)38    -0.256028511  0.113830929
factor(d$street)40    -0.703066339  0.578570147
factor(d$street)41    -0.352069207  0.377601807
factor(d$street)42    -0.465707540  0.176133647
factor(d$street)43    -0.245761775  0.352079096
factor(d$street)44    -0.096023833  0.365453711
factor(d$street)45    -0.285061851  0.207285040
factor(d$street)46    -0.186641881  0.206994024
factor(d$street)47    -0.211856963  1.050823109
factor(d$street)53    -0.368038445  0.440543792
factor(d$street)58    -0.157924080  0.096957927
factor(d$street)60    -0.105788740  0.161854774
factor(d$street)61    -0.396078731  0.379250405
factor(d$street)62    -0.204385241  0.319430919
factor(d$street)63    -0.206704159  0.149760683
factor(d$street)64    -0.145353946  0.247317758
factor(d$street)66    -0.175723675  0.055245712
factor(d$street)67    -0.309096773  0.283138958
factor(d$street)68    -0.161840335  0.307259839
factor(d$street)69    -0.084686684  0.240939964
factor(d$street)70    -0.166448592  0.149459327
factor(d$street)72    -0.211750762  0.117176598
factor(d$street)73    -0.292594592 -0.049705394
factor(d$street)74    -0.201895211  0.071717984
factor(d$street)75    -0.159337286  0.046637145
factor(d$street)77    -0.378695617  0.312253582
factor(d$street)78    -0.209623969  0.106122483
factor(d$street)79    -0.721231771  1.334744047
factor(d$street)80    -0.086952014  0.110168662
factor(d$street)81    -0.298641734  0.236594699
factor(d$street)82    -0.395560883  0.461261475
factor(d$street)83    -0.365377428  0.127876726
factor(d$street)84    -0.150023286  0.114284208
factor(d$street)85    -0.081438340  0.211530743
factor(d$street)86    -0.212132734  0.078107917
factor(d$street)88    -0.307692601  0.112591511
factor(d$street)89    -0.210911598  0.203801353
factor(d$street)91    -0.561263546  0.139605480
factor(d$street)93    -0.244052090  0.103706620
factor(d$street)94    -0.201683391  0.127663407
factor(d$street)96    -0.376642828  0.298702411
factor(d$street)98    -0.116280535  0.221147180
factor(d$street)99     0.038819642  0.424204060
factor(d$street)100   -0.218161692  0.366883526
factor(d$street)101   -0.038282449  0.432916300
factor(d$street)102   -0.476572047  0.305983785
factor(d$street)103   -0.198206570  0.370556999
factor(d$street)105   -0.062759216  0.294874137
factor(d$street)106    0.139861437  0.660985696
factor(d$street)107   -0.311551773  0.440251859
factor(d$street)109   -0.256367994  0.227255931
factor(d$street)110   -0.293505260  0.575088440
factor(d$street)111   -0.341643341  0.344458649
factor(d$street)112    0.006401596  0.363211254
factor(d$street)113   -0.266710806  0.430444881
factor(d$street)115    0.026327118  0.558427293
factor(d$street)117   -0.581350581  1.696149259
factor(d$street)118   -0.691836302  0.757618776
factor(d$street)119   -0.239970103  0.218096390
factor(d$street)120   -0.273908831  0.138441360
factor(d$street)121   -0.166743760  0.523086485
factor(d$street)122   -0.433918884  0.112513623
factor(d$street)124   -0.178341786  0.257346990
factor(d$street)125    0.016966561  0.582296153
factor(d$street)126   -0.380888778  0.173986448
factor(d$street)127   -0.679969647  0.827053351
factor(d$street)128   -0.079621542  0.370348383
```

```
factor(d$street)129   -0.894837687  1.649146099
factor(d$street)130   -0.104761602  0.456959685
factor(d$street)131   -0.041318047  0.592731495
factor(d$street)132    0.035286217  0.434071614
factor(d$street)133    0.055343184  0.576481350
factor(d$street)134   -0.284352668  0.263443156
factor(d$street)136   -0.114846902  0.476564165
factor(d$street)137   -0.052059604  0.562286456
factor(d$street)138   -0.224114017  0.466528698
factor(d$street)147    0.023596224  0.361991934
factor(d$street)148   -0.445904425  0.657937907
factor(d$street)149    0.039845258  0.523702856
factor(d$street)151   -0.710406390  0.183657406
factor(d$street)152   -0.541691248  0.790947583
factor(d$street)153   -0.327992627  0.272579616
factor(d$street)154   -0.284906812  0.166224134
factor(d$street)155   -0.727249453  0.237476315
factor(d$street)156   -0.294008175  0.685514803
factor(d$street)157   -0.019600804  0.317365908
factor(d$street)158   -0.064739457  0.860923350
factor(d$street)160   -0.288320769  0.045370396
factor(d$street)163   -0.250991904  0.095074234
factor(d$street)164   -0.221232066  0.025532537
factor(d$street)165   -0.271420549  0.290789113
factor(d$street)166   -0.435282111  0.139953420
factor(d$street)168   -0.894022616  1.196195600
factor(d$street)170   -0.142872545  0.128027162
factor(d$street)171   -0.264898361  0.170736997
factor(d$street)172   -0.119602928  0.069000300
factor(d$street)175   -0.081675188  0.584050181
factor(d$street)179   -0.341913645  0.349402841
factor(d$street)180   -0.226923067  0.031772852
factor(d$street)182   -0.392089427  0.100957580
factor(d$street)183   -0.232934634  0.066984462
factor(d$street)185   -0.262305820  0.138242839
factor(d$street)186   -0.183716700  0.190904241
factor(d$street)187   -0.160113308  0.140410913
factor(d$street)188   -0.133181639  0.076057332
factor(d$street)189   -0.177701866  0.067434834
factor(d$street)190   -0.231530764  0.254085083
factor(d$street)191   -0.203812903  0.172782345
factor(d$street)192   -0.232486886  0.022828449
factor(d$street)193   -0.162005822  0.285965423
factor(d$street)196   -0.478455100  0.023625754
factor(d$street)197   -0.864416147  1.006809070
factor(d$street)198   -0.130906787  0.297889277
factor(d$street)200   -0.170971860  0.082540729
factor(d$street)202   -0.285715278  0.299195809
factor(d$street)203   -0.395819248 -0.056206689
factor(d$street)206   -0.421381137 -0.054094692
factor(d$street)207   -0.336339615 -0.137270540
factor(d$street)208   -0.334131647  0.100322478
factor(d$street)209   -0.483951812 -0.135258958
factor(d$street)210   -0.397425173 -0.061836103
factor(d$street)213   -0.120166101  0.205039769
factor(d$street)215   -0.199264688  0.081413266
factor(d$street)216   -1.282598059  0.369920393
factor(d$street)217   -0.423056923  0.064060800
factor(d$street)220   -0.509363542  0.098814369
factor(d$street)221   -0.350949526  0.049723012
factor(d$street)222   -0.449897442  0.101166631
factor(d$street)223   -0.423468342  0.064479666
factor(d$street)225   -0.332232626  0.151885483
factor(d$street)227   -0.398778149  0.305627205
factor(d$street)228   -0.455057399  0.554454222
factor(d$street)229   -0.289989870  0.363391370
factor(d$street)230   -0.181888648  0.202076090
factor(d$street)232   -0.537405103  0.693187361
factor(d$street)233   -0.214580286  0.319462432
factor(d$street)235   -0.468540809  0.246511877
factor(d$street)236   -0.154680563  0.127766238
factor(d$street)238   -0.295687355  0.079417029
factor(d$street)240   -0.443692230 -0.150473097
factor(d$street)241   -0.545951725  0.338736046
factor(d$street)242   -0.109948844  0.189177065
factor(d$street)243   -0.487978887  0.804720911
factor(d$street)244   -0.129304409  0.310838326
factor(d$street)246   -0.344063215  0.069707601
factor(d$street)247   -0.111384475  0.105802368
factor(d$street)248   -0.153603670  0.142274170
factor(d$street)249   -0.162229933  0.335763426
factor(d$street)250   -0.051261523  0.142156377
factor(d$street)253   -0.059839891  0.277903676
factor(d$street)254   -0.148789640  0.351154286
factor(d$street)255   -0.177164396  0.232458905
factor(d$street)256   -0.096405303  0.146001724
factor(d$street)257   -0.243106638  0.035902265
factor(d$street)258   -0.015029937  0.208359703
factor(d$street)259    0.043367265  1.359914364
factor(d$street)260   -0.176952577  0.112649856
```

```
factor(d$street)261   -0.199729355  0.112773169
factor(d$street)262   -0.182718446  0.254075991
factor(d$street)263   -0.163426466  0.217228755
```

4. 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.

Hide

```
anova(mod_1,mod_3,test.statistic='F')
```

```
Error: $ operator is invalid for atomic vectors
```

Hide

```
test_fixed_effects <- 'fill this in'
```

> I could not get this test to work, but if we failed to reject the null hypothesis, this would indicate no statistically significant difference between the variances in the two models, in which case we would prefer the simpler model as the added complexity made no observable difference to the power of the model. If the test successfully rejected the null, I would have to weigh the tradeoffs between significantly increased complexity and additional predictive power.

5. 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.

Hide

```

d$no_cell_phone = 1-d$havecell
```

6. 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.

Hide

```
mod_5 <- lm(d$avg_bins_treat ~ d$bin + d$base_avg_bins_treat + factor(d$street) + d$havecell)
mod_5.vcovHC <- vcovHC(mod_5)
mod_3_robust_ci <- coefci(mod_5, vcov.=mod_5.vcovHC, level=0.95)
mod_3_robust_ci
```

```
                          2.5 %      97.5 %
(Intercept)            0.267315517  0.408651801
d$bin                  0.078012761  0.152188727
d$base_avg_bins_treat  0.313759946  0.433004654
factor(d$street)2     -0.217184229  0.015155885
factor(d$street)3     -0.264213537  0.154550368
factor(d$street)4     -0.264951353  0.010078934
factor(d$street)5     -0.305054815  0.099003185
factor(d$street)6     -0.289151044  0.172085701
factor(d$street)7     -0.380151801  0.231090127
factor(d$street)8     -0.168296553  0.055958114
factor(d$street)9     -0.286471174  0.106048822
factor(d$street)10    -0.188401514  0.124729278
factor(d$street)11    -0.298728314  0.141934028
factor(d$street)15    -0.272734245  0.216456434
factor(d$street)17    -0.166793401  0.239124900
factor(d$street)20    -0.376119114 -0.085762523
factor(d$street)21    -0.225749356  0.115908851
factor(d$street)22    -0.249077401  0.013933124
factor(d$street)23    -0.263443286 -0.047021400
factor(d$street)26    -0.209785328  0.180650410
factor(d$street)32    -0.367255592  0.704752816
factor(d$street)37    -0.253336184  0.558836113
factor(d$street)38    -0.265652472  0.103016225
factor(d$street)40    -0.746909398  0.614708850
factor(d$street)41    -0.339779212  0.384711709
factor(d$street)42    -0.437494918  0.205821403
factor(d$street)43    -0.262263870  0.348272268
factor(d$street)44    -0.116621601  0.345642246
factor(d$street)45    -0.277911559  0.185086198
factor(d$street)46    -0.209109115  0.206621257
factor(d$street)47    -0.202776932  1.026485224
factor(d$street)53    -0.389157418  0.421807524
factor(d$street)58    -0.153741710  0.090801118
factor(d$street)60    -0.110076133  0.163355693
factor(d$street)61    -0.367490413  0.341909815
factor(d$street)62    -0.225384039  0.300247612
factor(d$street)63    -0.211292378  0.151677824
factor(d$street)64    -0.167403519  0.246003372
factor(d$street)66    -0.165789534  0.062528032
factor(d$street)67    -0.309982684  0.299513900
factor(d$street)68    -0.150325095  0.317945593
factor(d$street)69    -0.092027273  0.225779525
factor(d$street)70    -0.184864078  0.141990864
factor(d$street)72    -0.209255293  0.120361438
factor(d$street)73    -0.294720211 -0.063180005
factor(d$street)74    -0.209472355  0.064132638
factor(d$street)75    -0.148587975  0.054065808
factor(d$street)77    -0.375250639  0.296479052
factor(d$street)78    -0.217719756  0.099040809
factor(d$street)79    -0.716586353  1.364026306
factor(d$street)80    -0.086779566  0.110803944
factor(d$street)81    -0.314696489  0.261587931
factor(d$street)82    -0.411441429  0.469269481
factor(d$street)83    -0.368523210  0.131950721
factor(d$street)84    -0.141410972  0.136564335
factor(d$street)85    -0.082038164  0.215688769
factor(d$street)86    -0.217473239  0.069614053
factor(d$street)88    -0.306519758  0.101427008
factor(d$street)89    -0.208668142  0.185795489
factor(d$street)91    -0.582152173  0.120398601
factor(d$street)93    -0.252428881  0.115040923
factor(d$street)94    -0.215810008  0.112435507
factor(d$street)96    -0.374085620  0.299082120
factor(d$street)98    -0.127118711  0.207895525
factor(d$street)99     0.049536701  0.416688737
factor(d$street)100   -0.217501403  0.356382943
factor(d$street)101   -0.022699036  0.443570916
factor(d$street)102   -0.446934979  0.336256066
factor(d$street)103   -0.186872336  0.363192109
factor(d$street)105   -0.076683064  0.308466007
factor(d$street)106    0.127856574  0.650250358
factor(d$street)107   -0.332492975  0.421551745
factor(d$street)109   -0.273120461  0.229460935
factor(d$street)110   -0.313935873  0.555463838
factor(d$street)111   -0.361538055  0.324384644
factor(d$street)112    0.008855785  0.358536482
factor(d$street)113   -0.287424264  0.410304554
factor(d$street)115    0.006778160  0.546920642
factor(d$street)117   -0.583323655  1.725169250
factor(d$street)118   -0.711100470  0.737075755
factor(d$street)119   -0.243694566  0.206640621
factor(d$street)120   -0.291863161  0.129560855
factor(d$street)121   -0.187646743  0.529337353
factor(d$street)122   -0.442586195  0.106384661
factor(d$street)124   -0.190558280  0.238558859
factor(d$street)125    0.005887504  0.584873020
factor(d$street)126   -0.400940353  0.154055259
factor(d$street)127   -0.681341928  0.821937377
factor(d$street)128   -0.066855543  0.367047130
```

```
factor(d$street)129   -0.871412580  1.618088716
factor(d$street)130   -0.110500599  0.459977012
factor(d$street)131   -0.053921930  0.590122573
factor(d$street)132    0.031502711  0.431227512
factor(d$street)133    0.035808027  0.605207260
factor(d$street)134   -0.268827658  0.250379026
factor(d$street)136   -0.135140289  0.457486534
factor(d$street)137   -0.021285433  0.524711687
factor(d$street)138   -0.239905416  0.450937739
factor(d$street)147    0.055518801  0.322292923
factor(d$street)148   -0.435412332  0.632559441
factor(d$street)149    0.032724686  0.512586312
factor(d$street)151   -0.753056648  0.219771356
factor(d$street)152   -0.498513964  0.773856161
factor(d$street)153   -0.295334248  0.266199309
factor(d$street)154   -0.285987325  0.176701655
factor(d$street)155   -0.746295900  0.216715573
factor(d$street)156   -0.266259531  0.683827994
factor(d$street)157   -0.004695226  0.310702967
factor(d$street)158   -0.065529160  0.840833045
factor(d$street)160   -0.287236584  0.049250771
factor(d$street)163   -0.271864840  0.075285067
factor(d$street)164   -0.221175132  0.044789005
factor(d$street)165   -0.261721111  0.289386373
factor(d$street)166   -0.459728232  0.190605535
factor(d$street)168   -0.844237835  1.170569033
factor(d$street)170   -0.133485885  0.125237707
factor(d$street)171   -0.251023636  0.183020214
factor(d$street)172   -0.104734047  0.088443055
factor(d$street)175   -0.067910770  0.591726730
factor(d$street)179   -0.362251332  0.329565553
factor(d$street)180   -0.223781027  0.036122166
factor(d$street)182   -0.402046519  0.104421794
factor(d$street)183   -0.219893141  0.078107835
factor(d$street)185   -0.250329340  0.135836912
factor(d$street)186   -0.188184234  0.180281642
factor(d$street)187   -0.142663093  0.146708220
factor(d$street)188   -0.132404968  0.068432010
factor(d$street)189   -0.174116208  0.071590561
factor(d$street)190   -0.213116785  0.256471411
factor(d$street)191   -0.216957912  0.179315885
factor(d$street)192   -0.227413062  0.023506916
factor(d$street)193   -0.150904881  0.309194137
factor(d$street)196   -0.485494271  0.007409085
factor(d$street)197   -0.884465119  0.987098906
factor(d$street)198   -0.126005765  0.306372336
factor(d$street)200   -0.165986597  0.093589358
factor(d$street)202   -0.278318598  0.331732642
factor(d$street)203   -0.410121149 -0.067512698
factor(d$street)206   -0.428323640 -0.047549038
factor(d$street)207   -0.351236202 -0.112667559
factor(d$street)208   -0.319933411  0.115433325
factor(d$street)209   -0.493027453 -0.135477390
factor(d$street)210   -0.379436715 -0.040169384
factor(d$street)213   -0.110031363  0.220319081
factor(d$street)215   -0.192960993  0.086587085
factor(d$street)216   -1.287356586  0.375024485
factor(d$street)217   -0.420879072  0.047236377
factor(d$street)220   -0.490176381  0.096415156
factor(d$street)221   -0.356716327  0.037662992
factor(d$street)222   -0.444070481  0.134883956
factor(d$street)223   -0.430294635  0.080907700
factor(d$street)225   -0.321536637  0.144423265
factor(d$street)227   -0.412872777  0.299528238
factor(d$street)228   -0.478011085  0.570367047
factor(d$street)229   -0.263951921  0.363028066
factor(d$street)230   -0.165683717  0.188208476
factor(d$street)232   -0.508353662  0.672648856
factor(d$street)233   -0.220304099  0.318205450
factor(d$street)235   -0.469043737  0.226431982
factor(d$street)236   -0.140112812  0.139557571
factor(d$street)238   -0.289385362  0.087793333
factor(d$street)240   -0.454769992 -0.159176420
factor(d$street)241   -0.578087597  0.355199609
factor(d$street)242   -0.092713427  0.198051537
factor(d$street)243   -0.509578665  0.785215618
factor(d$street)244   -0.116912908  0.329309406
factor(d$street)246   -0.349719191  0.100545681
factor(d$street)247   -0.107426742  0.100134200
factor(d$street)248   -0.144415104  0.144937177
factor(d$street)249   -0.148875727  0.338644992
factor(d$street)250   -0.043495213  0.148106978
factor(d$street)253   -0.047373826  0.289674099
factor(d$street)254   -0.136944190  0.355476588
factor(d$street)255   -0.168969342  0.243220554
factor(d$street)256   -0.086972123  0.156058609
factor(d$street)257   -0.241829124  0.039037043
factor(d$street)258   -0.011735347  0.203123372
factor(d$street)259    0.072372983  1.389217472
factor(d$street)260   -0.179792432  0.108900646
```

```
factor(d$street)261   -0.190624077  0.124834137
factor(d$street)262   -0.176574265  0.273740446
factor(d$street)263   -0.157496667  0.218184325
d$havecell             0.014662584  0.084357190
```

7. 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.

Hide

```
mod_6 <- lm(d$avg_bins_treat ~ d$bin + d$base_avg_bins_treat+ d$havecell + d$sms + factor(d$street))
mod_6.vcovHC <- vcovHC(mod_6)
mod_6_robust_ci <- coefci(mod_6, vcov.=mod_6.vcovHC, level=0.95)
mod_6_robust_ci
```

```
                            2.5 %       97.5 %
(Intercept)            0.267216289  0.408664231
d$bin                  0.077940398  0.152166899
d$base_avg_bins_treat  0.313926871  0.433038849
d$havecell             0.001597533  0.091806576
d$sms                 -0.041132580  0.051381329
factor(d$street)2     -0.217196698  0.015530725
factor(d$street)3     -0.264874361  0.154568068
factor(d$street)4     -0.265472394  0.009711886
factor(d$street)5     -0.304681784  0.098424923
factor(d$street)6     -0.290662264  0.172683060
factor(d$street)7     -0.381797969  0.232237954
factor(d$street)8     -0.168794341  0.056403148
factor(d$street)9     -0.287120735  0.107016841
factor(d$street)10    -0.188383821  0.123679469
factor(d$street)11    -0.298819481  0.141012884
factor(d$street)15    -0.271001970  0.216405462
factor(d$street)17    -0.165317830  0.237960970
factor(d$street)20    -0.377348474 -0.085671955
factor(d$street)21    -0.225998004  0.115422842
factor(d$street)22    -0.249123252  0.013798841
factor(d$street)23    -0.263763961 -0.046927843
factor(d$street)26    -0.211853129  0.181099304
factor(d$street)32    -0.369887538  0.705864292
factor(d$street)37    -0.252996833  0.557985819
factor(d$street)38    -0.267495036  0.103246098
factor(d$street)40    -0.746784497  0.614885088
factor(d$street)41    -0.339119442  0.384212033
factor(d$street)42    -0.437434573  0.205877130
factor(d$street)43    -0.264621478  0.351021684
factor(d$street)44    -0.118212427  0.346457748
factor(d$street)45    -0.277640894  0.183925263
factor(d$street)46    -0.210638697  0.207641082
factor(d$street)47    -0.201906060  1.027140121
factor(d$street)53    -0.387158144  0.418558770
factor(d$street)58    -0.154093328  0.089871795
factor(d$street)60    -0.110035483  0.162989370
factor(d$street)61    -0.369844700  0.344638538
factor(d$street)62    -0.224985830  0.301070622
factor(d$street)63    -0.211836928  0.150630146
factor(d$street)64    -0.166581061  0.246415930
factor(d$street)66    -0.167141441  0.063322218
factor(d$street)67    -0.309842190  0.299540267
factor(d$street)68    -0.150799675  0.317945363
factor(d$street)69    -0.092739152  0.226435421
factor(d$street)70    -0.185382021  0.141473115
factor(d$street)72    -0.209349680  0.119708613
factor(d$street)73    -0.295241119 -0.063576099
factor(d$street)74    -0.209355813  0.064146611
factor(d$street)75    -0.149691523  0.054111663
factor(d$street)77    -0.374673717  0.295546694
factor(d$street)78    -0.217931754  0.098993259
factor(d$street)79    -0.715560584  1.364393791
factor(d$street)80    -0.087024030  0.111324602
factor(d$street)81    -0.316981315  0.261627258
factor(d$street)82    -0.409104542  0.470708644
factor(d$street)83    -0.369002507  0.131331580
factor(d$street)84    -0.141042821  0.136338928
factor(d$street)85    -0.082329333  0.215944230
factor(d$street)86    -0.217857946  0.069231437
factor(d$street)88    -0.306350588  0.100006018
factor(d$street)89    -0.209270887  0.185537797
factor(d$street)91    -0.578250492  0.118747491
factor(d$street)93    -0.253413866  0.116309931
factor(d$street)94    -0.216093105  0.110922667
factor(d$street)96    -0.373657981  0.298945686
factor(d$street)98    -0.126262667  0.210016645
factor(d$street)99     0.049290588  0.417784295
factor(d$street)100   -0.217752999  0.356354801
factor(d$street)101   -0.022340536  0.445023726
factor(d$street)102   -0.447006714  0.336035376
factor(d$street)103   -0.187723285  0.363778225
factor(d$street)105   -0.076266213  0.307257494
factor(d$street)106    0.128015041  0.652182562
factor(d$street)107   -0.330101107  0.422179147
factor(d$street)109   -0.271393451  0.230621740
factor(d$street)110   -0.313278121  0.558350664
factor(d$street)111   -0.361954102  0.326265351
factor(d$street)112    0.008691813  0.360794381
factor(d$street)113   -0.285525039  0.411074469
factor(d$street)115    0.007032359  0.547466149
factor(d$street)117   -0.582239206  1.725613984
factor(d$street)118   -0.711488311  0.739613607
factor(d$street)119   -0.243078754  0.208914175
factor(d$street)120   -0.290931293  0.132187334
factor(d$street)121   -0.186636292  0.531141643
factor(d$street)122   -0.439859240  0.105241633
factor(d$street)124   -0.190413043  0.241648884
factor(d$street)125    0.006318397  0.585583544
factor(d$street)126   -0.399724630  0.156612809
```

```
factor(d$street)127   -0.681319632  0.823792388
factor(d$street)128   -0.066657604  0.367909438
factor(d$street)129   -0.875233358  1.622185493
factor(d$street)130   -0.110830207  0.462465743
factor(d$street)131   -0.053064808  0.592056602
factor(d$street)132    0.032328521  0.432669238
factor(d$street)133    0.038130716  0.605675457
factor(d$street)134   -0.269439953  0.251198104
factor(d$street)136   -0.135229375  0.457938426
factor(d$street)137   -0.023129309  0.526833685
factor(d$street)138   -0.238932658  0.453092918
factor(d$street)147    0.053539468  0.328014785
factor(d$street)148   -0.435251431  0.636576080
factor(d$street)149    0.033768806  0.514678461
factor(d$street)151   -0.753946108  0.220846758
factor(d$street)152   -0.499655773  0.776862354
factor(d$street)153   -0.295580149  0.264826305
factor(d$street)154   -0.287186420  0.177632732
factor(d$street)155   -0.752225313  0.221378747
factor(d$street)156   -0.266602432  0.682627410
factor(d$street)157   -0.005886329  0.312234307
factor(d$street)158   -0.067286754  0.843006311
factor(d$street)160   -0.287588990  0.049877930
factor(d$street)163   -0.271084517  0.075768150
factor(d$street)164   -0.222727862  0.044482248
factor(d$street)165   -0.263192315  0.291178958
factor(d$street)166   -0.462228296  0.191511833
factor(d$street)168   -0.845849636  1.174133151
factor(d$street)170   -0.134122147  0.124710455
factor(d$street)171   -0.251736674  0.183864370
factor(d$street)172   -0.104438771  0.088235596
factor(d$street)175   -0.069147862  0.592757393
factor(d$street)179   -0.367222254  0.332398351
factor(d$street)180   -0.224045001  0.036340786
factor(d$street)182   -0.403091045  0.104075090
factor(d$street)183   -0.220437707  0.078040326
factor(d$street)185   -0.250074170  0.134970010
factor(d$street)186   -0.188389869  0.180783030
factor(d$street)187   -0.142733819  0.146615896
factor(d$street)188   -0.131929806  0.068251730
factor(d$street)189   -0.174518693  0.071985713
factor(d$street)190   -0.213261656  0.256191711
factor(d$street)191   -0.216246287  0.178913817
factor(d$street)192   -0.227871630  0.023824711
factor(d$street)193   -0.150522822  0.308893610
factor(d$street)196   -0.486012245  0.009124365
factor(d$street)197   -0.885069465  0.986420253
factor(d$street)198   -0.126233725  0.306422625
factor(d$street)200   -0.166323500  0.094869423
factor(d$street)202   -0.279522446  0.331894277
factor(d$street)203   -0.409948623 -0.065826367
factor(d$street)206   -0.428974863 -0.047666852
factor(d$street)207   -0.351653379 -0.111982382
factor(d$street)208   -0.319968290  0.115064217
factor(d$street)209   -0.493468696 -0.135921366
factor(d$street)210   -0.379142499 -0.040471519
factor(d$street)213   -0.110451173  0.220056254
factor(d$street)215   -0.193037991  0.086029881
factor(d$street)216   -1.286346643  0.375036313
factor(d$street)217   -0.422625854  0.048062260
factor(d$street)220   -0.489840183  0.095545467
factor(d$street)221   -0.357677743  0.037257373
factor(d$street)222   -0.442851444  0.134808212
factor(d$street)223   -0.431128836  0.081936829
factor(d$street)225   -0.321843023  0.143290424
factor(d$street)227   -0.411692375  0.298699418
factor(d$street)228   -0.480724673  0.570026307
factor(d$street)229   -0.264364011  0.363447640
factor(d$street)230   -0.166040885  0.188815524
factor(d$street)232   -0.508430642  0.670409610
factor(d$street)233   -0.220496475  0.317884736
factor(d$street)235   -0.468837240  0.226712823
factor(d$street)236   -0.139845201  0.139448323
factor(d$street)238   -0.289194929  0.087857732
factor(d$street)240   -0.455300234 -0.160312902
factor(d$street)241   -0.580424965  0.356645115
factor(d$street)242   -0.093784701  0.197562565
factor(d$street)243   -0.508411031  0.787139488
factor(d$street)244   -0.116961464  0.329511121
factor(d$street)246   -0.349055113  0.100060878
factor(d$street)247   -0.107965480  0.100155714
factor(d$street)248   -0.144541886  0.145273426
factor(d$street)249   -0.148858223  0.337839940
factor(d$street)250   -0.043856270  0.148079180
factor(d$street)253   -0.047415015  0.289883402
factor(d$street)254   -0.136705775  0.356134648
factor(d$street)255   -0.169197200  0.243646206
factor(d$street)256   -0.087109729  0.156138219
factor(d$street)257   -0.242333698  0.039170989
factor(d$street)258   -0.011475434  0.203437337
```

```
factor(d$street)259    0.072370880  1.389193187
factor(d$street)260   -0.179706798  0.109099046
factor(d$street)261   -0.191276347  0.125043776
factor(d$street)262   -0.176390161  0.274281267
factor(d$street)263   -0.158047219  0.218645369
```

8. 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.

Hide

```
mod_7 <- 'fill this in'
```

4. 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.

Hide

```
d <- fread("ebola_rct2.csv")
head(d)
```

temperature_day0 <dbl>	dehydrated_day0 <int>	treat_zmapp <int>	temperature_day14 <dbl>	dehydrated_day14 <int>	male <int>
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

6 rows

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

1. 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?

Hide

```
zmapp_1 <- lm(d$dehydrated_day14 ~ d$treat_zmapp)
summary(zmapp_1)
```

```
Call:
lm(formula = d$dehydrated_day14 ~ d$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 ***
d$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
```

P = 0.006 << 0.05, ATE = -0.23

2. 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.

Hide

```
zmapp_2 <- lm(d$dehydrated_day14 ~ d$treat_zmapp + d$dehydrated_day0 + d$temperature_day0)
summary(zmapp_2)
```

```
Call:
lm(formula = d$dehydrated_day14 ~ d$treat_zmapp + d$dehydrated_day0 +
    d$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 *
d$treat_zmapp       -0.16554    0.07567  -2.188  0.03113 *
d$dehydrated_day0    0.06456    0.14635   0.441  0.66013
d$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
```

> ATE = -16.6% reduction in patient dehydration with a p-value = 7.6e-8 << 0.05

3. Do you prefer the estimate of the ATE reported in part (a) or part (b)? Why? Report the results of the F-test that you used to form this opinion.

Hide

```
zmapp_test_object <- anova(zmapp_1, zmapp_2, test.statistic='F')
```

```
Error: $ operator is invalid for atomic vectors
```

4. The regression from part (2) 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.

Hide

```
zmapp_3 <- lm(d$dehydrated_day14 ~ d$treat_zmapp + d$dehydrated_day0 + d$temperature_day0 + d$temperature_day14)
summary(zmapp_3)
```

```
Call:
lm(formula = d$dehydrated_day14 ~ d$treat_zmapp + d$dehydrated_day0 +
    d$temperature_day0 + d$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 **
d$treat_zmapp        -0.12010    0.07768  -1.546  0.12541
d$dehydrated_day0     0.04604    0.14426   0.319  0.75033
d$temperature_day0    0.17664    0.07642   2.312  0.02296 *
d$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
```

5. Do you prefer the estimate of the ATE reported in part (b) or part (d)? What is this preference based on?

6. 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?

Hide

```
zmapp_4 <- lm(d$temperature_day14 ~ d$treat_zmapp + d$dehydrated_day0 + d$temperature_day0 + d$male + d$male*d$treat_zmapp)
zmapp_4.vcovHC <- vcovHC(zmapp_4)
summary(zmapp_4)
```

```
Call:
lm(formula = d$temperature_day14 ~ d$treat_zmapp + d$dehydrated_day0 +
    d$temperature_day0 + d$male + d$male * d$treat_zmapp)

Residuals:
     Min       1Q   Median       3Q      Max
-0.70157 -0.37725 -0.02702  0.34687  0.73968

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          48.71269    9.26618   5.257 9.14e-07 ***
d$treat_zmapp        -0.23087    0.11871  -1.945   0.0548 .
d$dehydrated_day0     0.04113    0.18208   0.226   0.8218
d$temperature_day0    0.50480    0.09508   5.309 7.34e-07 ***
d$male                3.08549    0.12644  24.403  < 2e-16 ***
d$treat_zmapp:d$male -2.07669    0.19164 -10.836  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4518 on 94 degrees of freedom
Multiple R-squared:  0.9059,    Adjusted R-squared:  0.9009
F-statistic:   181 on 5 and 94 DF,  p-value: < 2.2e-16
```

In order to test the hypothesis that Zmapp is especially likely to reduce mens' temperatures, I added the male covariate and the male-ZMapp treatment interaction covariate. The coefficient of the interaction term of male and ZMapp treatment could let us know the effect of being male on the outcome of a treated individual. In this model, the coefficient of the interaction term is -2.0766863 (0.1983862) and is statistically significant with p-value of 1.869369810^{-17}. This indicates that Zmapp treatment on males leads to a 2.0766863 reduction in temperature after 14 days as compared to the same treatment on women.

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

Hide

```
cont = d[d$treat_zmapp==0]

zmapp_5 = lm(cont$temperature_day14 ~ factor(cont$male))
summary(zmapp_5)
```

```
Call:
lm(formula = cont$temperature_day14 ~ factor(cont$male))

Residuals:
     Min       1Q   Median       3Q      Max
-1.39584 -0.49433 -0.00544  0.44406  1.29504

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         98.4865     0.1019  966.72   <2e-16 ***
factor(cont$male)1   3.2051     0.1795   17.85   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6443 on 57 degrees of freedom
Multiple R-squared:  0.8483,    Adjusted R-squared:  0.8456
F-statistic: 318.7 on 1 and 57 DF,  p-value: < 2.2e-16
```

Females have better health outcomes in control, which suggests that the heterogeneous treatment effects that translates to the zmapp treatment not being as effective on females may true, as it could be that females are better able to recover from ebola untreated and thus this higher proportion of recovering females cancels out the treatment effect on sicker individuals.

8. Suppose that you had not run the regression in part (7). Instead, 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 gender 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 gender 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 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 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.)

Had I not run the regression in part (7), I likely would not have noticed that males and females appear to react differently to ebola (males getting sicker than females). Hearing several authority figures say that the results look good combined with the fact that there is a plausible theoretical explanation, I would be tempted to accept the results that Zmapp works especially well for men. I would, however, be suspicious of the results on account of the fact that my colleague had run 80 covariates and be worried of the increased potential for false discovery.I would be on the fence about whether it works especially well for men.

9. Now, imagine that your colleague's fishing expedition 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?

If I only tested this HTE, I wouldn't be worried about the issues associated with fishing expeditions. However, if in my own study I had run the regression in part 7, I would be very suspicious that this HTE truly exists as it appears that gender affects outcomes regardless of treatment.

10. Now, imagine that your colleague's fishing expedition **did** happen, but that you on your own tested this and only this HTE, discover a positive result and conclude there is an effect. How does your colleague's behavior change the interpretation of your test? Does this seem fair or reasonable?

This probably would not change my interpretation of the results, as ***I*** wasn't fishing for sex based effects when I found my significant result. This isn't really fair or reasonable, because how many covariates you test doesn't have any actual impact on which variables are truly causal in nature. Someone could analyze sex as the first covariate they test, or the 100th covariate they test and it wouldn't matter. I suppose its just part of the uncertainty of designing experiments.