Code ▾

W241 Class Project - Analysis

1. Load libraries, data

Load up the data and do simple analysis. This version uses the complete dataset.

Hide

```
# Libraries
library(lmtest)
library(sandwich)
library(ggplot2)
library(data.table)
library(stargazer)
library(ri)
library(multiwayvcov)
library(AER)
rm(list=ls())
d <- read.csv('~/Documents/mids-w241-final/Analysis/Combined Log.csv')
#d <- read.csv("C:/Users/Chris/OneDrive/Documents/MIDS/WS241/final/mids-w241-final/Analysis/Combined Log.csv")
d <- data.table(d)
# d <- d[complete.cases(d),] drops any row that is incomplete - too stringent since we have some cols which are not essential
d <- d[!is.na(no)] # Just drop row with missing values
# d <- d[complete.cases(d),] drops any row that is too incomplete - too stringent since we have some cols which are not essential
#d <- d[!is.na(no)] # Just drop row with missing values
# Base data
head(d)
```

n	city	title
<int>	<fctr>	<fctr>
6	seattle	Top Floor 2x2 Park & Lake Side/Corner Home!! USB Plugins!!
87	chicago	836 S Bishop St #G
62	houston	Sophisticated Living 2 BR APT with Excellent Amenities-Receive Up To $
89	chicago	ALL BRAND NEW OUTSTANDING EAST LAKEVIEW LOCATION
96	chicago	Upgrade your lifestyle. Inexpensive with great character!
26	seattle	LUXURY 2 BEDROOM ***2 Master Suites*Coming Soon! Call for Availability

6 rows | 1-4 of 17 columns

Hide

```
# Convert, transform data for analysis
# Drop some cols
d[,c('title','full_URL', 'reply_email_TO_BE_FILLED_IN_standard','posting_ID','notes') :=NULL]
# Set gender = 1 for Jane
d[treatment_assignment=='Jane_Control' | treatment_assignment=='Jane_Treat_High' | treatment_assignment=='Jane_Treat_Low',gender:=1]
d[treatment_assignment=='John_Control' | treatment_assignment=='John_Treat_High' | treatment_assignment=='John_Treat_Low',gender:=0]
# Set treatment variable = 0 for control, 1 for low, 2 for high (treatment here is continuous)
d[treatment_assignment=='Jane_Control' | treatment_assignment=='John_Control', treatment:=0]
d[treatment_assignment=='Jane_Treat_Low' | treatment_assignment=='John_Treat_Low', treatment:=1]
d[treatment_assignment=='Jane_Treat_High' | treatment_assignment=='John_Treat_High', treatment:=2]
# Alternatively, treat treatment types as categorical variables instead of continuous
d[treatment_assignment=='Jane_Treat_Low' | treatment_assignment=='John_Treat_Low', low_treatment:=1]
d[treatment_assignment=='Jane_Treat_High' | treatment_assignment=='John_Treat_High', high_treatment:=1]
d$low_treatment[is.na(d$low_treatment)] <- 0
d$high_treatment[is.na(d$high_treatment)] <- 0
d[low_treatment==1 | high_treatment==0, assigned:=1]
d$assigned[is.na(d$assigned)] <- 0
# Capture complier
#d[sent!='', compliers:=1]
#d$compliers[is.na(d$compliers)] <- 0
# Labeling data
d$gender <- factor(d$gender,labels = c("Male", "Female"))
d$outcome_f <- factor(d$outcome, labels = c("No Response", "Response"))
d$bedrooms <- factor(d$bedrooms, labels = c("1-bedroom", "2-bedroom"))
d$professional <- factor(d$professional, labels = c("Non-professional", "Professional"))
d$treatment_f <- factor(d$treatment, labels = c("Control","Low","High"))
head(d)
```

n	**city**	**posting_date**	**bedrooms**	**sqft**	**price**	**treatment_assignment**	**professional**
<int>	<fctr>	<fctr>	<fctr>	<int>	<int>	<fctr>	<fctr>
6	seattle	3/17/2017 13:32	2-bedroom	974	1791	Jane_Treat_High	Professional
87	chicago	3/17/2017 15:22	2-bedroom	*NA*	2050	Jane_Treat_Low	Non-professional
62	houston	3/17/2017 15:24	2-bedroom	1141	1598	Jane_Treat_High	Professional
89	chicago	3/17/2017 15:27	1-bedroom	1100	1995	Jane_Treat_Low	Professional
96	chicago	3/17/2017 15:26	1-bedroom	*NA*	1650	Jane_Control	Non-professional
26	seattle	3/17/2017 13:11	2-bedroom	1234	1655	John_Treat_Low	Professional

6 rows | 1-8 of 19 columns

Hide

```
NA
```

2. Check data, do simple tables to check for balance

Recode missing sqft values. It’s not necessary but we use this in some of our model specifications.

Hide

```
# Recode missing sqft with mean of cluster (city)
d[, Mean:=mean(sqft, na.rm=TRUE), by=city]
d[is.na(sqft)]$sqft <- d[is.na(sqft)]$Mean
```

```
Coerced 'double' RHS to 'integer' to match the column's type; may have truncated precis
ion. Either change the target column to 'double' first (by creating a new 'double' vect
or length 483 (nrows of entire table) and assign that; i.e. 'replace' column), or coerc
e RHS to 'integer' (e.g. 1L, NA_[real|integer]_, as.*, etc) to make your intent clear a
nd for speed. Or, set the column type correctly up front when you create the table and
 stick to it, please.
```

Hide

```
d[,c('Mean') :=NULL]
```

Remove duplicate emails

Hide

```
# If duplicate, only retain first email sent, sets up for exploration later; in our bas
e exploration we did not exclude duplicate emails
#d <- d[duplicate_email == 0]
```

For the most part it looks like we have a balanced dataset.

Hide

```
cat('Table of Outcomes:')
```

```
Table of Outcomes:
```

Hide

```
table(d$outcome_f)
```

```
No Response    Response 
        264         219 
```

Hide

```
cat('\nTable of Outcomes (By Gender):')
```

```

Table of Outcomes (By Gender):
```

Hide

```
table(d$outcome_f, d$gender)
```

```
             
              Male Female
  No Response  135    129
  Response     104    115
```

Hide

```
cat('\nTable of Outcomes (By Treatment):')
```

```

Table of Outcomes (By Treatment):
```

Hide

```
table(d$outcome_f, d$treatment_f)
```

```
             
              Control Low High
  No Response      88  89   87
  Response         72  72   75
```

Hide

```
cat('\nTable of Outcomes (By Treatment and Gender):')
```

```

Table of Outcomes (By Treatment and Gender):
```

Hide

```
table(d$outcome_f, factor(d$treatment_assignment))
```

```
             Jane_Control Jane_Treat_High Jane_Treat_Low John_Control John_Treat_High
John_Treat_Low
  No Response           39              41             49           49              46
           40
  Response              39              40             36           33              35
           36
```

Hide

```
cat('\nTable of Outcomes (By City):')
```

```

Table of Outcomes (By City):
```

Hide

```
table(d$outcome_f,factor(d$city))
```

```
             chicago houston sandiego seattle
  No Response      63      79       65      57
  Response         61      40       52      66
```

Hide

```
cat('\nTable of Outcomes (By Rooms):')
```

```

Table of Outcomes (By Rooms):
```

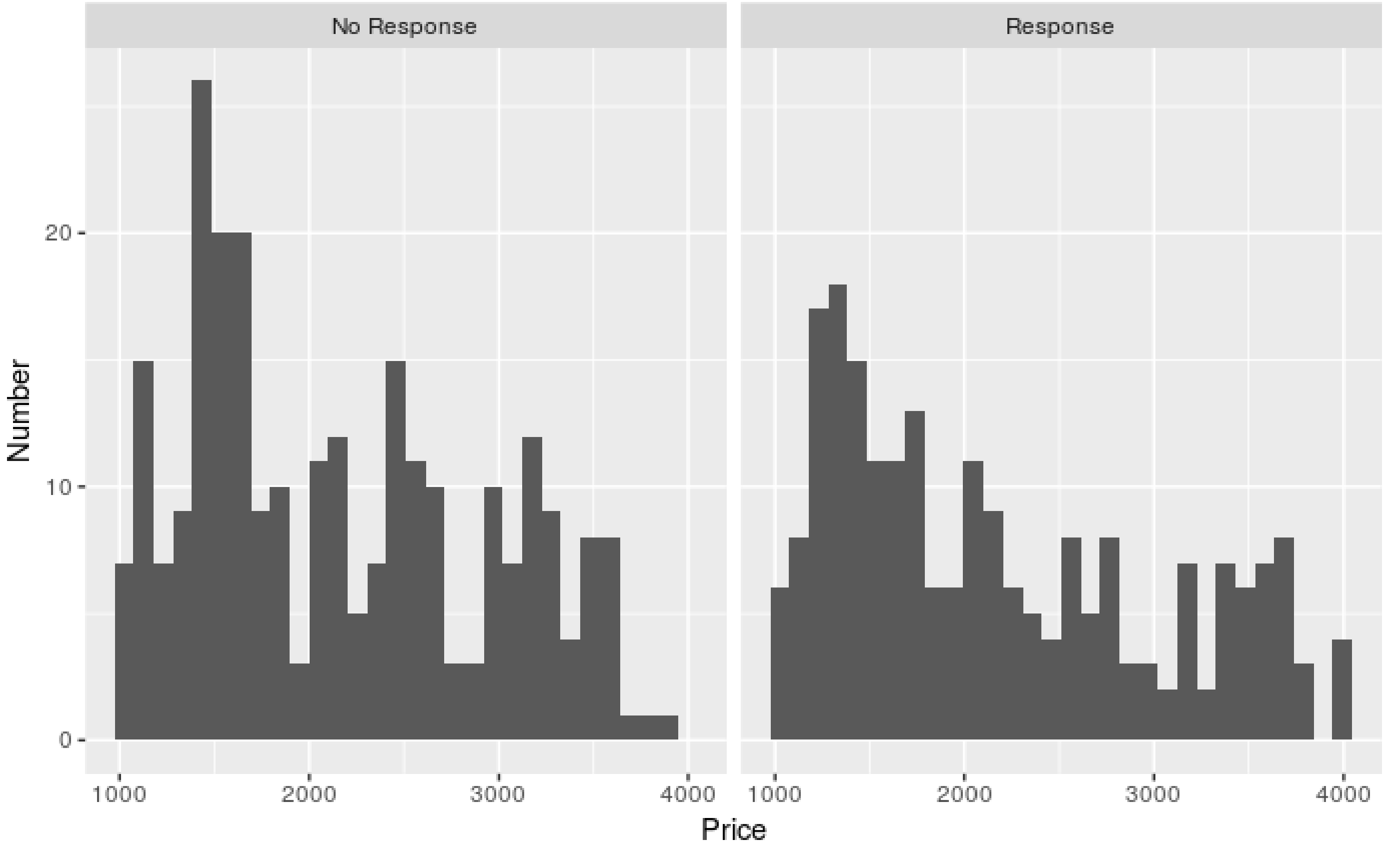
Hide

```
table(d$outcome_f,factor(d$bedrooms))
```

```
             1-bedroom 2-bedroom
  No Response       130       134
  Response          103       116
```

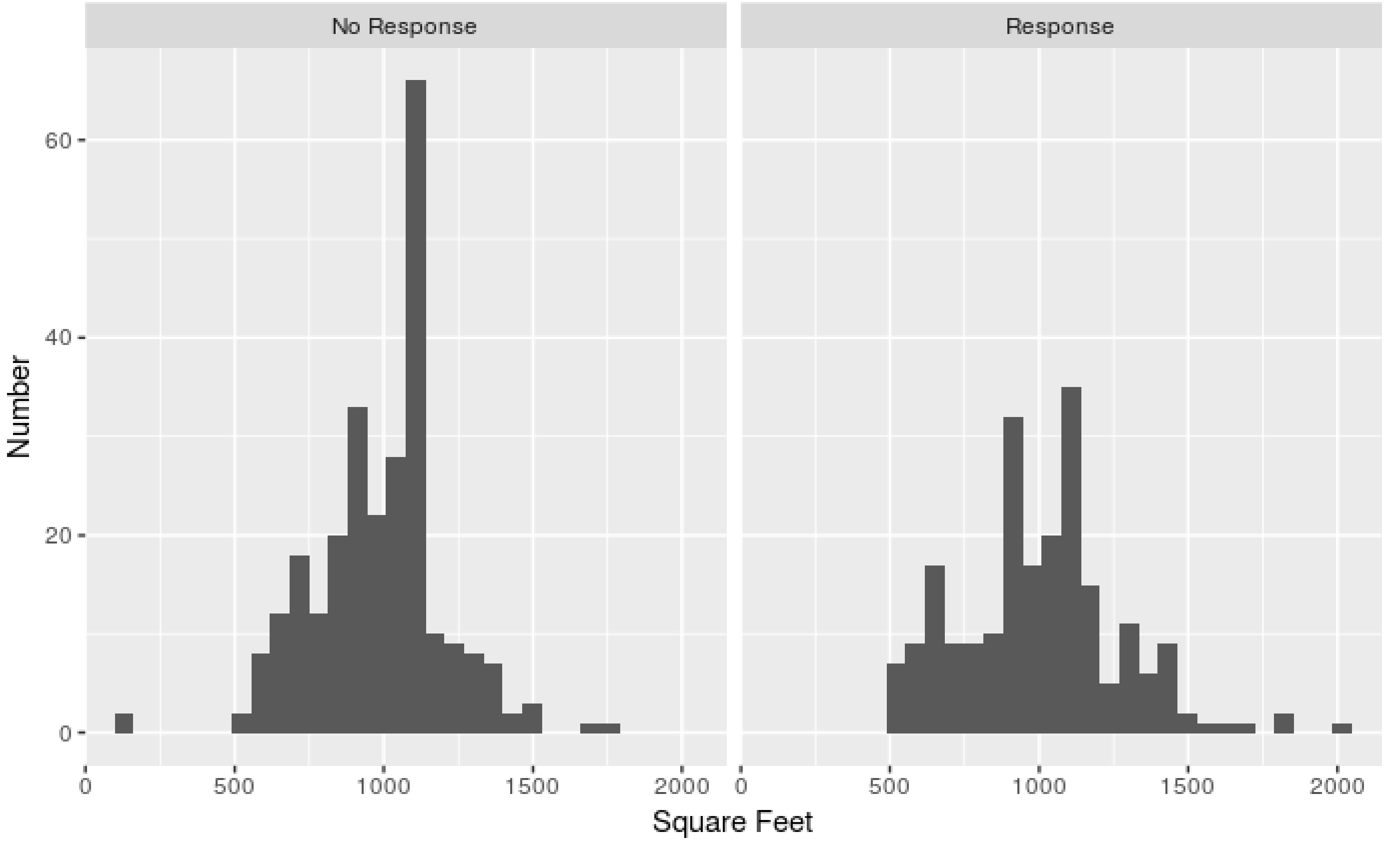
Hide

```
ggplot(d,aes(x=price))+geom_histogram()+facet_grid(~outcome_f)+labs(x="Price",y="Numbe
r")
```



Hide

```
# Sqft info has missing values => we can drop all cases (see above) but for now leave this alone or we use the clustering estimate
# Similar but somewhat worse issue for professional, same.email info
ggplot(d,aes(x=sqft))+geom_histogram()+facet_grid(~outcome_f)+labs(x="Square Feet",y="Number")
```



Hide

```
cat('\nTable of Outcomes (By Professional):')
```

```


Table of Outcomes (By Professional):
```

Hide

```
table(d$outcome_f,factor(d$professional))
```

```

              
               Non-professional Professional
  No Response                55          209
  Response                   35          184
```

3. Analysis

Simple Analysis

We do a chi-squared test of independence to see if the observations are independent. We cannot reject the hypothesis that the observations are indpendent. This is true for even the professional category.

Hide

```
# For Outcome and Gender
tbl <- table(d$outcome_f,d$gender)
tbl
```

```
              Male Female
  No Response  135    129
  Response     104    115
```

Hide

```
chisq.test(tbl)
```

```
	Pearson's Chi-squared test with Yates' continuity correction

data:  tbl
X-squared = 0.49961, df = 1, p-value = 0.4797
```

Hide

```
# On Outcome and Treatment
tbl <- table(d$outcome_f,d$treatment)
tbl
```

```
               0  1  2
  No Response 88 89 87
  Response    72 72 75
```

Hide

```
chisq.test(tbl)
```

```
	Pearson's Chi-squared test

data:  tbl
X-squared = 0.092173, df = 2, p-value = 0.955
```

Hide

```
# On Outcome and Treatment Assignment
tbl <- table(d$outcome_f,factor(d$treatment_assignment))
tbl
```

```
             Jane_Control Jane_Treat_High Jane_Treat_Low John_Control John_Treat_High
 John_Treat_Low
  No Response           39              41             49           49              46
            40
  Response              39              40             36           33              35
            36
```

Hide

```
chisq.test(tbl)
```

```

    Pearson's Chi-squared test

data:  tbl
X-squared = 2.6574, df = 5, p-value = 0.7526
```

Hide

```
# On Outcome and Professional
tbl <- table(d$outcome_f,d$professional)
tbl
```

```
             Non-professional Professional
  No Response               55          209
  Response                  35          184
```

Hide

```
chisq.test(tbl)
```

```

    Pearson's Chi-squared test with Yates' continuity correction

data:  tbl
X-squared = 1.5521, df = 1, p-value = 0.2128
```

Regression

We run regression on treatment as a factor (control, low, high) with and without gender as another factor. Other co-variates are added including city, price, bedrooms.

Basic model

```
Outcome variable = alpha + B_high + B_low + gender + covariates
```

Hide

```
# First we treat treatment as a continous variable
# Model 1a - Basic model
m1 <- lm(outcome~treatment,data=d)
stargazer(m1,type='text')
```

```

===============================================
                        Dependent variable:
                    ---------------------------
                              outcome
-----------------------------------------------
treatment                      0.005
                              (0.048)

Constant                     0.450***
                              (0.039)

-----------------------------------------------
Observations                    483
R2                            0.00002
Adjusted R2                   -0.002
Residual Std. Error       0.499 (df = 481)
F Statistic             0.011 (df = 1; 481)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m1, vcovHC(m1)) # Robust se
```

```

t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.4500000  0.0395777 11.3700   <2e-16 ***
treatment   0.0051084  0.0483624  0.1056   0.9159
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 2a - Treatment & gender
m2 <- lm(outcome~treatment*gender,data=d)
stargazer(m2,type='text')
```

```

==================================================
                           Dependent variable:
                       ---------------------------
                                 outcome
--------------------------------------------------
treatment                         0.050
                                 (0.068)

genderFemale                      0.098
                                 (0.079)

treatment:genderFemale           -0.092
                                 (0.097)

Constant                        0.402***
                                 (0.055)

--------------------------------------------------
Observations                       483
R2                                0.003
Adjusted R2                      -0.003
Residual Std. Error         0.499 (df = 479)
F Statistic               0.516 (df = 3; 479)
==================================================
Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m2, vcovHC(m2)) # Robust se
```

```


t test of coefficients:

                        Estimate Std. Error t value  Pr(>|t|)    
(Intercept)             0.402439   0.054823  7.3407 9.182e-13 ***
treatment               0.049790   0.067850  0.7338    0.4634    
genderFemale            0.097561   0.079338  1.2297    0.2194    
treatment:genderFemale -0.091959   0.096985 -0.9482    0.3435    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 3a - Treatment & gender + covariates
m3 <- lm(outcome~treatment*gender+factor(city)+factor(bedrooms)+price,data=d)
stargazer(m3,type='text')
```

```
=====================================================
                              Dependent variable:    
                          ---------------------------
                                    outcome          
-----------------------------------------------------
treatment                            0.051           
                                    (0.068)          
                                                     
genderFemale                         0.097           
                                    (0.079)          
                                                     
factor(city)houston                -0.155**          
                                    (0.064)          
                                                     
factor(city)sandiego                -0.045           
                                    (0.064)          
                                                     
factor(city)seattle                  0.047           
                                    (0.064)          
                                                     
factor(bedrooms)2-bedroom            0.031           
                                    (0.049)          
                                                     
price                              -0.00000          
                                   (0.00003)         
                                                     
treatment:genderFemale              -0.093           
                                    (0.096)          
                                                     
Constant                           0.432***          
                                    (0.090)          
                                                     
-----------------------------------------------------
Observations                          483            
R2                                   0.026           
Adjusted R2                          0.010           
Residual Std. Error            0.496 (df = 474)      
F Statistic                  1.605 (df = 8; 474)     
=====================================================
Note:                     *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m3, vcovHC(m3)) # Robust se
```

```

t test of coefficients:

                             Estimate  Std. Error t value Pr(>|t|)    
(Intercept)                4.3187e-01  9.2994e-02  4.6441 4.43e-06 ***
treatment                  5.1256e-02  6.7703e-02  0.7571  0.44938    
genderFemale               9.6820e-02  7.9159e-02  1.2231  0.22190    
factor(city)houston       -1.5538e-01  6.3663e-02 -2.4406  0.01503 *  
factor(city)sandiego      -4.5435e-02  6.5768e-02 -0.6908  0.49001    
factor(city)seattle        4.6560e-02  6.4985e-02  0.7165  0.47406    
factor(bedrooms)2-bedroom  3.1126e-02  4.8262e-02  0.6449  0.51928    
price                     -3.8356e-06  3.0690e-05 -0.1250  0.90059    
treatment:genderFemale    -9.3189e-02  9.6840e-02 -0.9623  0.33639    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 3a-1 - Treatment & gender + all covariates including less reliable sqft and pro
fessoinal
m3.1 <- lm(outcome~treatment*gender+factor(city)+factor(bedrooms)+price+sqft+profession
al,data=d)
stargazer(m3.1,type='text')
```

```
=====================================================
                              Dependent variable:
                          ---------------------------
                                    outcome
-----------------------------------------------------
treatment                            0.045
                                    (0.068)

genderFemale                         0.103
                                    (0.078)

factor(city)houston                -0.175***
                                    (0.066)

factor(city)sandiego                -0.029
                                    (0.068)

factor(city)seattle                  0.051
                                    (0.069)

factor(bedrooms)2-bedroom           -0.0004
                                    (0.054)

price                              -0.00003
                                   (0.00003)

sqft                                0.0002
                                   (0.0001)

professionalProfessional            0.125**
                                    (0.062)

treatment:genderFemale              -0.092
                                    (0.096)

Constant                             0.205
                                    (0.144)

-----------------------------------------------------
Observations                          483
R2                                   0.038
Adjusted R2                          0.018
Residual Std. Error            0.494 (df = 472)
F Statistic                 1.864** (df = 10; 472)
=====================================================
Note:                     *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m3.1, vcovHC(m3.1)) # Robust se
```

```
t test of coefficients:

                            Estimate  Std. Error t value Pr(>|t|)
(Intercept)               2.0546e-01  1.4527e-01  1.4143 0.157936
treatment                 4.4953e-02  6.8155e-02  0.6596 0.509857
genderFemale              1.0259e-01  7.9008e-02  1.2985 0.194752
factor(city)houston      -1.7456e-01  6.6128e-02 -2.6397 0.008572 **
factor(city)sandiego     -2.9036e-02  6.9619e-02 -0.4171 0.676819
factor(city)seattle       5.0883e-02  7.0620e-02  0.7205 0.471566
factor(bedrooms)2-bedroom -4.2180e-04  5.2647e-02 -0.0080 0.993611
price                    -3.0302e-05  3.3411e-05 -0.9069 0.364902
sqft                      1.9973e-04  1.2202e-04  1.6368 0.102329
professionalProfessional  1.2542e-01  6.3676e-02  1.9696 0.049465 *
treatment:genderFemale   -9.1867e-02  9.6821e-02 -0.9488 0.343192
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Next we treat treatment as a categorical variable (effect might not be linear)
# Model 1b - Basic model
m4 <- lm(outcome~treatment_f,data=d)
stargazer(m4,type='text')
```

```
===============================================
                        Dependent variable:
                    ---------------------------
                              outcome
-----------------------------------------------
treatment_fLow                -0.003
                              (0.056)

treatment_fHigh                0.013
                              (0.056)

Constant                     0.450***
                              (0.039)

-----------------------------------------------
Observations                    483
R2                            0.0002
Adjusted R2                   -0.004
Residual Std. Error      0.499 (df = 480)
F Statistic            0.046 (df = 2; 480)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m4, vcovHC(m4)) # Robust se
```

```
t test of coefficients:

                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.450000   0.039578 11.3700   <2e-16 ***
treatment_fLow  -0.002795   0.055867 -0.0500   0.9601    
treatment_fHigh  0.012963   0.055859  0.2321   0.8166    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 2b - Treatment & gender
m5 <- lm(outcome~treatment_f*gender,data=d)
stargazer(m5,type='text')
```

```
=========================================================
                                 Dependent variable:    
                             ---------------------------
                                       outcome          
---------------------------------------------------------
treatment_fLow                          0.071           
                                       (0.080)          
                                                        
treatment_fHigh                         0.030           
                                       (0.078)          
                                                        
genderFemale                            0.098           
                                       (0.079)          
                                                        
treatment_fLow:genderFemale            -0.148           
                                       (0.112)          
                                                        
treatment_fHigh:genderFemale           -0.036           
                                       (0.111)          
                                                        
Constant                              0.402***          
                                       (0.055)          
                                                        
---------------------------------------------------------
Observations                             483            
R2                                      0.006           
Adjusted R2                            -0.005           
Residual Std. Error               0.500 (df = 477)      
F Statistic                     0.528 (df = 5; 477)     
=========================================================
Note:                         *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m5, vcovHC(m5)) # Robust se
```

```
t test of coefficients:

                              Estimate Std. Error t value  Pr(>|t|)    
(Intercept)                   0.402439   0.054823  7.3407 9.234e-13 ***
treatment_fLow                0.071245   0.079837  0.8924    0.3726    
treatment_fHigh               0.029660   0.078175  0.3794    0.7046    
genderFemale                  0.097561   0.079338  1.2297    0.2194    
treatment_fLow:genderFemale  -0.147716   0.112268 -1.3157    0.1889    
treatment_fHigh:genderFemale -0.035833   0.112088 -0.3197    0.7493    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 3b - Treatment & gender + covariates
m6 <- lm(outcome~treatment_f*gender+factor(city)+factor(bedrooms)+price,data=d)
stargazer(m6,type='text')
```

```

========================================================
                                 Dependent variable:
                             ---------------------------
                                       outcome
--------------------------------------------------------
treatment_fLow                          0.074
                                       (0.079)

treatment_fHigh                         0.031
                                       (0.078)

genderFemale                            0.097
                                       (0.079)

factor(city)houston                   -0.157**
                                       (0.064)

factor(city)sandiego                   -0.047
                                       (0.064)

factor(city)seattle                     0.044
                                       (0.064)

factor(bedrooms)2-bedroom               0.032
                                       (0.049)

price                                 -0.00000
                                      (0.00003)

treatment_fLow:genderFemale            -0.151
                                       (0.111)

treatment_fHigh:genderFemale           -0.035
                                       (0.111)

Constant                              0.433***
                                       (0.091)

--------------------------------------------------------
Observations                             483
R2                                      0.029
Adjusted R2                             0.008
Residual Std. Error               0.496 (df = 472)
F Statistic                     1.399 (df = 10; 472)
========================================================
Note:                        *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m6, vcovHC(m6)) # Robust se
```

```

t test of coefficients:

                                Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)                   4.3327e-01  9.3187e-02  4.6495 4.325e-06 ***
treatment_fLow                7.3544e-02  7.9784e-02  0.9218    0.3571    
treatment_fHigh               3.0518e-02  7.8207e-02  0.3902    0.6966    
genderFemale                  9.6823e-02  7.9161e-02  1.2231    0.2219    
factor(city)houston          -1.5723e-01  6.3812e-02 -2.4639    0.0141 *  
factor(city)sandiego         -4.6826e-02  6.5934e-02 -0.7102    0.4779    
factor(city)seattle           4.4475e-02  6.5228e-02  0.6818    0.4957    
factor(bedrooms)2-bedroom     3.2102e-02  4.8609e-02  0.6604    0.5093    
price                        -4.1162e-06  3.0934e-05 -0.1331    0.8942    
treatment_fLow:genderFemale  -1.5084e-01  1.1212e-01 -1.3453    0.1792    
treatment_fHigh:genderFemale -3.5359e-02  1.1210e-01 -0.3154    0.7526    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 3b-1 - Treatment & gender + all covariates including less reliable sqft and pro
fessoinal
m6.1 <- lm(outcome~treatment_f*gender+factor(city)+factor(bedrooms)+price+sqft+professi
onal,data=d)
stargazer(m6.1,type='text')
```

```
========================================================
                                 Dependent variable:    
                             ---------------------------
                                       outcome          
--------------------------------------------------------
treatment_fLow                          0.063           
                                       (0.079)          
                                                        
treatment_fHigh                         0.028           
                                       (0.078)          
                                                        
genderFemale                            0.102           
                                       (0.078)          
                                                        
factor(city)houston                   -0.177***         
                                       (0.066)          
                                                        
factor(city)sandiego                   -0.032           
                                       (0.068)          
                                                        
factor(city)seattle                     0.047           
                                       (0.069)          
                                                        
factor(bedrooms)2-bedroom               0.002           
                                       (0.054)          
                                                        
price                                 -0.00003          
                                      (0.00003)         
                                                        
sqft                                   0.0002           
                                      (0.0001)          
                                                        
professionalProfessional               0.128**          
                                       (0.063)          
                                                        
treatment_fLow:genderFemale            -0.147           
                                       (0.111)          
                                                        
treatment_fHigh:genderFemale           -0.035           
                                       (0.111)          
                                                        
Constant                                0.212           
                                       (0.144)          
                                                        
--------------------------------------------------------
Observations                             483            
R2                                      0.040           
Adjusted R2                             0.016           
Residual Std. Error               0.494 (df = 470)      
F Statistic                     1.649* (df = 12; 470)   
========================================================
Note:                        *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m6.1, vcovHC(m6.1)) # Robust se
```

```


t test of coefficients:

                                Estimate  Std. Error t value Pr(>|t|)
(Intercept)                   2.1190e-01  1.4458e-01  1.4656 0.143427
treatment_fLow                6.2881e-02  8.0392e-02  0.7822 0.434503
treatment_fHigh               2.7746e-02  7.8568e-02  0.3531 0.724137
genderFemale                  1.0238e-01  7.9039e-02  1.2953 0.195852
factor(city)houston          -1.7745e-01  6.6167e-02 -2.6819 0.007579 **
factor(city)sandiego         -3.2071e-02  6.9557e-02 -0.4611 0.644961
factor(city)seattle           4.6910e-02  7.0773e-02  0.6628 0.507773
factor(bedrooms)2-bedroom     2.2089e-03  5.2914e-02  0.0417 0.966719
price                        -3.0535e-05  3.3802e-05 -0.9033 0.366804
sqft                          1.9285e-04  1.2171e-04  1.5845 0.113755
professionalProfessional      1.2789e-01  6.3882e-02  2.0020 0.045859 *
treatment_fLow:genderFemale  -1.4664e-01  1.1200e-01 -1.3093 0.191081
treatment_fHigh:genderFemale -3.5093e-02  1.1223e-01 -0.3127 0.754666
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In all models, the coefficients on treatment, whether continous or as a factor, are not statistically significant. If we add gender, there is also no evidence of a the interaction term being statistically significant. Thus, there is no evidence that exclamation points have influenced the likelihood of receiving a response.

Hide

```
# We try an alternative specification for treatment (as dummy variables)
# Model 1c - Basic model
m7 <- lm(outcome ~ low_treatment + high_treatment, data=d)

stargazer(m7, type='text')
```

```

===============================================
                        Dependent variable:    
                    ---------------------------
                              outcome          
-----------------------------------------------
low_treatment                 -0.003           
                              (0.056)          
                                               
high_treatment                 0.013           
                              (0.056)          
                                               
Constant                     0.450***          
                              (0.039)          
                                               
-----------------------------------------------
Observations                    483            
R2                            0.0002           
Adjusted R2                   -0.004           
Residual Std. Error      0.499 (df = 480)      
F Statistic             0.046 (df = 2; 480)    
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m7, vcovHC(m7)) # Robust se
```

```

t test of coefficients:

                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     0.450000   0.039578 11.3700   <2e-16 ***
low_treatment  -0.002795   0.055867 -0.0500   0.9601    
high_treatment  0.012963   0.055859  0.2321   0.8166    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 2c - Treatment & gender
m8 <- lm(outcome ~ low_treatment + high_treatment*gender, data=d)
stargazer(m8, type='text')
```

```

=======================================================
                                Dependent variable:    
                            ---------------------------
                                      outcome          
-------------------------------------------------------
low_treatment                         -0.004           
                                      (0.056)          
                                                       
high_treatment                        -0.006           
                                      (0.073)          
                                                       
genderFemale                           0.024           
                                      (0.056)          
                                                       
high_treatment:genderFemale            0.038           
                                      (0.096)          
                                                       
Constant                             0.439***          
                                      (0.048)          
                                                       
-------------------------------------------------------
Observations                            483            
R2                                     0.002           
Adjusted R2                           -0.007           
Residual Std. Error              0.500 (df = 478)      
F Statistic                     0.222 (df = 4; 478)    
=======================================================
Note:                       *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m8, vcovHC(m8)) # Robust se
```

```


t test of coefficients:

                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  0.4385118  0.0478301  9.1681   <2e-16 ***
low_treatment               -0.0037483  0.0561084 -0.0668   0.9468    
high_treatment              -0.0064131  0.0734399 -0.0873   0.9305    
genderFemale                 0.0235655  0.0561118  0.4200   0.6747    
high_treatment:genderFemale  0.0381629  0.0970455  0.3932   0.6943    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# Model 3c - Treatment & gender + covariates
m9 <- lm(outcome ~ low_treatment + high_treatment + gender + factor(city) + factor(bedr
ooms) + price, data=d)
stargazer(m9, type='text')
```

```
=====================================================
                              Dependent variable:
                          ---------------------------
                                    outcome
-----------------------------------------------------
low_treatment                       -0.004
                                    (0.056)

high_treatment                       0.014
                                    (0.055)

genderFemale                         0.035
                                    (0.045)

factor(city)houston                -0.154**
                                    (0.064)

factor(city)sandiego                -0.046
                                    (0.064)

factor(city)seattle                  0.048
                                    (0.064)

factor(bedrooms)2-bedroom            0.031
                                    (0.049)

price                              -0.00001
                                   (0.00003)

Constant                           0.466***
                                    (0.084)

-----------------------------------------------------
Observations                          483
R2                                   0.025
Adjusted R2                          0.008
Residual Std. Error             0.496 (df = 474)
F Statistic                   1.497 (df = 8; 474)
=====================================================
Note:                     *p<0.1; **p<0.05; ***p<0.01
```

Hide

```
coeftest(m9, vcovHC(m9)) # Robust se
```

```

t test of coefficients:

                             Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)                4.6571e-01  8.6773e-02  5.3669 1.256e-07 ***
low_treatment             -3.7784e-03  5.5953e-02 -0.0675   0.94619    
high_treatment             1.3738e-02  5.5896e-02  0.2458   0.80596    
genderFemale               3.4752e-02  4.5695e-02  0.7605   0.44732    
factor(city)houston       -1.5429e-01  6.3676e-02 -2.4230   0.01577 *  
factor(city)sandiego      -4.5755e-02  6.5682e-02 -0.6966   0.48638    
factor(city)seattle        4.7974e-02  6.4963e-02  0.7385   0.46059    
factor(bedrooms)2-bedroom  3.0814e-02  4.8250e-02  0.6386   0.52337    
price                     -5.6698e-06  3.0873e-05 -0.1836   0.85437    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The coefficients on treatment are also statistically insignificant. There is no evidence that exclamation points have an effect.

Randomization Inference

Next we use randomization inference (assuming a Sharp Null of No Effect) to understand if our observation is consistent with an empirical null distribution. For this, we combine low and high treatment into treatment (since we have not learned more complex fixes for heterogenous effects).

Hide

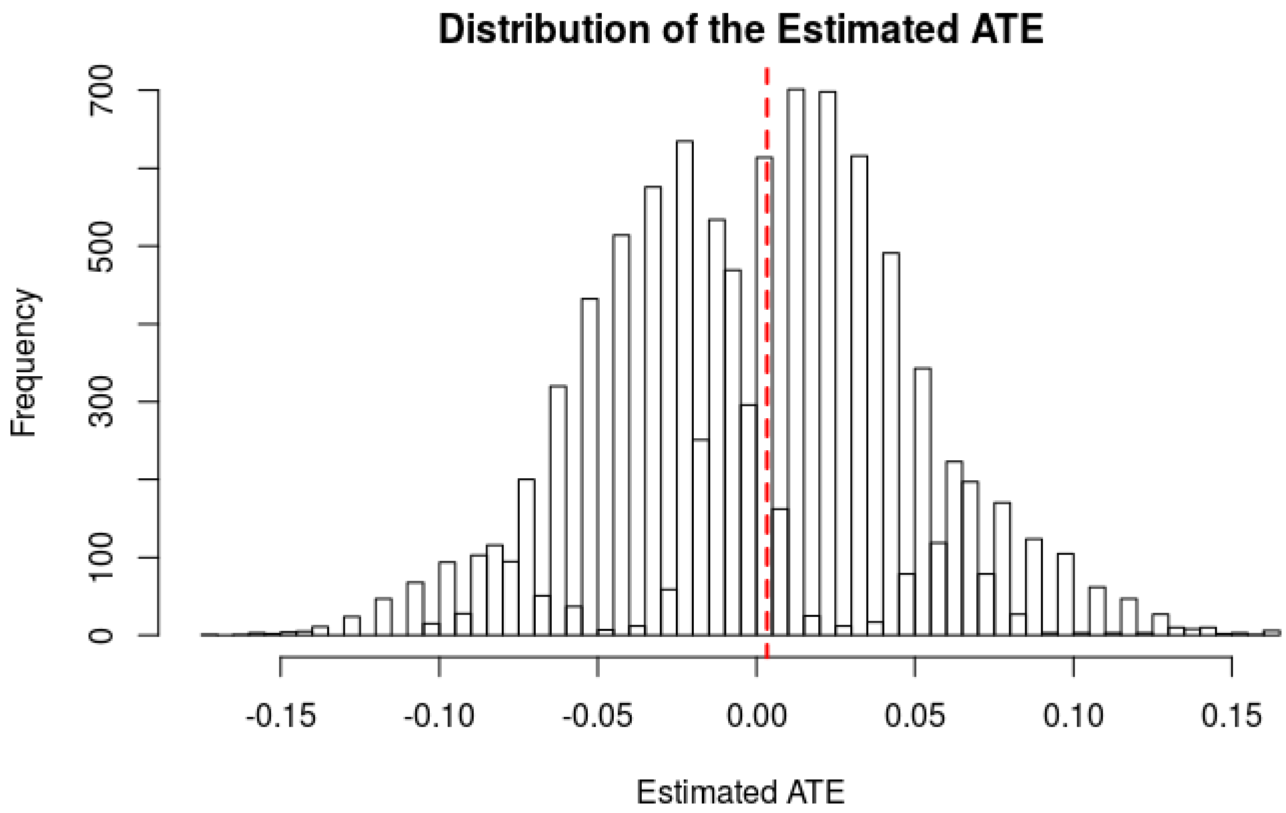
```
# Combining treatments
di <- d
di[treatment==2,treatment:=1]
# Define distributions
y <- di$outcome
Z <- di$treatment
blk1 <- as.numeric(di$gender) # We block by gender
blk2 <- as.numeric(di$city) # Block by city
blk3 <- as.numeric(di$bedrooms)
# By gender
perms <- genperms(Z, clustvar = NULL, blockvar = blk1)
```

```
Too many permutations to use exact method.
Defaulting to approximate method.
Increase maxiter to at least 4.19386819554668e+130 to perform exact estimation.
```

Hide

```
probs <- genprobexact(Z, clustvar = NULL, blockvar = blk1) # probability of treatment
ate <- estate(y,Z,prob=probs) # estimate the ATE
Ys <- genouts(y,Z,ate=0) # generate potential outcomes under sharp null of no effect
distout <- gendist(Ys,perms, prob=probs) # generate sampling dist. under sharp null
dispdist(distout, ate, quantiles = c(0.025, 0.975), display.plot = TRUE) # display char
acteristics of sampling dist. for inference
```

```
$two.tailed.p.value
[1] 0.9748

$two.tailed.p.value.abs
[1] 0.9874

$greater.p.value
[1] 0.4874

$lesser.p.value
[1] 0.5192

$quantile
       2.5%       97.5%
-0.09838053  0.09753228

$sd
[1] 0.0484482

$exp.val
[1] -0.0004189318
```



Hide

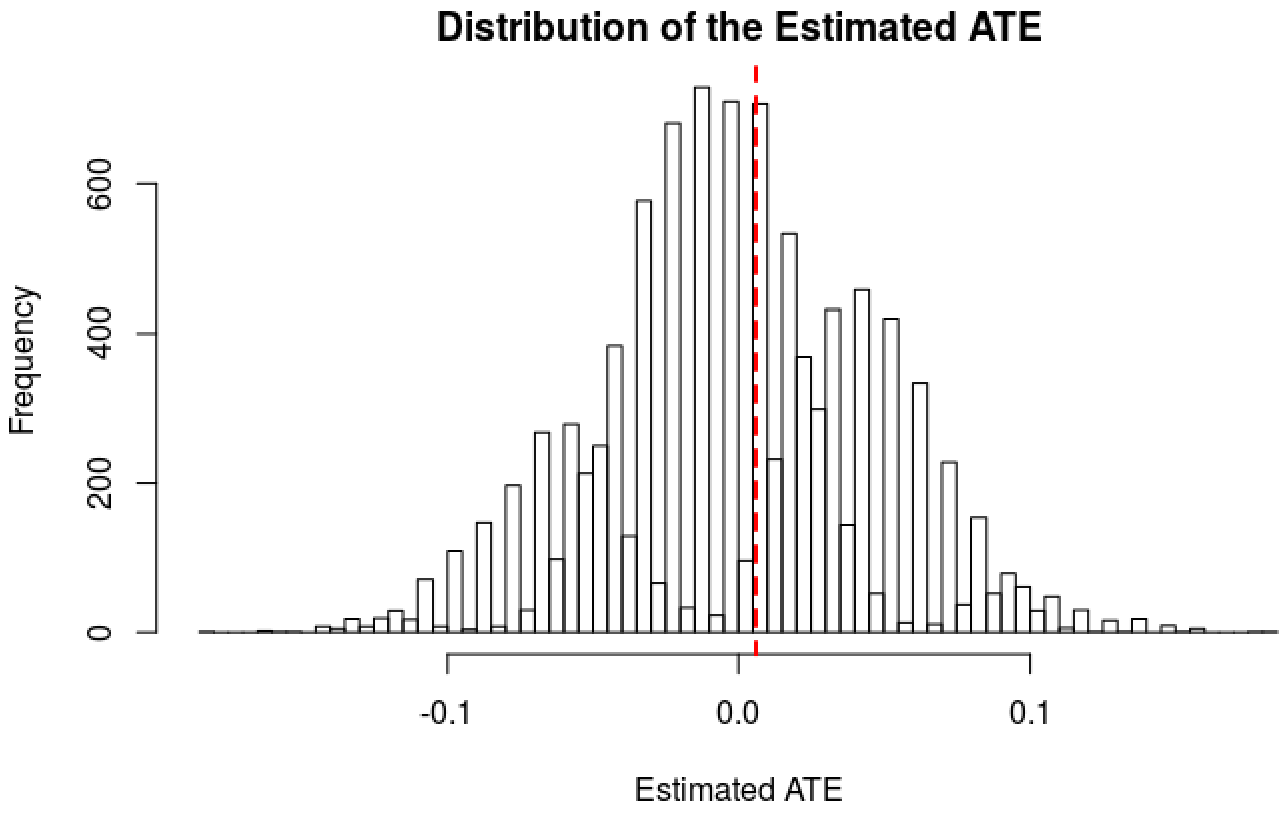
```
# By city
perms <- genperms(Z, clustvar = NULL, blockvar = blk2)
```

```
Too many permutations to use exact method.
Defaulting to approximate method.
Increase maxiter to at least 4.59676200506436e+128 to perform exact estimation.
```

Hide

```
probs <- genprobexact(Z, clustvar = NULL, blockvar = blk2) # probability of treatment
ate <- estate(y,Z,prob=probs) # estimate the ATE
Ys <- genouts(y,Z,ate=0) # generate potential outcomes under sharp null of no effect
distout <- gendist(Ys,perms, prob=probs) # generate sampling dist. under sharp null
dispdist(distout, ate, quantiles = c(0.025, 0.975), display.plot = TRUE) # display characteristics of sampling dist. for inference
```

```
$two.tailed.p.value
[1] 0.9064

$two.tailed.p.value.abs
[1] 0.893

$greater.p.value
[1] 0.4532

$lesser.p.value
[1] 0.5475

$quantile
       2.5%       97.5% 
-0.09654754  0.09120148 

$sd
[1] 0.04817207

$exp.val
[1] -0.0006351514
```



Hide

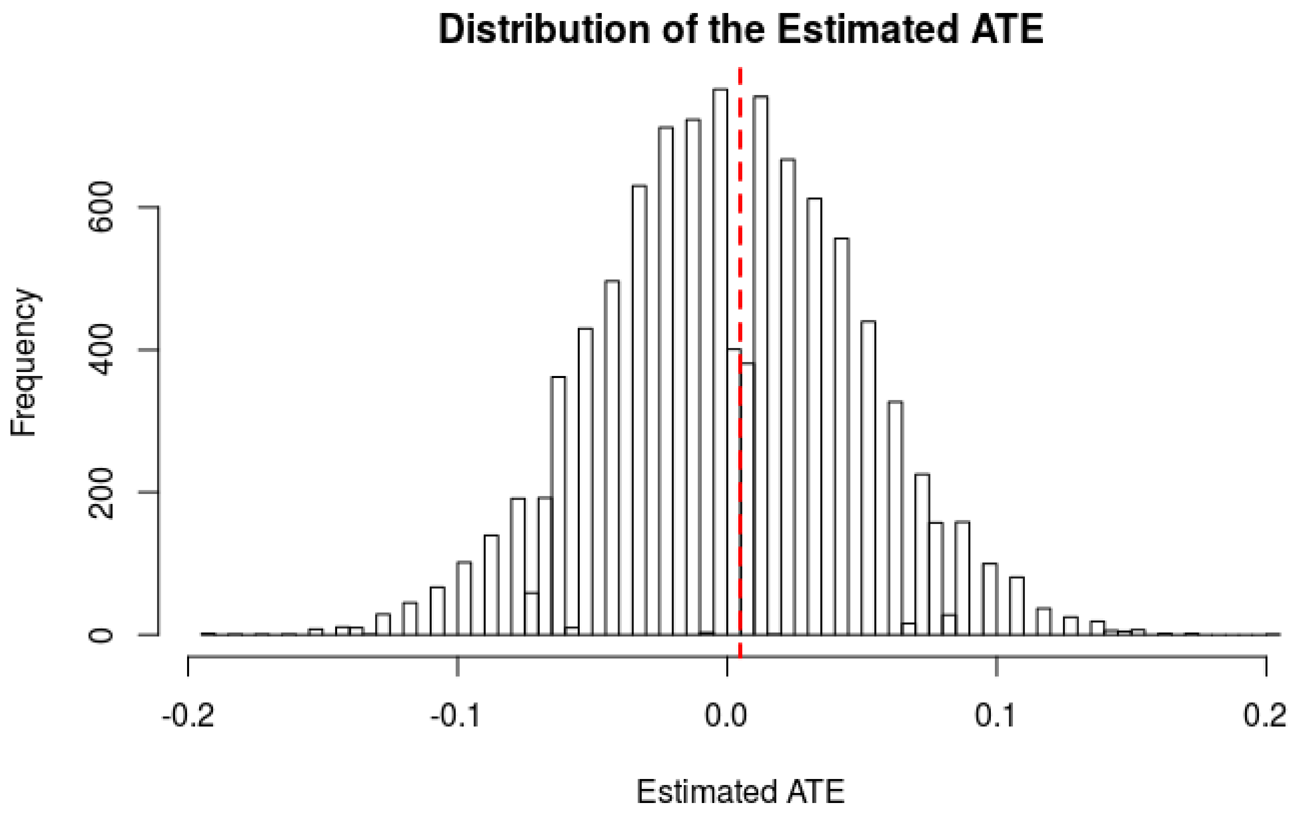
```
# By bedroom
perms <- genperms(Z, clustvar = NULL, blockvar = blk3)
```

```
Too many permutations to use exact method.
Defaulting to approximate method.
Increase maxiter to at least 4.80931839467142e+130 to perform exact estimation.
```

Hide

```
probs <- genprobexact(Z, clustvar = NULL, blockvar = blk3) # probability of treatment
ate <- estate(y,Z,prob=probs) # estimate the ATE
Ys <- genouts(y,Z,ate=0) # generate potential outcomes under sharp null of no effect
distout <- gendist(Ys,perms, prob=probs) # generate sampling dist. under sharp null
dispdist(distout, ate, quantiles = c(0.025, 0.975), display.plot = TRUE) # display char
acteristics of sampling dist. for inference
```

```
$two.tailed.p.value
[1] 0.9632

$two.tailed.p.value.abs
[1] 0.9082

$greater.p.value
[1] 0.4816

$lesser.p.value
[1] 0.528

$quantile
       2.5%       97.5% 
-0.09763483  0.09827367 

$sd
[1] 0.04858612

$exp.val
[1] 0.0003082015
```



Hide

```
#P-value for actual data
p.val.actual = sum(abs(distout) > ate) / length(distout)
p.val.actual
```

```
[1] 0.8986
```

Hide

```
#get respnse rate by treatment or control
actual.response.rate.by.treatment <- di[, mean(outcome), by = c("treatment")]
actual.response.rate.by.treatment
```

treatment	V1
<dbl>	<dbl>
1	0.4551084
0	0.4500000

2 rows

Hide

```
di[, sum(outcome > -100), by = c("treatment")]
```

treatment	V1
<dbl>	<int>
1	323
0	160

2 rows

Once again, we cannot reject the null hypothesis of no effect.

Other Analysis

1. Although not a signficant issue for this experiement we estimate the CACE. For this we define non-compliers as those for who we sent emails but did not received them - and we know this because we received a “bounced” email message.

Hide

```
# We calculate the CACE manually
# Manually compute CACE
itt <- mean(d$outcome[d$treatment != 0]) - mean(d$outcome[d$treatment == 0])
prop_treated <- 481/483
sprintf("\nThe estimated CACE is: %.5f", itt / prop_treated)
```

```
[1] "\nThe estimated CACE is: 0.00513"
```

Hide

```
# or 2SLS
#itt_fit <- ivreg(outcome ~treatment,~compliers,data=d)
#stargazer(itt_fit, type='text')
```

2. We also did some work Work to find treatment response rate required to reject null. That is, how much does the

Hide

```
# First, we use RI. create a new temp column of outcomes where the share of responses i
s n%
#To see what treatment response rate is required for significant result, adjust this va
riable.
#Found that a treatment response rate of about 0.6 would be required to observe signifi
cant result
treatment.response.rate <- 0.6
di$hypothetical.outcomes.temp <- sample(c(0,1), size = nrow(d), replace = TRUE, prob =
c(1-treatment.response.rate, treatment.response.rate))
#create new outcome column that takes original outcomes for control group, but new hypo
thetical outcomes with adjusted response rate for treatment rows
di$hypothetical.outcomes = d$outcome
di[treatment==1, hypothetical.outcomes:=hypothetical.outcomes.temp]
```

```
Coerced 'double' RHS to 'integer' to match the column's type; may have truncated precis
ion. Either change the target column to 'double' first (by creating a new 'double' vect
or length 483 (nrows of entire table) and assign that; i.e. 'replace' column), or coerc
e RHS to 'integer' (e.g. 1L, NA_[real|integer]_, as.*, etc) to make your intent clear a
nd for speed. Or, set the column type correctly up front when you create the table and
 stick to it, please.
```

Hide

```
fake.response.rate.by.treatment <- di[, mean(hypothetical.outcomes), by =
c("treatment")]
fake.response.rate.by.treatment
```

treatment <dbl>	V1 <dbl>
1	0.6037152
0	0.4500000

2 rows

Hide

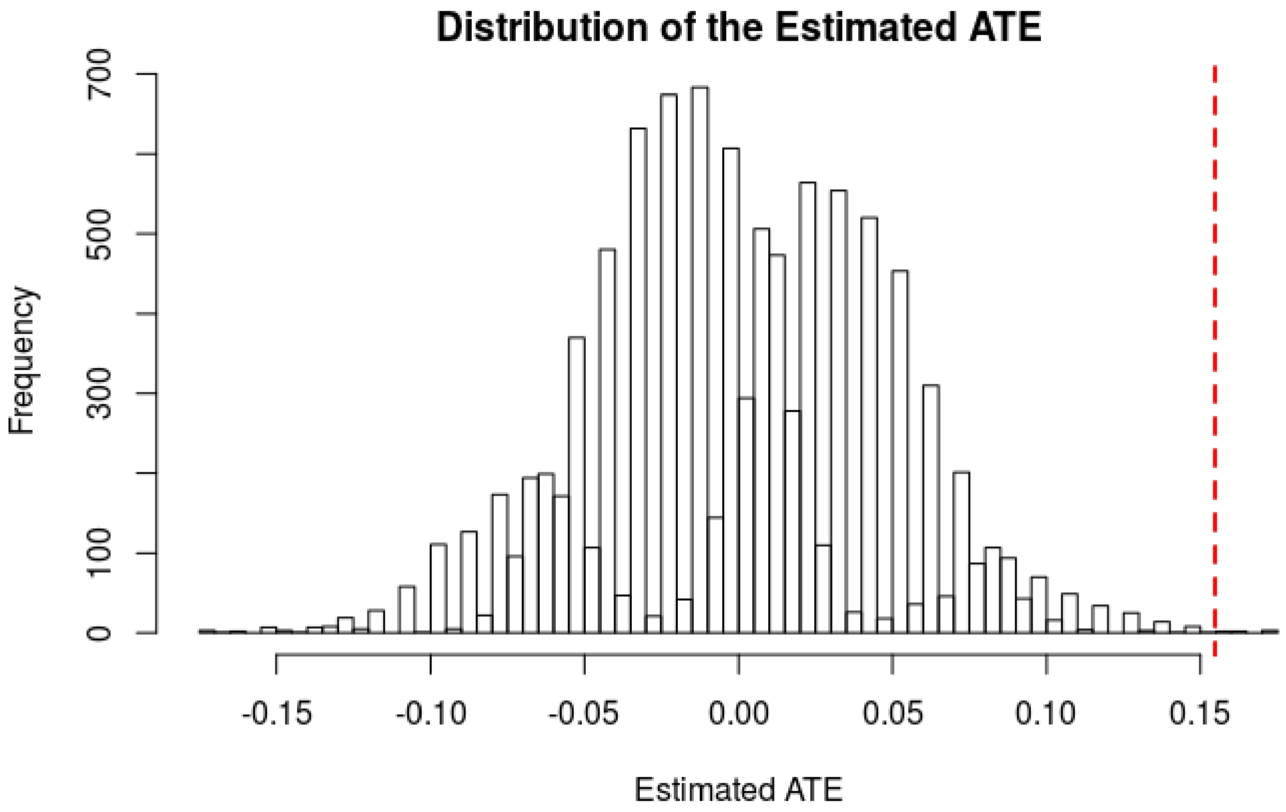
```
#run RI using the fake data
y.fake <- di$hypothetical.outcomes
Z.fake <- di$treatment
cls.fake <- di$gender
blk.fake <- di$city
perms.fake <- genperms(Z.fake, clustvar = NULL, blockvar = blk.fake)
```

```
Too many permutations to use exact method.
Defaulting to approximate method.
Increase maxiter to at least 4.59676200506436e+128 to perform exact estimation.
```

Hide

```
probs.fake <- genprobexact(Z.fake, clustvar = NULL, blockvar = blk.fake) # probability
 of treatment
ate.fake <- estate(y.fake,Z.fake,prob=probs.fake) # estimate the ATE
Ys.fake <- genouts(y.fake,Z.fake,ate=0) # generate potential outcomes under sharp null
 of no effect
distout.fake <- gendist(Ys.fake,perms.fake, prob=probs.fake) # generate sampling dist.
 under sharp null
dispdist(distout.fake, ate.fake, quantiles = c(0.025, 0.975), display.plot = TRUE) # di
splay characteristics of sampling dist. for inference
```

```
$two.tailed.p.value
[1] 0.0014

$two.tailed.p.value.abs
[1] 0.0013

$greater.p.value
[1] 7e-04

$lesser.p.value
[1] 0.9993

$quantile
       2.5%       97.5%
-0.09569790  0.09060323

$sd
[1] 0.04780999

$exp.val
[1] -0.0001980287
```



Hide

```
#P-value for actual data
p.val.fake = sum(abs(distout.fake) > ate.fake) / length(distout.fake)
p.val.fake
```

```
[1] 0.0013
```

Hide

```
# We can also do this using the regression estimate. For this we use the more general m
odel, m1. Given this model, if we would want to see treatment be statistically signific
ant we either have a larger coefficient or a lower standard error. Choosing a larger n
 might be one way to reduce the standard error.
z <- coeftest(m1, vcovHC(m1)) # Robust se
# We want Est/Stderr > 2
t.stderr <- z[2]/2
std <- z[4]*sqrt(483)
newn <-(std/t.stderr)^2
sprintf("\nTo attain enough power, i.e. to drive the standard error small enough (all t
hings unchanged), we would need a sample size of %.0f", newn)
```

```
[1] "\nTo attain enough power, i.e. to drive the standard error small enough (all thing
s unchanged), we would need a sample size of 35590"
```

Hide

```
# Conversely we can also estimate the required difference in coefficient:
newest <- 2*z[4]
sprintf("\nWe need to see an effect of greater than: %.3f", newest)
```

```
[1] "\nWe need to see an effect of greater than: 0.056"
```

Hide

```
sprintf("Which is about %0.f times more than what we see currently", 2/z[6])
```

```
[1] "Which is about 9 times more than what we see currently"
```

4. Conclusion

Despite running a few different models, we find no evidencce that the number of exclamation points affected response rates to our email.