W241 Class Project - Analysis

Code

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 <int>	city <fctr>	title <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' | trea
tment_assignment=='Jane_Treat_Low',gender:=1]
d[treatment_assignment=='John_Control' | treatment_assignment=='John_Treat_High' | trea
tment_assignment=='John_Treat_Low',gender:=0]
# Set treatment variable = 0 for control, 1 for low, 2 for high (treatment here is cont
inuous)
d[treatment_assignment=='Jane_Control' | treatment_assignment=='John_Control', treatmen
t:=0]
d[treatment_assignment=='Jane_Treat_Low' | treatment_assignment=='John_Treat_Low', trea
tment:=1]
d[treatment_assignment=='Jane_Treat_High' | treatment_assignment=='John_Treat_High', tr
eatment:=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', hi
gh_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 <int>	**city** <fctr>	**posting_date** <fctr>	**bedrooms** <fctr>	**sqft** <int>	**price** <int>	**treatment_assignment** <fctr>	**professional** <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
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
        256         207
```

Hide

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

```


Table of Outcomes (By Gender):
```

Hide

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

```

              Male Female
  No Response  131    125
  Response     100    107
```

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      87  87   82
  Response         68  68   71
```

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              38             48           48              44
           39
  Response              38              37             32           30              34
           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      78       65      50
  Response         61      38       51      57
```

Hide

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

```

Table of Outcomes (By Rooms):
```

Hide

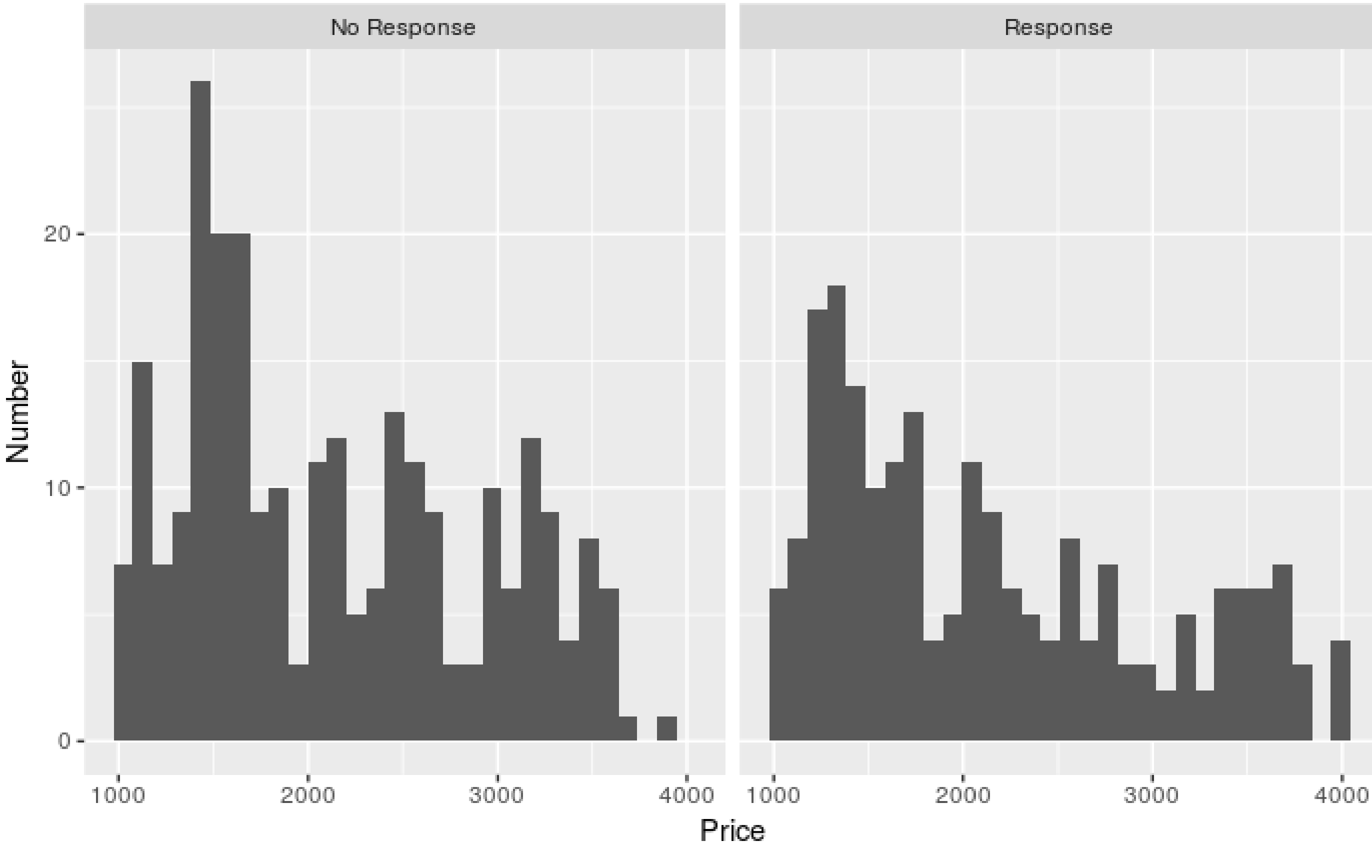
```
table(d$outcome_f,factor(d$bedrooms))
```

```

              1-bedroom 2-bedroom
  No Response       125       131
  Response           96       111
```

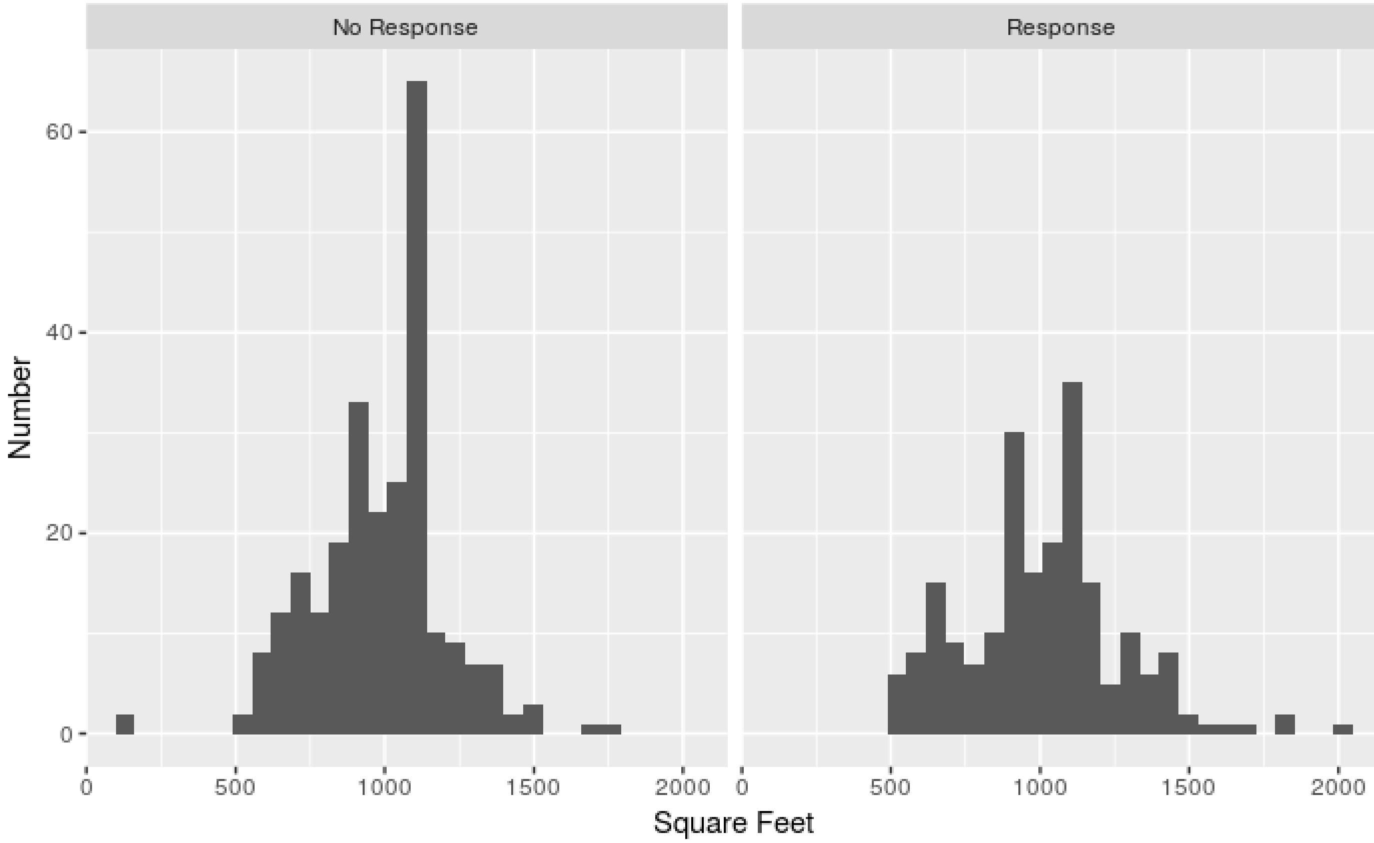
Hide

```
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 t
his 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="N
umber")
```



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          201
  Response                  35          172
```

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  131    125
  Response     100    107
```

Hide

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

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

data:  tbl
X-squared = 0.26941, df = 1, p-value = 0.6037
```

Hide

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

```
               0  1  2
  No Response 87 87 82
  Response    68 68 71
```

Hide

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

```
	Pearson's Chi-squared test

data:  tbl
X-squared = 0.26615, df = 2, p-value = 0.8754
```

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              38             48           48              44
            39
  Response              38              37             32           30              34
            36
```

Hide

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

```
    Pearson's Chi-squared test

data:  tbl
X-squared = 3.6372, df = 5, p-value = 0.6027
```

Hide

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

```
             Non-professional Professional
  No Response               55          201
  Response                  35          172
```

Hide

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

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

data:  tbl
X-squared = 1.2523, df = 1, p-value = 0.2631
```

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                      0.013
                              (0.028)

Constant                     0.434***
                              (0.037)

-----------------------------------------------
Observations                    463
R2                            0.0004
Adjusted R2                   -0.002
Residual Std. Error      0.498 (df = 461)
F Statistic             0.198 (df = 1; 461)
===============================================
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) 0.434495   0.036582 11.8771   <2e-16 ***
treatment   0.012644   0.028502  0.4436   0.6575
---
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                       0.026
                               (0.040)

genderFemale                    0.054
                               (0.073)

treatment:genderFemale         -0.026
                               (0.057)

Constant                       0.407***
                               (0.052)

--------------------------------------------------
Observations                     463
R2                              0.002
Adjusted R2                    -0.005
Residual Std. Error        0.499 (df = 459)
F Statistic              0.262 (df = 3; 459)
==================================================
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)             0.407259   0.051210  7.9527 1.434e-14 ***
treatment               0.025641   0.039798  0.6443    0.5197
genderFemale            0.054454   0.073288  0.7430    0.4579
treatment:genderFemale -0.026152   0.057200 -0.4572    0.6477
---
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                            0.027
                                    (0.040)

genderFemale                         0.052
                                    (0.073)

factor(city)houston                -0.164**
                                    (0.064)

factor(city)sandiego                -0.051
                                    (0.064)

factor(city)seattle                  0.042
                                    (0.066)

factor(bedrooms)2-bedroom            0.027
                                    (0.050)

price                              -0.00000
                                   (0.00003)

treatment:genderFemale              -0.025
                                    (0.057)

Constant                           0.441***
                                    (0.088)

-------------------------------------------------------
Observations                          463
R2                                   0.026
Adjusted R2                          0.009
Residual Std. Error            0.496 (df = 454)
F Statistic                  1.516 (df = 8; 454)
=======================================================
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)                4.4134e-01  9.0292e-02  4.8880 1.416e-06 ***
treatment                  2.7172e-02  3.9686e-02  0.6847   0.49389    
genderFemale               5.2129e-02  7.3091e-02  0.7132   0.47609    
factor(city)houston       -1.6386e-01  6.3923e-02 -2.5633   0.01069 *  
factor(city)sandiego      -5.1371e-02  6.5858e-02 -0.7800   0.43578    
factor(city)seattle        4.2490e-02  6.7313e-02  0.6312   0.52821    
factor(bedrooms)2-bedroom  2.7215e-02  4.9288e-02  0.5522   0.58111    
price                     -2.4420e-06  3.1488e-05 -0.0776   0.93822    
treatment:genderFemale    -2.4833e-02  5.7221e-02 -0.4340   0.66451    
---
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                            0.026
                                    (0.040)

genderFemale                         0.059
                                    (0.073)

factor(city)houston                -0.181***
                                    (0.067)

factor(city)sandiego                -0.031
                                    (0.068)

factor(city)seattle                  0.051
                                    (0.071)

factor(bedrooms)2-bedroom           -0.009
                                    (0.055)

price                              -0.00003
                                   (0.00003)

sqft                                0.0002*
                                   (0.0001)

professionalProfessional            0.124**
                                    (0.062)

treatment:genderFemale              -0.027
                                    (0.057)

Constant                             0.193
                                    (0.144)

-----------------------------------------------------
Observations                          463
R2                                   0.039
Adjusted R2                          0.018
Residual Std. Error            0.493 (df = 452)
F Statistic                  1.832* (df = 10; 452)
=====================================================
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)               1.9334e-01  1.4554e-01  1.3285  0.18470
treatment                 2.6110e-02  3.9819e-02  0.6557  0.51235
genderFemale              5.9426e-02  7.2995e-02  0.8141  0.41601
factor(city)houston      -1.8053e-01  6.6516e-02 -2.7141  0.00690 **
factor(city)sandiego     -3.0741e-02  6.9931e-02 -0.4396  0.66044
factor(city)seattle       5.0970e-02  7.2326e-02  0.7047  0.48134
factor(bedrooms)2-bedroom -8.6209e-03  5.3405e-02 -0.1614  0.87183
price                    -3.0499e-05  3.4144e-05 -0.8932  0.37220
sqft                      2.2319e-04  1.2288e-04  1.8164  0.06998 .
professionalProfessional  1.2359e-01  6.3672e-02  1.9410  0.05288 .
treatment:genderFemale   -2.6580e-02  5.7256e-02 -0.4642  0.64271
---
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                 0.000
                              (0.057)

treatment_fHigh                0.025
                              (0.057)

Constant                      0.439***
                              (0.040)

-----------------------------------------------
Observations                    463
R2                             0.001
Adjusted R2                   -0.004
Residual Std. Error      0.499 (df = 460)
F Statistic            0.132 (df = 2; 460)
===============================================
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)     4.3871e-01 4.0117e-02 10.9358   <2e-16 ***
treatment_fLow  2.4600e-16 5.6734e-02  0.0000   1.0000
treatment_fHigh 2.5343e-02 5.7065e-02  0.4441   0.6572
---
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                          0.095
                                       (0.081)

treatment_fHigh                         0.051
                                       (0.080)

genderFemale                            0.109
                                       (0.080)

treatment_fLow:genderFemale            -0.189*
                                       (0.113)

treatment_fHigh:genderFemale           -0.051
                                       (0.114)

Constant                              0.385***
                                       (0.056)

---------------------------------------------------------
Observations                             463
R2                                      0.008
Adjusted R2                            -0.003
Residual Std. Error               0.498 (df = 457)
F Statistic                     0.724 (df = 5; 457)
=========================================================
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)                   0.384615   0.055801  6.8926 1.827e-11 ***
treatment_fLow                0.095385   0.080823  1.1802   0.23855    
treatment_fHigh               0.051282   0.079678  0.6436   0.52015    
genderFemale                  0.108891   0.080287  1.3563   0.17568    
treatment_fLow:genderFemale  -0.188891   0.113758 -1.6605   0.09751 .  
treatment_fHigh:genderFemale -0.051455   0.114474 -0.4495   0.65329    
---
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                           0.099
                                        (0.080)

treatment_fHigh                          0.054
                                        (0.079)

genderFemale                             0.109
                                        (0.080)

factor(city)houston                    -0.169***
                                        (0.064)

factor(city)sandiego                    -0.053
                                        (0.064)

factor(city)seattle                      0.038
                                        (0.066)

factor(bedrooms)2-bedroom                0.028
                                        (0.050)

price                                  -0.00000
                                       (0.00003)

treatment_fLow:genderFemale             -0.194*
                                        (0.113)

treatment_fHigh:genderFemale            -0.049
                                        (0.113)

Constant                               0.419***
                                        (0.092)

---------------------------------------------------------
Observations                              463
R2                                       0.033
Adjusted R2                              0.011
Residual Std. Error                0.495 (df = 452)
F Statistic                      1.527 (df = 10; 452)
=========================================================
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)                  4.1906e-01  9.4500e-02  4.4345 1.16e-05 ***
treatment_fLow               9.8680e-02  8.0530e-02  1.2254 0.221067    
treatment_fHigh              5.4357e-02  7.9457e-02  0.6841 0.494254    
genderFemale                 1.0870e-01  8.0054e-02  1.3578 0.175197    
factor(city)houston         -1.6851e-01  6.3987e-02 -2.6335 0.008742 ** 
factor(city)sandiego        -5.3147e-02  6.6088e-02 -0.8042 0.421713    
factor(city)seattle          3.8055e-02  6.7536e-02  0.5635 0.573391    
factor(bedrooms)2-bedroom    2.8279e-02  4.9665e-02  0.5694 0.569366    
price                       -1.8808e-06  3.1555e-05 -0.0596 0.952499    
treatment_fLow:genderFemale  -1.9405e-01  1.1360e-01 -1.7081 0.088303 .  
treatment_fHigh:genderFemale -4.8976e-02  1.1449e-01 -0.4278 0.669017    
---
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*gender+factor(city)+factor(bedrooms)+price+sqft+profession
al,data=d)
stargazer(m6.1,type='text')
```

```
=====================================================
                              Dependent variable:
                          ---------------------------
                                    outcome
-----------------------------------------------------
treatment                            0.026
                                    (0.040)

genderFemale                         0.059
                                    (0.073)

factor(city)houston                -0.181***
                                    (0.067)

factor(city)sandiego                -0.031
                                    (0.068)

factor(city)seattle                  0.051
                                    (0.071)

factor(bedrooms)2-bedroom           -0.009
                                    (0.055)

price                              -0.00003
                                   (0.00003)

sqft                                0.0002*
                                   (0.0001)

professionalProfessional            0.124**
                                    (0.062)

treatment:genderFemale              -0.027
                                    (0.057)

Constant                             0.193
                                    (0.144)

-----------------------------------------------------
Observations                          463
R2                                   0.039
Adjusted R2                          0.018
Residual Std. Error             0.493 (df = 452)
F Statistic                   1.832* (df = 10; 452)
=====================================================
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)                1.9334e-01  1.4554e-01  1.3285  0.18470
treatment                  2.6110e-02  3.9819e-02  0.6557  0.51235
genderFemale               5.9426e-02  7.2995e-02  0.8141  0.41601
factor(city)houston       -1.8053e-01  6.6516e-02 -2.7141  0.00690 **
factor(city)sandiego      -3.0741e-02  6.9931e-02 -0.4396  0.66044
factor(city)seattle        5.0970e-02  7.2326e-02  0.7047  0.48134
factor(bedrooms)2-bedroom -8.6209e-03  5.3405e-02 -0.1614  0.87183
price                     -3.0499e-05  3.4144e-05 -0.8932  0.37220
sqft                       2.2319e-04  1.2288e-04  1.8164  0.06998 .
professionalProfessional   1.2359e-01  6.3672e-02  1.9410  0.05288 .
treatment:genderFemale    -2.6580e-02  5.7256e-02 -0.4642  0.64271
---
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')
coeftest(m7, vcovHC(m7)) # Robust se

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

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

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)
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

# By city
perms <- genperms(Z, clustvar = NULL, blockvar = blk2)
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 char
acteristics of sampling dist. for inference

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

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

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

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

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)

# 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

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

```
[1] "Which is about 5 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.