W241 Class Project - Analysis

1. Load libraries, data

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

```
# 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/Anal
ysis/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

# Base data
head(d)

# 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 compliers
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$treatment_f <- factor(d$treatment, labels = c("Control","Low","High"))
```

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

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

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

```
## Table of Outcomes:
```

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

```
##
## No Response    Response
##         264         219
```

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

```
##
## Table of Outcomes (By Gender):
```

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

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

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

```
##
## Table of Outcomes (By Treatment):
```

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

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

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

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

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

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

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

```
##
## Table of Outcomes (By City):
```

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

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

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

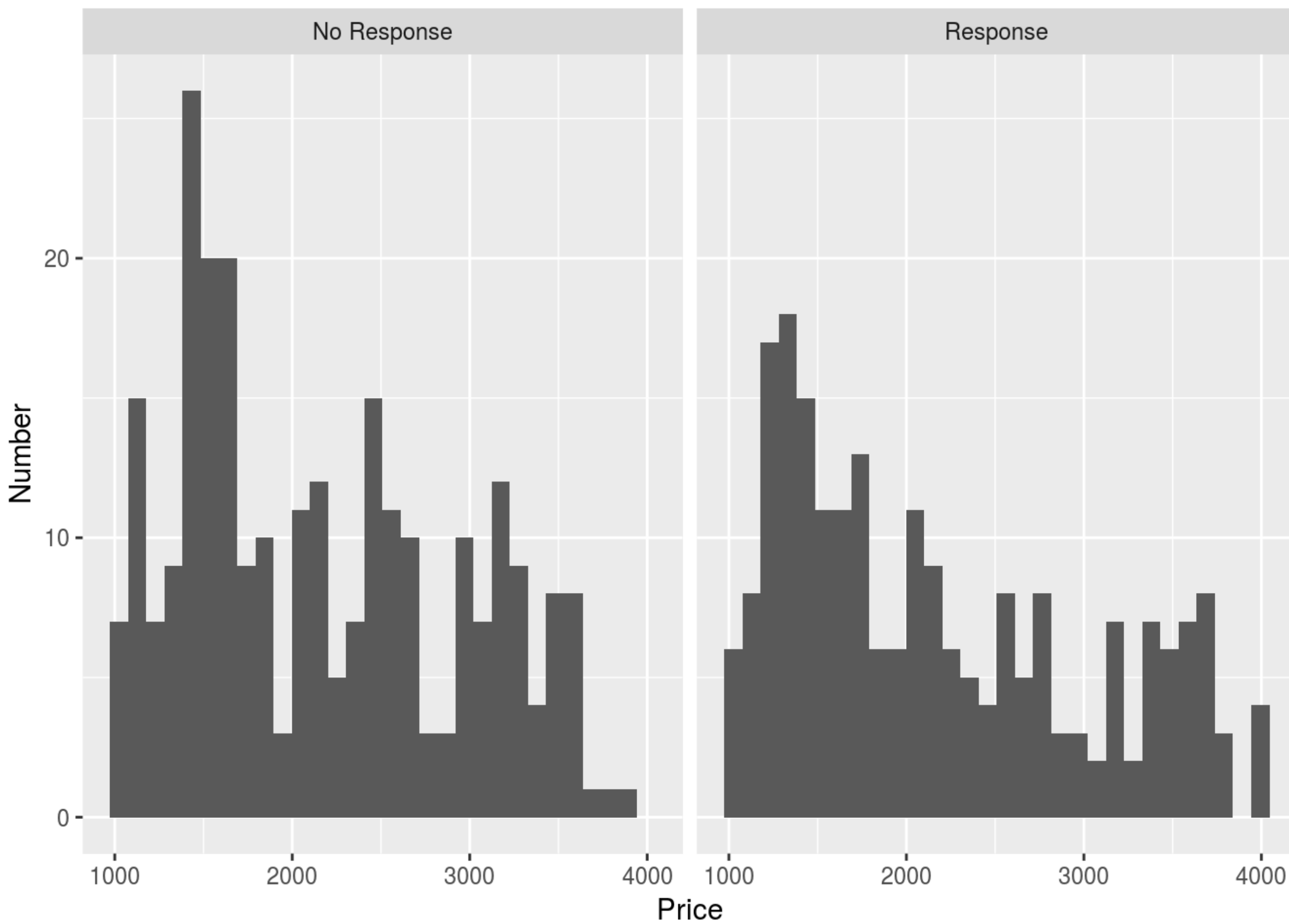
```
##
## Table of Outcomes (By Rooms):
```

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

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

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

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# Sqft info has missing values => we can drop all cases (see above) but for now leave t
his alone
# Similar but somewhat worse issue for professional, same.email info
```

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.

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

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

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

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tbl
## X-squared = 0.49961, df = 1, p-value = 0.4797
```

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

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

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

```
##
##  Pearson's Chi-squared test
##
## data:  tbl
## X-squared = 0.092173, df = 2, p-value = 0.955
```

```
# 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
##   No Response            39              41             49           49
##   Response               39              40             36           33
##
##                John_Treat_High John_Treat_Low
##   No Response              46             40
##   Response                 35             36
```

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

```
##
##  Pearson's Chi-squared test
##
## data:  tbl
## X-squared = 2.6574, df = 5, p-value = 0.7526
```

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
```

```
# 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                      0.007
##                               (0.028)
##
## Constant                     0.447***
##                               (0.036)
##
## -----------------------------------------------
## Observations                    483
## R2                            0.0001
## Adjusted R2                   -0.002
## Residual Std. Error      0.499 (df = 481)
## F Statistic            0.055 (df = 1; 481)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

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

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4468885  0.0360544  12.395   <2e-16 ***
## treatment   0.0065007  0.0279011   0.233   0.8159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## ==================================================
##                            Dependent variable:
##                        ---------------------------
##                                  outcome
## --------------------------------------------------
## treatment                         0.015
##                                  (0.039)
##
## genderFemale                      0.054
##                                  (0.072)
##
## treatment:genderFemale           -0.018
##                                  (0.056)
##
## Constant                        0.420***
##                                  (0.051)
##
## --------------------------------------------------
## Observations                       483
## R2                                0.002
## Adjusted R2                      -0.005
## Residual Std. Error          0.499 (df = 479)
## F Statistic                0.261 (df = 3; 479)
## ==================================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

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

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value  Pr(>|t|)
## (Intercept)              0.420269   0.050348  8.3472 7.548e-16 ***
## treatment                0.014940   0.039035  0.3827    0.7021
## genderFemale             0.053678   0.072275  0.7427    0.4580
## treatment:genderFemale -0.017544   0.055980 -0.3134    0.7541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 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                           0.015
##                                    (0.039)
##
## genderFemale                        0.052
##                                    (0.072)
##
## factor(city)houston               -0.155**
##                                    (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                             -0.00000
##                                   (0.00003)
##
## treatment:genderFemale             -0.017
##                                    (0.055)
##
## Constant                          0.452***
##                                    (0.087)
##
## -----------------------------------------------------
## Observations                         483
## R2                                  0.025
## Adjusted R2                         0.008
## Residual Std. Error           0.496 (df = 474)
## F Statistic                 1.503 (df = 8; 474)
## =====================================================
## Note:                     *p<0.1; **p<0.05; ***p<0.01
```

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

```
##
## t test of coefficients:
##
##                           Estimate  Std. Error t value  Pr(>|t|)    
## (Intercept)              4.5237e-01  8.8895e-02  5.0887 5.205e-07 ***
## treatment                1.5360e-02  3.9044e-02  0.3934   0.69420    
## genderFemale             5.1775e-02  7.2094e-02  0.7182   0.47301    
## factor(city)houston     -1.5451e-01  6.3680e-02 -2.4263   0.01563 *  
## factor(city)sandiego    -4.5650e-02  6.5718e-02 -0.6946   0.48763    
## factor(city)seattle      4.7687e-02  6.4965e-02  0.7340   0.46329    
## factor(bedrooms)2-bedroom  3.0561e-02  4.8223e-02  0.6337   0.52655    
## price                   -4.8779e-06  3.0726e-05 -0.1588   0.87393    
## treatment:genderFemale  -1.7236e-02  5.5989e-02 -0.3078   0.75834    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 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
```

```
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
```

```
# 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
```

```
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
```

```
# 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
```

```
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
```

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.

```
# 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
```

```
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
```

```
# 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
```

```
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
```

```
# Model 3c - Treatment & gender + covariates
m9 <- lm(outcome~low_treatment + high_treatment + gender + factor(city) + factor(bedroo
ms) + 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
```

```
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.

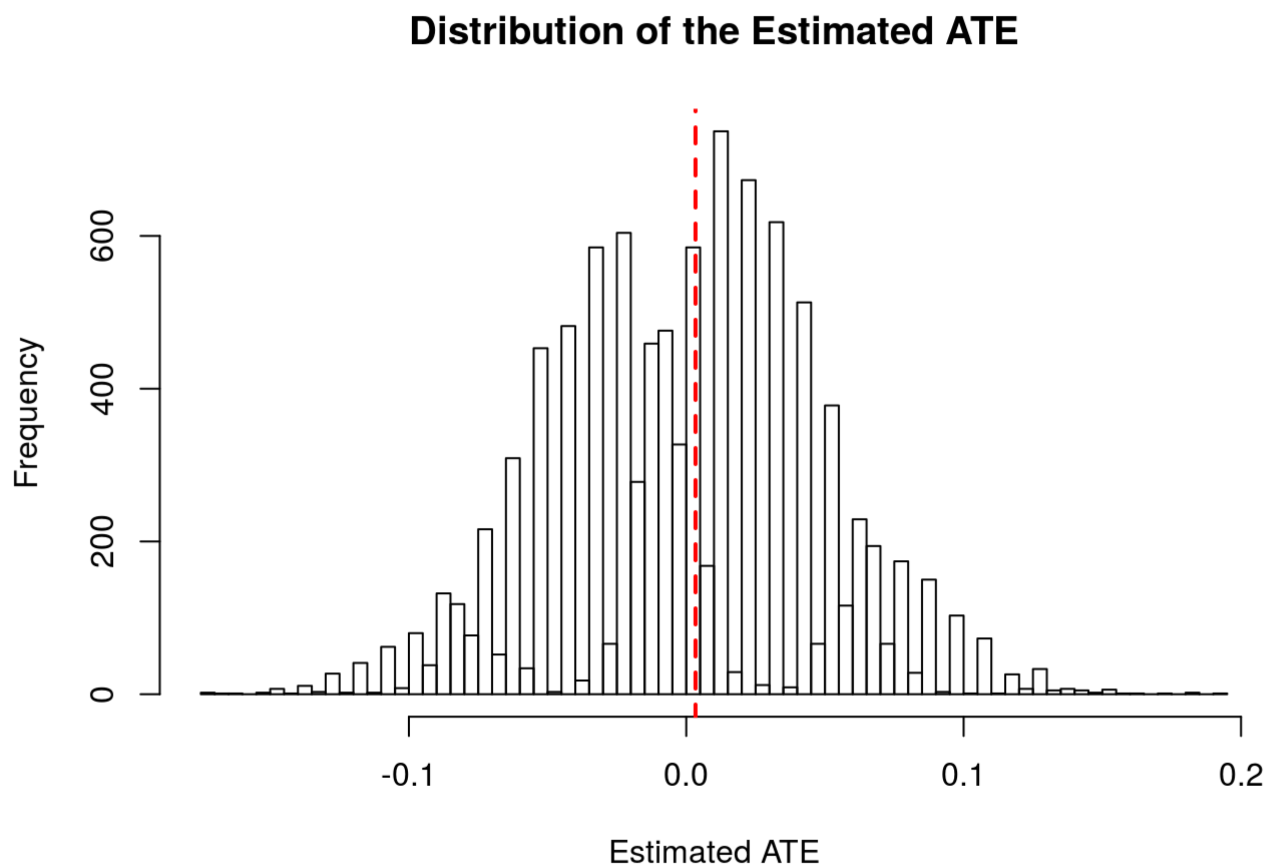
Randomization Inference

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).

```
# 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.
```

```
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
```

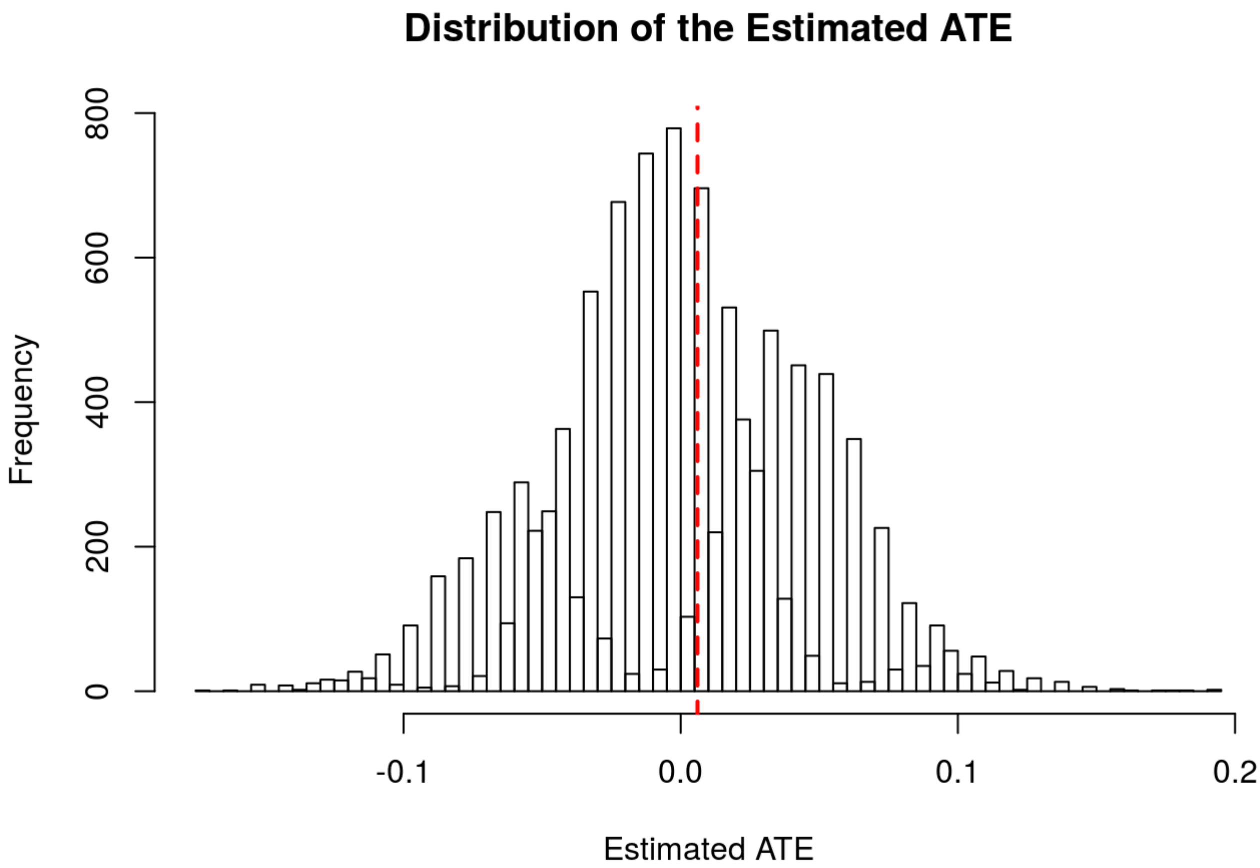


```
## $two.tailed.p.value
## [1] 0.9866
##
## $two.tailed.p.value.abs
## [1] 0.9898
##
## $greater.p.value
## [1] 0.4933
##
## $lesser.p.value
## [1] 0.5146
##
## $quantile
##        2.5%       97.5%
## -0.09195981  0.09753228
##
## $sd
## [1] 0.04848565
##
## $exp.val
## [1] -0.0001730871
```

```
# 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.
```

```
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
```

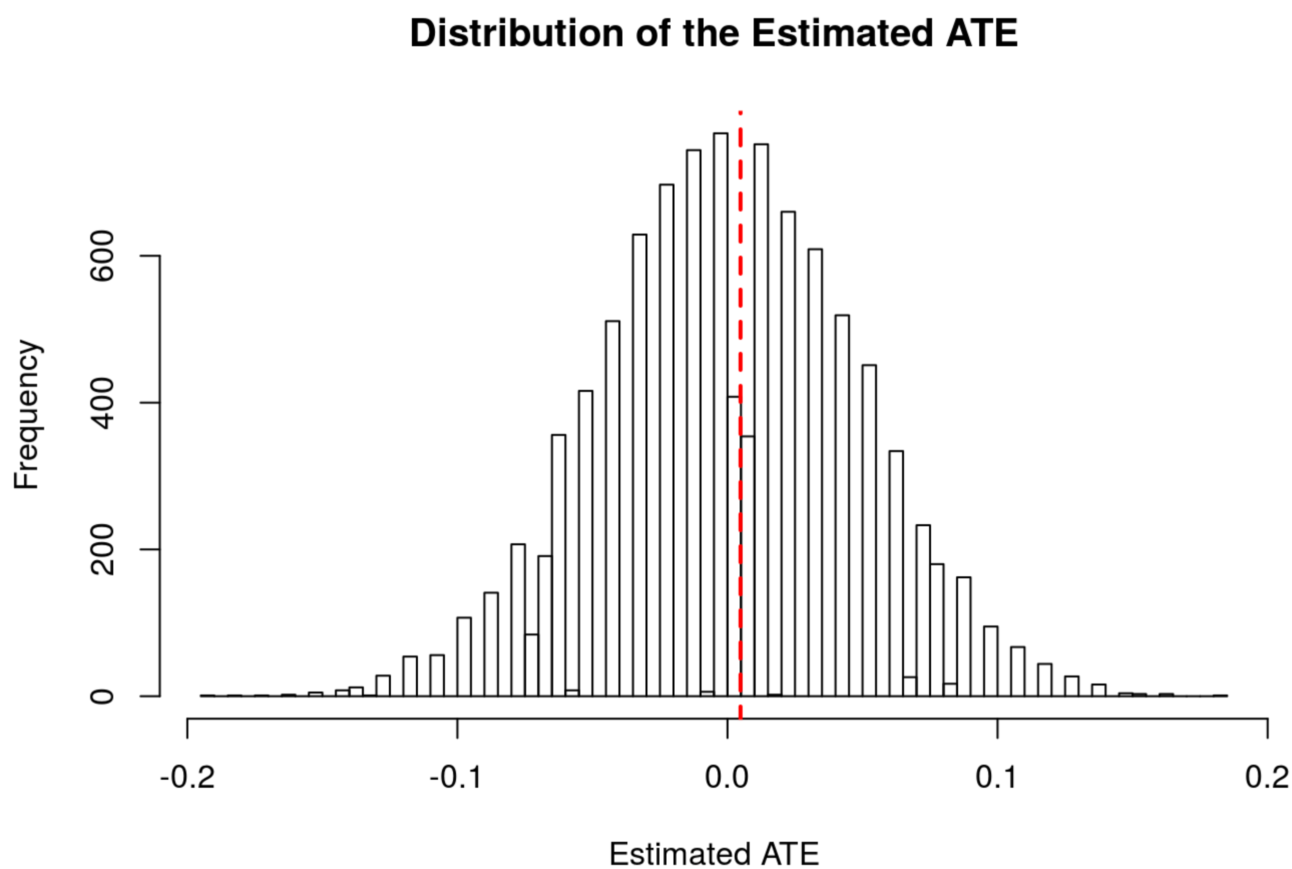



```
## $two.tailed.p.value
## [1] 0.908
##
## $two.tailed.p.value.abs
## [1] 0.8851
##
## $greater.p.value
## [1] 0.454
##
## $lesser.p.value
## [1] 0.5467
##
## $quantile
##        2.5%       97.5%
## -0.09580659  0.09096136
##
## $sd
## [1] 0.04736246
##
## $exp.val
## [1] -0.0001792889
```

```
# 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.
```

```
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] 0.9574
##
## $two.tailed.p.value.abs
## [1] 0.9078
##
## $greater.p.value
## [1] 0.4787
##
## $lesser.p.value
## [1] 0.5326
##
## $quantile
##        2.5%       97.5%
## -0.09773104  0.09808125
##
## $sd
## [1] 0.04844126
##
## $exp.val
## [1] -4.608262e-05
```

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

```
## [1] 0.8965
```

```
#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
## 1:         0 0.4500000
## 2:         1 0.4551084
```

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

```
##    treatment  V1
## 1:         0 160
## 2:         1 323
```

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

Other Analysis

CACE

```
# Using Models # NEEDS WORK
itt_fit <- lm(outcome ~ treatment, data = d)
summary(itt_fit)
```

```
##
## Call:
## lm(formula = outcome ~ treatment, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4551 -0.4551 -0.4500  0.5449  0.5500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.450000   0.039438  11.410   <2e-16 ***
## treatment   0.005108   0.048226   0.106    0.916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4989 on 481 degrees of freedom
## Multiple R-squared:  2.333e-05,  Adjusted R-squared:  -0.002056
## F-statistic: 0.01122 on 1 and 481 DF,  p-value: 0.9157
```

```
coeftest(itt_fit, vcovHC(itt_fit))
```

```
##
## 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
```

```
itt_d_fit <- lm(compliers ~ treatment, data = d)
coeftest(itt_d_fit)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7000000  0.0363817 19.2405   <2e-16 ***
## treatment   -0.0034056  0.0444892 -0.0765    0.939
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(itt_d_fit,vcovHC(itt_d_fit))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7000000  0.0364563 19.2011   <2e-16 ***
## treatment   -0.0034056  0.0445810 -0.0764   0.9391
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
itt_fit$coefficients[2] / itt_d_fit$coefficients[2]
```

```
## treatment
##      -1.5
```

```
# Manually compute CACE
itt <- d[, mean(outcome[assigned == 1]) - mean(outcome[assigned == 0])]

prop_treated <- d[ , mean(compliers/assigned, na.rm = T)]
prop_treated <- d[assigned == 1, mean(compliers)]

sprintf("%.10f", itt / prop_treated)
```

```
## [1] "-0.0207707708"
```

Added by NC. Work to find treatment response rate required to reject null

```
#create a new temp column of outcomes where the share of responses is 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]
```

```
## Warning in `[.data.table`(di, treatment == 1, `:=`(hypothetical.outcomes, :
## Coerced 'double' RHS to 'integer' to match the column's type; may have
## truncated precision. Either change the target column to 'double' first
## (by creating a new 'double' vector length 483 (nrows of entire table) and
## assign that; i.e. 'replace' column), or coerce RHS to 'integer' (e.g. 1L,
## NA_[real|integer]_, as.*, etc) to make your intent clear and for speed. Or,
## set the column type correctly up front when you create the table and stick
## to it, please.
```

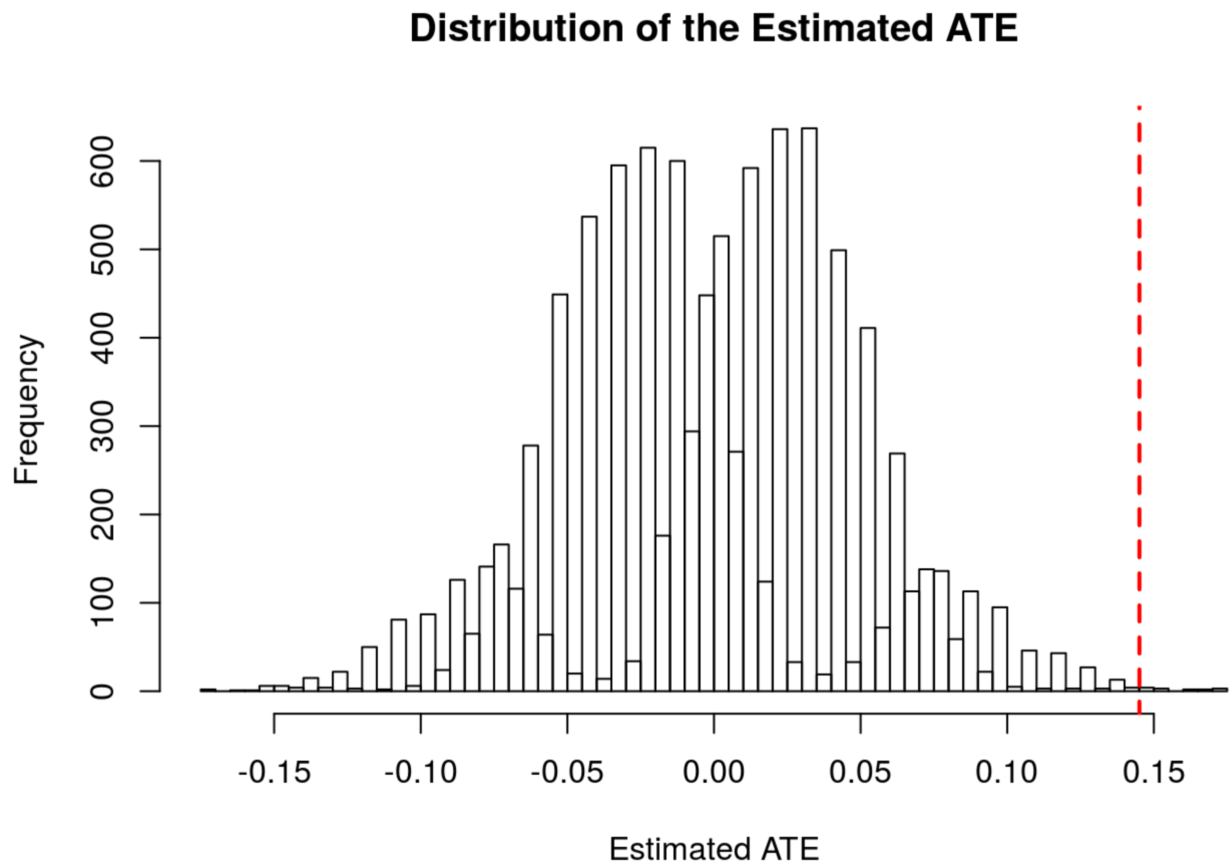
```
fake.response.rate.by.treatment <- di[, mean(hypothetical.outcomes), by =
c("treatment")]
fake.response.rate.by.treatment
```

```
##    treatment        V1
## 1:         0 0.4500000
## 2:         1 0.5944272
```

```
#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.
```

```
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] 0.0026
##
## $two.tailed.p.value.abs
## [1] 0.0029
##
## $greater.p.value
## [1] 0.0013
##
## $lesser.p.value
## [1] 0.9987
##
## $quantile
##        2.5%       97.5%
## -0.09788821  0.09629793
##
## $sd
## [1] 0.04870111
##
## $exp.val
## [1] -0.00103068
```

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

```
## [1] 0.0029
```

4. Conclusion

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