W241_Project_Analysis

Jeffrey Hsu

04/10/2017

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Warning: package 'Hmisc' was built under R version 3.3.2
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

Load Sales Data

```
SOColumnName <- c("OrderID", "Date", "BuyerID", "NrItems",
                "OrderValue", "ShippingCost", "OrderTotal",
                "City", "Country")

# Load Sales Order Data for 2017-03
SO1703 <- read.csv("./data/SoldOrders201703.csv", sep=",")
SO1703 <- subset(SO1703,
                 select=c(Order.ID,Sale.Date,Buyer.User.ID,
                          Number.of.Items,Order.Value,Shipping,
                          Order.Total,Ship.City,Ship.Country))
colnames(SO1703) <- SOColumnName
SO1703$Date <- as.Date(SO1703$Date,format = "%m/%d/%y")

# Load Sales Order Data for 2017-04
SO1704 <- read.csv("./data/SoldOrders201704.csv", sep=",")
SO1704 <- subset(SO1704,
                 select=c(Order.ID,Sale.Date,Buyer.User.ID,
                          Number.of.Items,Order.Value,Shipping,
                          Order.Total,Ship.City,Ship.Country))
colnames(SO1704) <- SOColumnName
SO1704$Date <- as.Date(SO1704$Date,format = "%m/%d/%y")
```

```r
SO1617 <- read.csv("./data/SoldOrders201609-201702.csv",sep=",")
SO1617 <- subset(SO1617,
                 select=c(Order.ID,Sale.Date,Buyer.User.ID,
                          Number.of.Items,Order.Value,Shipping,
                          Order.Total,Ship.City,Ship.Country))
colnames(SO1617) <- SOColumnName
SO1617$Date <- as.Date(SO1617$Date,format = "%m/%d/%y")

# Merge into 1 sales order dataset
SO <- rbind(SO1704, SO1703)
SO <- rbind(SO, SO1617)

## Load Sales Order Item data
SOI1703 <- read.csv("./data/SoldOrderItems201703.csv", sep=",")
SOI1704 <- read.csv("./data/SoldOrderItems201704.csv", sep=",")
```

transform into treatment and control datasets

```r
str(SO)
```

```
## 'data.frame':    1161 obs. of  9 variables:
##  $ OrderID     : chr  "1185591770" "1191482821" "1185447738" "1185425334" ...
##  $ Date        : Date, format: "2017-04-09" "2017-04-09" ...
##  $ BuyerID     : Factor w/ 1107 levels "amusing01","andrearod5o3",..: 20 15 6 8 12 18 9 14 13 3 ...
##  $ NrItems     : int  1 2 1 1 3 1 1 1 7 1 ...
##  $ OrderValue  : num  9.99 13.98 6.99 6.99 30.97 ...
##  $ ShippingCost: num  3.75 4.25 3.75 3.75 4.75 3.75 3.75 3.75 6.75 3.75 ...
##  $ OrderTotal  : num  13.7 18.2 10.7 10.7 35.7 ...
##  $ City        : Factor w/ 880 levels "Auburn","Calgary",..: 13 21 22 5 9 19 18 10 6 8 ...
##  $ Country     : Factor w/ 48 levels "Austria","Canada",..: 10 7 10 9 6 10 10 10 3 6 ...
```

```r
SO_cont1 <- subset(SO, SO$Date > "2017-03-12" & SO$Date < "2017-03-20")
SO_treat <- subset(SO, SO$Date > "2017-03-19" & SO$Date < "2017-03-27")
SO_cont2 <- subset(SO, SO$Date > "2017-03-26" & SO$Date < "2017-04-03")
SO_cont3 <- subset(SO, SO$Date > "2017-04-02" & SO$Date < "2017-04-10")
SO_base <- subset(SO, SO$Date < "2017-03-13")
```

Load Websession Data

```r
session_control1 <- read.csv("./data/websession_control1.csv", sep=",")
session_control2 <- read.csv("./data/websession_control2.csv", sep=",")
session_control3 <- read.csv("./data/websession_control3.csv", sep=",")
session_treatment <- read.csv("./data/websession_treatment.csv", sep=",")
```

Transform and merge session data

Load Demographic Data

```r
d_age <- read.csv("./data/GA_age.csv", sep=",")
d_gender <- read.csv("./data/GA_gender.csv", sep=",")
```

Data Combine

Covariate Balance

Covariate Balance checks that other factors of sales stays in similar level across control and treatment weeks. This assures the randomness of treatment or control given a single website visit.

Age

```
## -----Contorl week 1 V.s. Treatment Week-----

# Check balance cont1 v.s. treat of age group 18-24
t.test(as.numeric(d_age[1,2:8]),as.numeric(d_age[1,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[1, 2:8]) and as.numeric(d_age[1, 9:15])
## t = 0.85797, df = 6, p-value = 0.4238
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.143321 14.857606
## sample estimates:
## mean of the differences
##                3.857143
```

```
wilcox.test(as.numeric(d_age[1,2:8]),as.numeric(d_age[1,9:15]), paired = TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  as.numeric(d_age[1, 2:8]) and as.numeric(d_age[1, 9:15])
## V = 19, p-value = 0.4688
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check balance cont1 v.s. treat of age group 15-34
t.test(as.numeric(d_age[2,2:8]),as.numeric(d_age[2,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[2, 2:8]) and as.numeric(d_age[2, 9:15])
## t = -1.8297, df = 6, p-value = 0.117
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.688407   1.831264
## sample estimates:
## mean of the differences
##               -5.428571
```

```
wilcox.test(as.numeric(d_age[2,2:8]),as.numeric(d_age[2,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[2, 2:8]),
## as.numeric(d_age[2, : cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(as.numeric(d_age[2, 2:8]),
## as.numeric(d_age[2, : cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[2, 2:8]) and as.numeric(d_age[2, 9:15])
## V = 3.5, p-value = 0.1718
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check balance cont1 v.s. treat of age group 35-44
t.test(as.numeric(d_age[3,2:8]),as.numeric(d_age[3,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[3, 2:8]) and as.numeric(d_age[3, 9:15])
## t = -0.6563, df = 6, p-value = 0.536
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.807593   6.236164
## sample estimates:
## mean of the differences
##               -2.285714
```

```
wilcox.test(as.numeric(d_age[3,2:8]),as.numeric(d_age[3,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[3, 2:8]),
## as.numeric(d_age[3, : cannot compute exact p-value with ties

## Warning in wilcox.test.default(as.numeric(d_age[3, 2:8]),
## as.numeric(d_age[3, : cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[3, 2:8]) and as.numeric(d_age[3, 9:15])
## V = 7, p-value = 0.5271
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check balance cont1 v.s. treat of age group 45-54
t.test(as.numeric(d_age[4,2:8]),as.numeric(d_age[4,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[4, 2:8]) and as.numeric(d_age[4, 9:15])
## t = -1.082, df = 6, p-value = 0.3208
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.045855   5.045855
## sample estimates:
## mean of the differences
##                      -4
```

```
wilcox.test(as.numeric(d_age[4,2:8]),as.numeric(d_age[4,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[4, 2:8]),
```

```
## as.numeric(d_age[4, : cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[4, 2:8]) and as.numeric(d_age[4, 9:15])
## V = 2, p-value = 0.3613
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check balance cont1 v.s. treat of all age groups
t.test(as.numeric(d_age[5,2:8]),as.numeric(d_age[5,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[5, 2:8]) and as.numeric(d_age[5, 9:15])
## t = -0.74001, df = 6, p-value = 0.4872
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -33.83762  18.12333
## sample estimates:
## mean of the differences
##               -7.857143
```

```r
wilcox.test(as.numeric(d_age[5,2:8]),as.numeric(d_age[5,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[5, 2:8]),
## as.numeric(d_age[5, : cannot compute exact p-value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[5, 2:8]) and as.numeric(d_age[5, 9:15])
## V = 10.5, p-value = 0.6108
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check age composition balance of cont1 v.s. treat
cont1_age_comp <- c( rep(21,sum(d_age[1,2:8])),
                     rep(29.5,sum(d_age[2,2:8])),
                     rep(39.5,sum(d_age[3,2:8])),
                     rep(49.5,sum(d_age[4,2:8])) )
treat_age_comp <- c( rep(21,sum(d_age[1,9:16])),
                     rep(29.5,sum(d_age[2,9:16])),
                     rep(39.5,sum(d_age[3,9:16])),
                     rep(49.5,sum(d_age[4,9:16])) )
t.test(cont1_age_comp, treat_age_comp)
```

```
##
##  Welch Two Sample t-test
##
## data:  cont1_age_comp and treat_age_comp
## t = -3.6312, df = 1209.9, p-value = 0.0002939
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.450018 -0.731212
## sample estimates:
## mean of x mean of y
```

```
## 27.75503 29.34564
```

```r
wilcox.test(cont1_age_comp, treat_age_comp)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: cont1_age_comp and treat_age_comp
## W = 166060, p-value = 0.0008514
## alternative hypothesis: true location shift is not equal to 0
```

```r
## -----Contorl week 2 V.s. Treatment Week-----

# Check balance cont2 v.s. treat of age group 18-24
t.test(as.numeric(d_age[1,16:22]),as.numeric(d_age[1,9:15]), paired = TRUE)
```

```
##
## Paired t-test
##
## data: as.numeric(d_age[1, 16:22]) and as.numeric(d_age[1, 9:15])
## t = -0.036572, df = 6, p-value = 0.972
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -9.701017 9.415303
## sample estimates:
## mean of the differences
##              -0.1428571
```

```r
wilcox.test(as.numeric(d_age[1,16:22]),as.numeric(d_age[1,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[1, 16:22]),
## as.numeric(d_age[1, : cannot compute exact p-value with zeroes
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: as.numeric(d_age[1, 16:22]) and as.numeric(d_age[1, 9:15])
## V = 10, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check balance cont2 v.s. treat of age group 25-34
t.test(as.numeric(d_age[2,16:22]),as.numeric(d_age[2,9:15]), paired = TRUE)
```

```
##
## Paired t-test
##
## data: as.numeric(d_age[2, 16:22]) and as.numeric(d_age[2, 9:15])
## t = -1.2718, df = 6, p-value = 0.2505
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -15.455013 4.883584
## sample estimates:
## mean of the differences
##               -5.285714
```

```r
wilcox.test(as.numeric(d_age[2,16:22]),as.numeric(d_age[2,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[2, 16:22]),
```

```
## as.numeric(d_age[2, : cannot compute exact p-value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[2, 16:22]) and as.numeric(d_age[2, 9:15])
## V = 7, p-value = 0.2685
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check balance cont2 v.s. treat of age group 35-44
t.test(as.numeric(d_age[3,16:22]),as.numeric(d_age[3,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[3, 16:22]) and as.numeric(d_age[3, 9:15])
## t = -2.2529, df = 6, p-value = 0.06518
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.3909616   0.6766759
## sample estimates:
## mean of the differences
##               -7.857143
```

```
wilcox.test(as.numeric(d_age[3,16:22]),as.numeric(d_age[3,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[3, 16:22]),
## as.numeric(d_age[3, : cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[3, 16:22]) and as.numeric(d_age[3, 9:15])
## V = 3, p-value = 0.1422
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check balance cont2 v.s. treat of age group 45-54
t.test(as.numeric(d_age[4,16:22]),as.numeric(d_age[4,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[4, 16:22]) and as.numeric(d_age[4, 9:15])
## t = -0.036515, df = 6, p-value = 0.9721
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.715920  9.430206
## sample estimates:
## mean of the differences
##              -0.1428571
```

```
wilcox.test(as.numeric(d_age[4,16:22]),as.numeric(d_age[4,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[4, 16:22]),
## as.numeric(d_age[4, : cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
```

```
##
## data:  as.numeric(d_age[4, 16:22]) and as.numeric(d_age[4, 9:15])
## V = 7, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check balance cont2 v.s. treat of all age groups
t.test(as.numeric(d_age[5,16:22]),as.numeric(d_age[5,9:15]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[5, 16:22]) and as.numeric(d_age[5, 9:15])
## t = -1.0906, df = 6, p-value = 0.3173
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -43.55762  16.70048
## sample estimates:
## mean of the differences
##               -13.42857
```

```r
wilcox.test(as.numeric(d_age[5,16:22]),as.numeric(d_age[5,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[5, 16:22]),
## as.numeric(d_age[5, : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[5, 16:22]) and as.numeric(d_age[5, 9:15])
## V = 8.5, p-value = 0.3972
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check age composition balance of cont2 v.s. treat
cont2_age_comp <- c( rep(21,sum(d_age[1,16:22])),
                     rep(29.5,sum(d_age[2,16:22])),
                     rep(39.5,sum(d_age[3,16:22])),
                     rep(49.5,sum(d_age[4,16:22])) )
t.test(cont2_age_comp, treat_age_comp)
```

```
##
##  Welch Two Sample t-test
##
## data:  cont2_age_comp and treat_age_comp
## t = -2.0032, df = 1081.6, p-value = 0.0454
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.90189326 -0.01970679
## sample estimates:
## mean of x mean of y
##  28.38484  29.34564
```

```r
wilcox.test(cont2_age_comp, treat_age_comp)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  cont2_age_comp and treat_age_comp
```

```
## W = 158180, p-value = 0.01112
## alternative hypothesis: true location shift is not equal to 0
```

```
## -----Contorl week 2 V.s. Control Week 1-----

# Check balance cont2 v.s. cont1 of age group 18-24
t.test(as.numeric(d_age[1,16:22]),as.numeric(d_age[1,2:8]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[1, 16:22]) and as.numeric(d_age[1, 2:8])
## t = -0.96077, df = 6, p-value = 0.3738
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.187306   6.187306
## sample estimates:
## mean of the differences
##                      -4
```

```
wilcox.test(as.numeric(d_age[1,16:22]),as.numeric(d_age[1,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[1, 16:22]),
## as.numeric(d_age[1, : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[1, 16:22]) and as.numeric(d_age[1, 2:8])
## V = 8, p-value = 0.3517
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check balance cont2 v.s. cont1 of age group 25-34
t.test(as.numeric(d_age[2,16:22]),as.numeric(d_age[2,2:8]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[2, 16:22]) and as.numeric(d_age[2, 2:8])
## t = 0.036629, df = 6, p-value = 0.972
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.400377  9.686091
## sample estimates:
## mean of the differences
##               0.1428571
```

```
wilcox.test(as.numeric(d_age[2,16:22]),as.numeric(d_age[2,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[2, 16:22]),
## as.numeric(d_age[2, : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[2, 16:22]) and as.numeric(d_age[2, 2:8])
## V = 15, p-value = 0.9325
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check balance cont2 v.s. cont1 of age group 35-44
t.test(as.numeric(d_age[3,16:22]),as.numeric(d_age[3,2:8]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[3, 16:22]) and as.numeric(d_age[3, 2:8])
## t = -1.4945, df = 6, p-value = 0.1857
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.693508   3.550651
## sample estimates:
## mean of the differences
##               -5.571429
```

```r
wilcox.test(as.numeric(d_age[3,16:22]),as.numeric(d_age[3,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[3, 16:22]),
## as.numeric(d_age[3, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[3, 16:22]) and as.numeric(d_age[3, 2:8])
## V = 1, p-value = 0.2012
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check balance cont2 v.s. cont1 of age group 45-54
t.test(as.numeric(d_age[4,16:22]),as.numeric(d_age[4,2:8]), paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  as.numeric(d_age[4, 16:22]) and as.numeric(d_age[4, 2:8])
## t = 1.3955, df = 6, p-value = 0.2123
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.906015 10.620301
## sample estimates:
## mean of the differences
##                3.857143
```

```r
wilcox.test(as.numeric(d_age[4,16:22]),as.numeric(d_age[4,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[4, 16:22]),
## as.numeric(d_age[4, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[4, 16:22]) and as.numeric(d_age[4, 2:8])
## V = 5, p-value = 0.4227
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Check balance cont2 v.s. cont1 of age group 45-54
t.test(as.numeric(d_age[5,16:22]),as.numeric(d_age[5,2:8]), paired = TRUE)
```

```
##
```

```
##  Paired t-test
##
## data:  as.numeric(d_age[5, 16:22]) and as.numeric(d_age[5, 2:8])
## t = -0.8228, df = 6, p-value = 0.4421
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -22.14017  10.99732
## sample estimates:
## mean of the differences
##               -5.571429
```

```
wilcox.test(as.numeric(d_age[5,16:22]),as.numeric(d_age[5,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_age[5, 16:22]),
## as.numeric(d_age[5, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_age[5, 16:22]) and as.numeric(d_age[5, 2:8])
## V = 8, p-value = 0.675
## alternative hypothesis: true location shift is not equal to 0
```

```
# Check age composition balance of cont2 v.s. cont1
t.test(cont2_age_comp, cont1_age_comp)
```

```
##
##  Welch Two Sample t-test
##
## data:  cont2_age_comp and cont1_age_comp
## t = 1.3169, df = 1010.9, p-value = 0.1882
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3086649  1.5682951
## sample estimates:
## mean of x mean of y
##  28.38484  27.75503
```

```
wilcox.test(cont2_age_comp, cont1_age_comp)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  cont2_age_comp and cont1_age_comp
## W = 141970, p-value = 0.5066
## alternative hypothesis: true location shift is not equal to 0
```

Gender

```
## -----Control 1 v.s. Treatment-----
wilcox.test(as.numeric(d_gender[1,2:8]),as.numeric(d_gender[1,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[1, 2:8]),
## as.numeric(d_gender[1, : cannot compute exact p-value with ties
```

```
##
```

```
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_gender[1, 2:8]) and as.numeric(d_gender[1, 9:15])
## V = 10.5, p-value = 0.6115
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(d_gender[2,2:8]),as.numeric(d_gender[2,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[2, 2:8]),
## as.numeric(d_gender[2, : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_gender[2, 2:8]) and as.numeric(d_gender[2, 9:15])
## V = 20, p-value = 0.3517
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(d_gender[3,2:8]),as.numeric(d_gender[3,9:15]), paired = TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  as.numeric(d_gender[3, 2:8]) and as.numeric(d_gender[3, 9:15])
## V = 13, p-value = 0.9375
## alternative hypothesis: true location shift is not equal to 0
```

```r
cont1_gender_comp <- c(rep(0,sum(d_gender[1,2:8])), rep(1,sum(d_gender[2,2:8])))
treat_gender_comp <- c(rep(0,sum(d_gender[1,9:15])), rep(1,sum(d_gender[2,9:15])))
t.test(cont1_gender_comp, treat_gender_comp)
```

```
##
##  Welch Two Sample t-test
##
## data:  cont1_gender_comp and treat_gender_comp
## t = 2.3173, df = 1378.6, p-value = 0.02063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.006894197 0.082953327
## sample estimates:
## mean of x mean of y
## 0.1788269 0.1339031
```

```r
## -----Control 2 v.s. Treatment-----
wilcox.test(as.numeric(d_gender[1,16:22]),as.numeric(d_gender[1,9:15]), paired = TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  as.numeric(d_gender[1, 16:22]) and as.numeric(d_gender[1, 9:15])
## V = 12, p-value = 0.8125
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(d_gender[2,16:22]),as.numeric(d_gender[2,9:15]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[2, 16:22]),
## as.numeric(d_gender[2, : cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[2, 16:22]),
```

```
## as.numeric(d_gender[2, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_gender[2, 16:22]) and as.numeric(d_gender[2, 9:15])
## V = 5, p-value = 0.5879
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(d_gender[3,16:22]),as.numeric(d_gender[3,9:15]), paired = TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  as.numeric(d_gender[3, 16:22]) and as.numeric(d_gender[3, 9:15])
## V = 10, p-value = 0.5781
## alternative hypothesis: true location shift is not equal to 0
```

```r
cont2_gender_comp <- c(rep(0,sum(d_gender[1,16:22])), rep(1,sum(d_gender[2,16:22])))
t.test(cont2_gender_comp, treat_gender_comp)
```

```
##
##  Welch Two Sample t-test
##
## data:  cont2_gender_comp and treat_gender_comp
## t = -0.39902, df = 1333.7, p-value = 0.6899
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.04343009  0.02874882
## sample estimates:
## mean of x mean of y
## 0.1265625 0.1339031
```

```r
## -----Control 2 v.s. Control 1-----
wilcox.test(as.numeric(d_gender[1,16:22]),as.numeric(d_gender[1,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[1, 16:22]),
## as.numeric(d_gender[1, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_gender[1, 16:22]) and as.numeric(d_gender[1, 2:8])
## V = 10, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(d_gender[2,16:22]),as.numeric(d_gender[2,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[2, 16:22]),
## as.numeric(d_gender[2, : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_gender[2, 16:22]) and as.numeric(d_gender[2, 2:8])
## V = 1, p-value = 0.1056
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(d_gender[3,16:22]),as.numeric(d_gender[3,2:8]), paired = TRUE)
```

```
## Warning in wilcox.test.default(as.numeric(d_gender[3, 16:22]),
## as.numeric(d_gender[3, : cannot compute exact p-value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  as.numeric(d_gender[3, 16:22]) and as.numeric(d_gender[3, 2:8])
## V = 11.5, p-value = 0.7349
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(cont1_gender_comp, cont2_gender_comp)
```

```
##
##  Welch Two Sample t-test
##
## data:  cont1_gender_comp and cont2_gender_comp
## t = 2.6693, df = 1333.2, p-value = 0.007694
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.01385326 0.09067553
## sample estimates:
## mean of x mean of y
## 0.1788269 0.1265625
```

Country

Web Sessions of returning and new visitor composition

```r
t.test(session_control1$Sessions, session_treatment$Sessions)
```

```
##
##  Welch Two Sample t-test
##
## data:  session_control1$Sessions and session_treatment$Sessions
## t = 0.031802, df = 2019.6, p-value = 0.9746
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08921820  0.09215946
## sample estimates:
## mean of x mean of y
##  1.187117  1.185646
```

```r
t.test(session_control2$Sessions, session_treatment$Sessions)
```

```
##
##  Welch Two Sample t-test
##
## data:  session_control2$Sessions and session_treatment$Sessions
## t = 0.20635, df = 1957.7, p-value = 0.8365
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07707255  0.09519868
## sample estimates:
```

```
## mean of x mean of y
##  1.194709  1.185646
```

```r
t.test(session_control1$Sessions, session_control2$Sessions)
```

```
##
##  Welch Two Sample t-test
##
## data:  session_control1$Sessions and session_control2$Sessions
## t = -0.17888, df = 1902.3, p-value = 0.8581
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.09083619  0.07565133
## sample estimates:
## mean of x mean of y
##  1.187117  1.194709
```

Effect Size

```r
cohensD(cont1_gender_comp, treat_gender_comp) ## small effect
```

```
## [1] 0.1238527
```

```r
cohensD(cont1_age_comp, treat_age_comp) ## small effect
```

```
## [1] 0.2062712
```

```r
cohensD(session_control1$Sessions, session_treatment$Sessions) ## no effect
```

```
## [1] 0.001410579
```

Treatment Effects

Below are estimations for treatement effects on sales performance.

```r
past_customer <- unique(SO[SO$Date < "2017-03-13",]$BuyerID)
new_customer <- setdiff(unique(SO[SO$Date > "2017-03-12" & SO$Date < "2017-03-27",]$BuyerID),
                                  past_customer)
intersect(unique(SO[SO$Date > "2017-03-12" & SO$Date < "2017-03-20",]$BuyerID),
          unique(SO[SO$Date > "2017-03-19" & SO$Date < "2017-03-27",]$BuyerID))
```

```
## character(0)
```

```r
# there's no customer who placed an order in both control 1 week and treatment week
intersect(unique(SO[SO$Date > "2017-03-26" & SO$Date < "2017-04-09",]$BuyerID),
          unique(SO[SO$Date > "2017-03-19" & SO$Date < "2017-03-27",]$BuyerID))
```

```
## [1] "danielornelas2"
```

```r
# there's only 1 buyer that placed an order post treatment
# that also placed a order during the treatment week

# add new indicator on returning customer orders
SO$Returning_Customer <- as.numeric(SO$BuyerID %in% past_customer)

SO$treat <- as.numeric(SO$Date > "2017-03-19" & SO$Date < "2017-03-27")
```

Order Value

Control 1 v.s. Treatment

```
# average effect on order value of treatment v.s. control 1
SO_cont1_treat <- SO[SO$Date > "2017-03-12" & SO$Date < "2017-03-27",]
m1_ordervalue <- lm(OrderValue~treat, data=SO_cont1_treat)
summary(m1_ordervalue)
```

```
##
## Call:
## lm(formula = OrderValue ~ treat, data = SO_cont1_treat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.249 -13.849  -7.815  -0.825 113.931
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.805      5.381   2.751  0.00896 **
## treat          9.034      7.905   1.143  0.26005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.24 on 39 degrees of freedom
## Multiple R-squared:  0.03241,    Adjusted R-squared:  0.007597
## F-statistic: 1.306 on 1 and 39 DF,  p-value: 0.2601
```

```
m1_ordervalue$vcov <- vcov(m1_ordervalue)
coeftest(m1_ordervalue, m1_ordervalue$vcov)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.8050     5.3813  2.7512 0.008959 **
## treat         9.0345     7.9050  1.1429 0.260053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

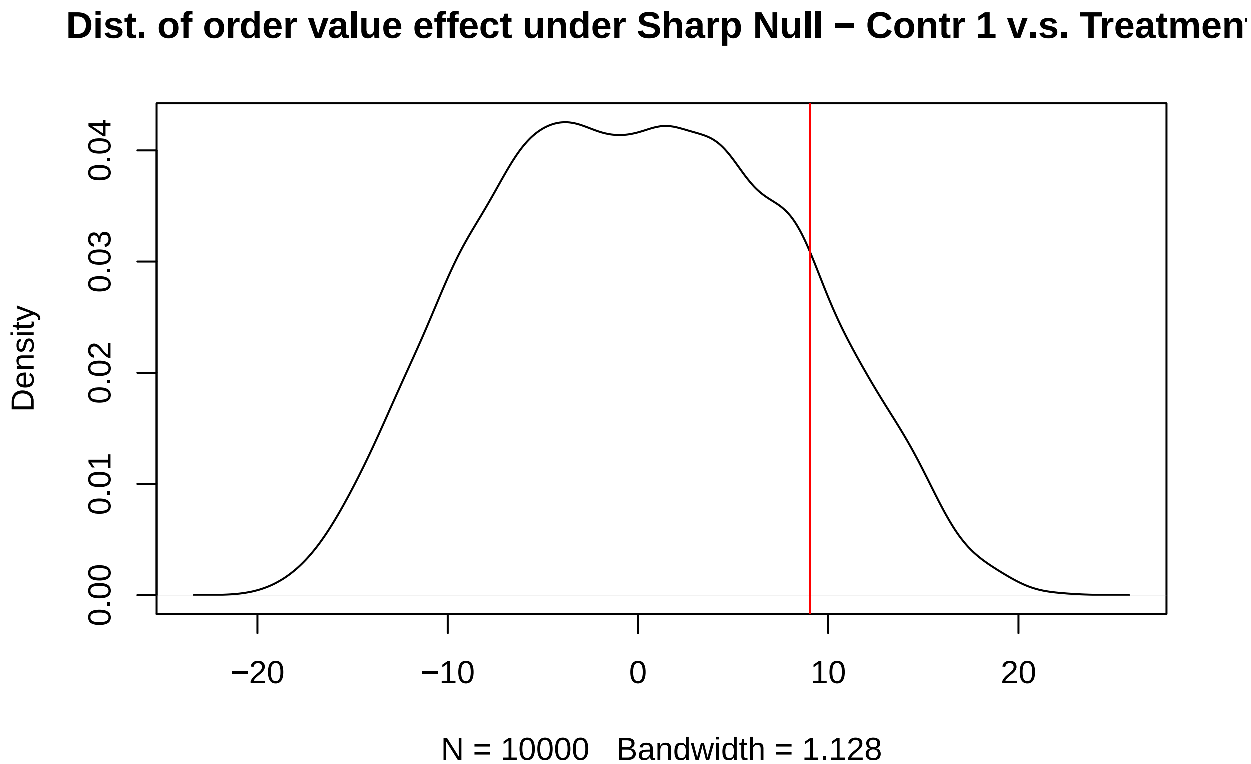
```
# non-parametric test
wilcox.test(SO_cont1$OrderValue,SO_treat$OrderValue)
```

```
## Warning in wilcox.test.default(SO_cont1$OrderValue, SO_treat$OrderValue):
## cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  SO_cont1$OrderValue and SO_treat$OrderValue
## W = 228, p-value = 0.6237
## alternative hypothesis: true location shift is not equal to 0
```

```
# ramdomization inference
ate <- mean(SO_treat$OrderValue) - mean(SO_cont1$OrderValue)
dist_sharpnull_ordervalue <- rep(0,10000)
for (i in 1:10000) {
```

```
  treat <- sample(SO_cont1_treat$treat)
  dist_sharpnull_ordervalue[i] <- mean(SO_cont1_treat[treat==1,]$OrderValue) -
    mean(SO_cont1_treat[treat==0,]$OrderValue)
}
plot(density(dist_sharpnull_ordervalue),
     main="Dist. of order value effect under Sharp Null - Contr 1 v.s. Treatment")
abline(v=ate, col="red")
```




```
mean(abs(ate) < abs(dist_sharpnull_ordervalue))
```

```
## [1] 0.2842
```

```
# Controlling for returning customer
m2_ordervalue <- lm(OrderValue~treat+Returning_Customer, data=SO_cont1_treat)
summary(m2_ordervalue)
```

```
##
## Call:
## lm(formula = OrderValue ~ treat + Returning_Customer, data = SO_cont1_treat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.147 -11.587  -3.499   1.501  88.945
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          10.489      4.945   2.121  0.04050 *
## treat                 6.687      7.053   0.948  0.34907
## Returning_Customer   31.648      9.348   3.386  0.00166 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.41 on 38 degrees of freedom
## Multiple R-squared:  0.2566, Adjusted R-squared:  0.2175
## F-statistic:  6.56 on 2 and 38 DF,  p-value: 0.003571
```

```
m2_ordervalue$vcov <- vcov(m2_ordervalue)
coeftest(m2_ordervalue, m2_ordervalue$vcov)
```

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          10.4894     4.9454  2.1210 0.040503 *
## treat                 6.6874     7.0534  0.9481 0.349068
## Returning_Customer   31.6478     9.3476  3.3857 0.001662 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare control 2 to treatment

```
SO_cont2_treat <- SO[SO$Date > "2017-03-19" & SO$Date < "2017-04-03",]
m3_ordervalue <- lm(OrderValue~treat, data=SO_cont2_treat)
summary(m2_ordervalue)
```

```
##
## Call:
## lm(formula = OrderValue ~ treat + Returning_Customer, data = SO_cont1_treat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.147 -11.587  -3.499   1.501  88.945
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           10.489      4.945   2.121  0.04050 *
## treat                  6.687      7.053   0.948  0.34907
## Returning_Customer    31.648      9.348   3.386  0.00166 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.41 on 38 degrees of freedom
## Multiple R-squared:  0.2566, Adjusted R-squared:  0.2175
## F-statistic:  6.56 on 2 and 38 DF,  p-value: 0.003571
```

```
m2_ordervalue$vcov <- vcov(m2_ordervalue)
coeftest(m2_ordervalue, m2_ordervalue$vcov)
```

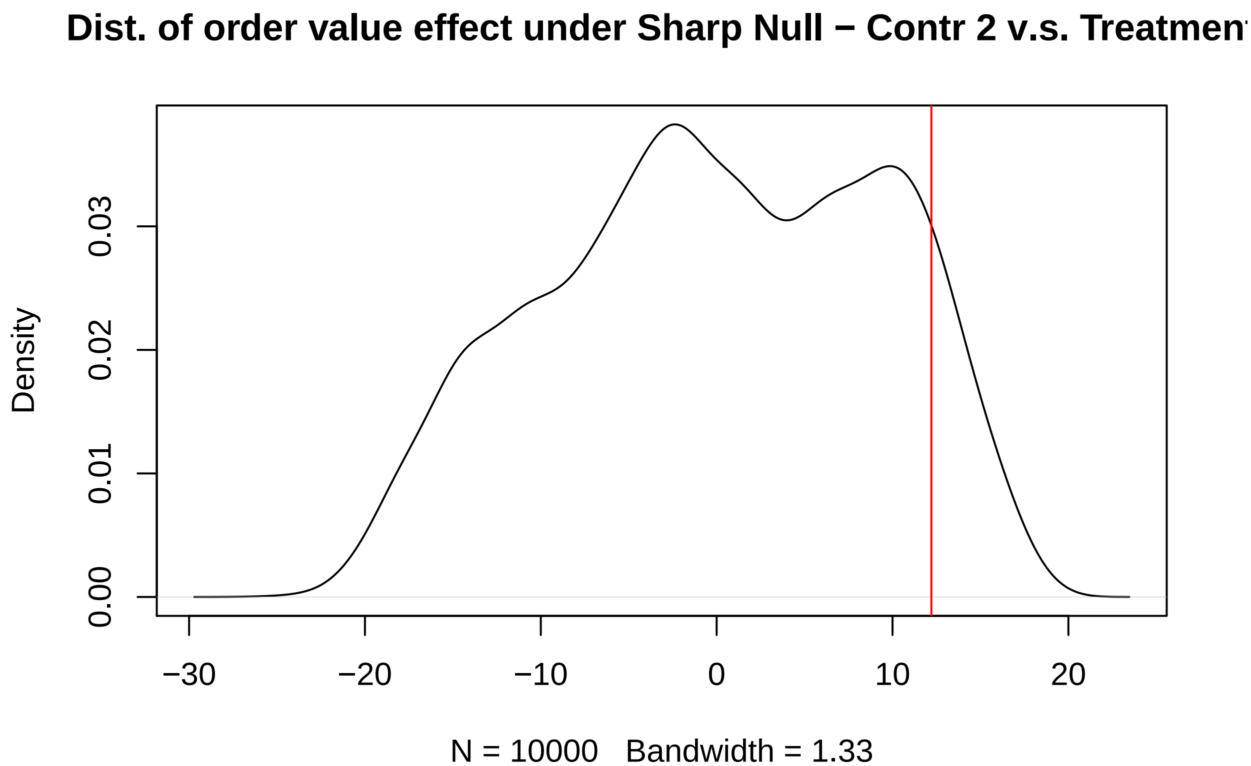
```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          10.4894     4.9454  2.1210 0.040503 *
## treat                 6.6874     7.0534  0.9481 0.349068
## Returning_Customer   31.6478     9.3476  3.3857 0.001662 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# non-parametric test
wilcox.test(SO_cont2$OrderValue,SO_treat$OrderValue)
```

```
## Warning in wilcox.test.default(SO_cont2$OrderValue, SO_treat$OrderValue):
## cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  SO_cont2$OrderValue and SO_treat$OrderValue
## W = 143, p-value = 0.7262
## alternative hypothesis: true location shift is not equal to 0
```

```r
# ramdomization inference
ate <- mean(SO_treat$OrderValue) - mean(SO_cont2$OrderValue)
dist_sharpnull_ordervalue <- rep(0,10000)
for (i in 1:10000) {
  treat <- sample(SO_cont2_treat$treat)
  dist_sharpnull_ordervalue[i] <- mean(SO_cont2_treat[treat==1,]$OrderValue) -
    mean(SO_cont2_treat[treat==0,]$OrderValue)
}
plot(density(dist_sharpnull_ordervalue),
     main="Dist. of order value effect under Sharp Null - Contr 2 v.s. Treatment")
abline(v=ate, col="red")
```




```r
mean(abs(ate) < abs(dist_sharpnull_ordervalue)) # p value under sharp null
```

```
## [1] 0.2198
```

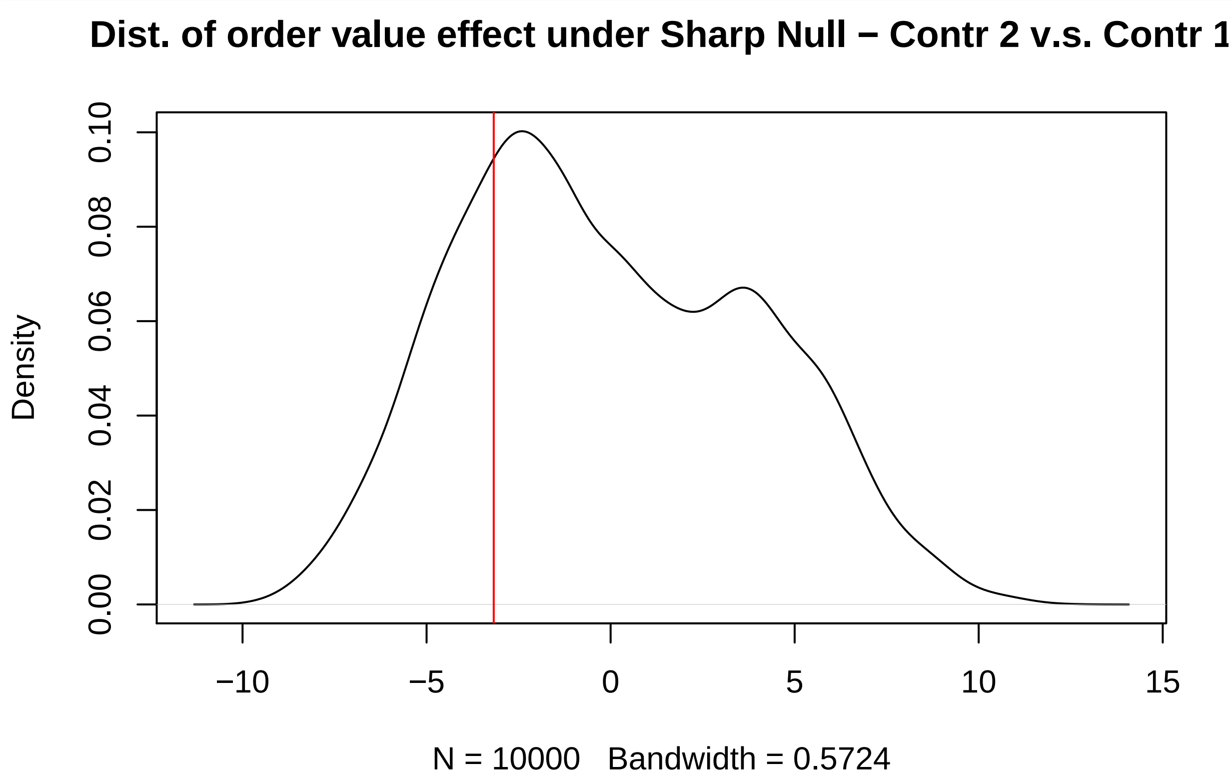
Intertemperal substitution check

```r
wilcox.test(SO_cont2$OrderValue,SO_cont1$OrderValue)
```

```
## Warning in wilcox.test.default(SO_cont2$OrderValue, SO_cont1$OrderValue):
## cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  SO_cont2$OrderValue and SO_cont1$OrderValue
## W = 146.5, p-value = 0.8098
## alternative hypothesis: true location shift is not equal to 0
```

```r
SO_cont1_cont2 <- rbind(SO_cont2,SO_cont1)
ate <- mean(SO_cont2$OrderValue) - mean(SO_cont1$OrderValue)
post_treatment <- c(rep(0,length(SO_cont1$OrderID)), rep(1,length(SO_cont2$OrderID)))
for (i in 1:10000){
        treat <- sample(post_treatment)
        dist_sharpnull_ordervalue[i] <- mean(SO_cont1_cont2[treat==1,]$OrderValue) -
    mean(SO_cont1_cont2[treat==0,]$OrderValue)
}
plot(density(dist_sharpnull_ordervalue),
     main="Dist. of order value effect under Sharp Null - Contr 2 v.s. Contr 1")
abline(v=ate, col="red")
```




```r
mean(abs(ate) < abs(dist_sharpnull_ordervalue)) # p value under sharp null
```

```
## [1] 0.4931
```

Sales value per Buyer

Conversion Rate

average number of items per order