Etsy Data Final Analysis

Ramya Balasubramaniam

4/15/2017

Preparing the Data

```
etsy_data1 <- read.csv(file = "~/downloads/etsy_data/etsy_aft_treat1.csv")
names(etsy_data1)
```

```
##  [1] "Sale.Date"          "Item.Name"          "Buyer"
##  [4] "Quantity"           "Price"              "Coupon.Code"
##  [7] "Coupon.Details"     "Coupon.Discount"    "Order.Shipping"
## [10] "Order.Sales.Tax"    "Item.Total"         "Currency"
## [13] "Transaction.ID"     "Listing.ID"         "Date.Paid"
## [16] "Date.Shipped"       "Ship.Name"          "Ship.Address1"
## [19] "Ship.Address2"      "Ship.City"          "Ship.State"
## [22] "Ship.Zipcode"       "Ship.Country"       "Order.ID"
## [25] "Variations"         "Order.Type"         "Listings.Type"
## [28] "Payment.Type"       "InPerson.Discount" "InPerson.Location"
## [31] "VAT.Paid.by.Buyer"
```

```
etsysales1 <- etsy_data1[, c("Sale.Date", "Quantity", "Price",
    "Buyer", "Item.Total")]
unique(etsysales1$Sale.Date)  ##Ensuring data is from treatment week
```

```
## [1] 3/26/17  3/25/17  3/24/17  3/23/17  3/22/17  3/21/17  3/20/17  03/20/17
## 8 Levels: 03/20/17 3/20/17 3/21/17 3/22/17 3/23/17 3/24/17 ... 3/26/17
```

```
etsysales1[etsysales1$Sale.Date == "03/20/17", "Sale.Date"] <- "3/20/17"
table(etsysales1$Price)
```

```
##
## 5.59 6.99 7.99 9.59 9.99   25   29
##   47    4    4    4    1    1    1
```

```
etsysales1_buyer <- aggregate(x = etsysales1[, c(2:3, 5)], by = list(etsysales1$Buyer),
    FUN = sum)  ##number of items & order value per buyer
etsysales1_date <- aggregate(x = etsysales1[, c(2:3, 5)], by = list(etsysales1$Sale.Date),
    FUN = sum)  ##number of items & order value per day
etsy1_totsale <- sum(etsysales1[, 2] * etsysales1[, 3])  ## total sales in terms of discounted price
etsy1_totsale
```

```
## [1] 452.95
```

```
etsy_data2 <- read.csv(file = "~/downloads/etsy_data/etsy_bef_treat1.csv")
names(etsy_data2)
```

```
##  [1] "Sale.Date"          "Item.Name"          "Buyer"
##  [4] "Quantity"           "Price"              "Coupon.Code"
##  [7] "Coupon.Details"     "Coupon.Discount"    "Order.Shipping"
## [10] "Order.Sales.Tax"    "Item.Total"         "Currency"
## [13] "Transaction.ID"     "Listing.ID"         "Date.Paid"
## [16] "Date.Shipped"       "Ship.Name"          "Ship.Address1"
```

```
## [19] "Ship.Address2"     "Ship.City"         "Ship.State"
## [22] "Ship.Zipcode"      "Ship.Country"      "Order.ID"
## [25] "Variations"        "Order.Type"        "Listings.Type"
## [28] "Payment.Type"      "InPerson.Discount" "InPerson.Location"
## [31] "VAT.Paid.by.Buyer"
```

```r
date.int <- c("03/19/17", "03/18/17", "03/17/17", "03/16/17",
    "03/15/17", "03/14/17", "03/13/17")
etsysales2 <- etsy_data2[which(etsy_data2$Sale.Date %in% date.int),
    c("Sale.Date", "Quantity", "Price", "Buyer", "Item.Total")]
unique(etsysales2$Sale.Date)
```

```
## [1] 03/19/17 03/18/17 03/17/17 03/16/17 03/15/17 03/14/17 03/13/17
## 26 Levels: 03/02/17 03/03/17 03/05/17 03/06/17 03/07/17 ... 03/28/17
```

```r
table(etsysales2$Price)
```

```
##
##  6.99  9.99 11.99 19.99    29
##    23     1     4     1     3
```

```r
etsysales2_buyer <- aggregate(x = etsysales2[, c(2:3, 5)], by = list(etsysales2$Buyer),
    FUN = sum)  ##number of items & order value per buyer
etsysales2_date <- aggregate(x = etsysales2[, c(2:3, 5)], by = list(etsysales2$Sale.Date),
    FUN = sum)  ##number of items & order value per day
etsy2_totsale <- sum(etsysales2[, 2] * etsysales2[, 3])  ## total sales in terms of actual price
etsy2_totsale
```

```
## [1] 325.71
```

```r
etsy_data3 <- read.csv(file = "~/downloads/etsy_data/etsy_final1.csv")
names(etsy_data3)
```

```
##  [1] "Sale.Date"         "Item.Name"         "Buyer"
##  [4] "Quantity"          "Price"             "Coupon.Code"
##  [7] "Coupon.Details"    "Coupon.Discount"   "Order.Shipping"
## [10] "Order.Sales.Tax"   "Item.Total"        "Currency"
## [13] "Transaction.ID"    "Listing.ID"        "Date.Paid"
## [16] "Date.Shipped"      "Ship.Name"         "Ship.Address1"
## [19] "Ship.Address2"     "Ship.City"         "Ship.State"
## [22] "Ship.Zipcode"      "Ship.Country"      "Order.ID"
## [25] "Variations"        "Order.Type"        "Listings.Type"
## [28] "Payment.Type"      "InPerson.Discount" "InPerson.Location"
## [31] "VAT.Paid.by.Buyer"
```

```r
date.int <- c("03/27/17", "03/28/17", "03/29/17", "03/30/17",
    "03/31/17", "04/01/17", "04/02/17")
etsysales3 <- etsy_data3[which(etsy_data3$Sale.Date %in% date.int),
    c("Sale.Date", "Quantity", "Price", "Buyer", "Item.Total")]
unique(etsysales3$Sale.Date)
```

```
## [1] 04/02/17 03/31/17 03/30/17 03/29/17 03/28/17 03/27/17
## 11 Levels: 03/27/17 03/28/17 03/29/17 03/30/17 03/31/17 ... 04/09/17
```

```r
table(etsysales3$Price)
```

```
##
##  6.99  9.99 11.99 19.99
```

```
##    12     4     1     1
```

```
etsysales3_buyer <- aggregate(x = etsysales3[, c(2:3, 5)], by = list(etsysales3$Buyer),
    FUN = sum)  ##number of items & order value per buyer
etsysales3_date <- aggregate(x = etsysales3[, c(2:3, 5)], by = list(etsysales3$Sale.Date),
    FUN = sum)  ##number of items & order value per day
etsy3_totsale <- sum(etsysales3[, 2] * etsysales3[, 3])  ## total sales in terms of actual price
etsy3_totsale
```

```
## [1] 162.81
```

```
etsy1_totquan <- sum(etsysales1[, 2])
etsy2_totquan <- sum(etsysales2[, 2])
etsy3_totquan <- sum(etsysales3[, 2])
```

Tests of significance

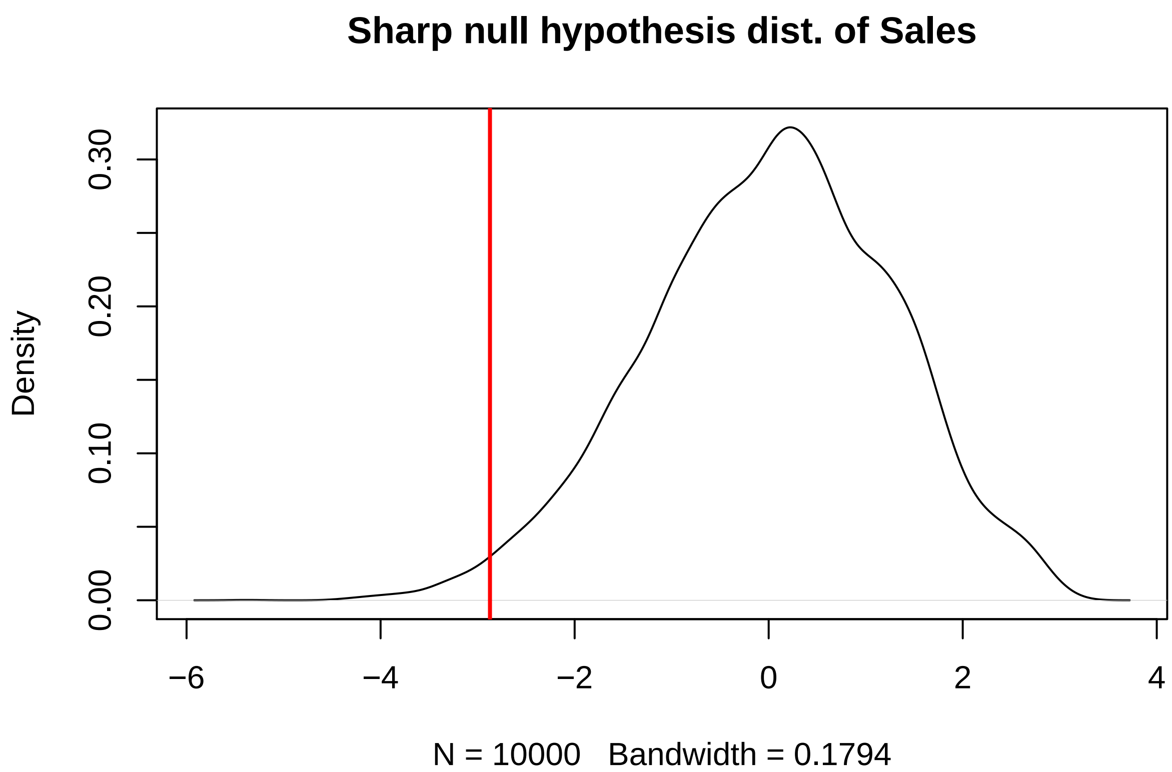
```
a <- c(47, 0, 4, 4, 0)
b <- c(23, 0, 1, 4, 1)
c <- c(12, 0, 4, 1, 1)
wilcox.test(a, b, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  a and b
## V = 5, p-value = 0.4227
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(b, c, paired = T)  ## this was for number of units sold
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  b and c
## V = 4.5, p-value = 0.5862
## alternative hypothesis: true location shift is not equal to 0
```

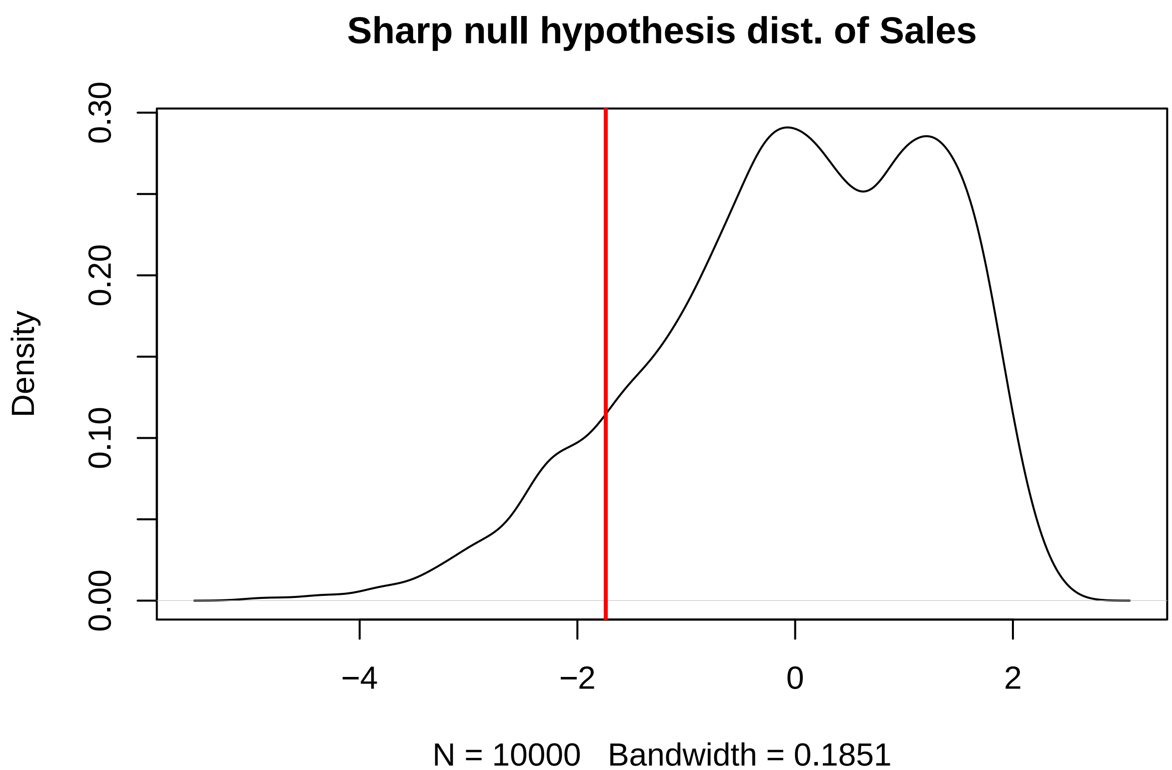
```
e1 <- data.frame(date = c(etsysales2[, 1], etsysales1[, 1]),
    price = c(etsysales2[, 2] * etsysales2[, 3], etsysales1[,
        2] * etsysales1[, 3]), quantity = c(etsysales2[, 2],
        etsysales1[, 2]))
po.control <- e1$price
po.treat <- po.control
treatment <- c(rep(x = 0, each = 32), rep(x = 1, each = 62))
ate <- mean(po.treat[treatment == 1]) - mean(po.control[treatment ==
    0])
etsy_dist1 <- vector(mode = "numeric", length = 10000)
for (i in 1:10000) {
    treat <- sample(x = treatment, size = 94)
    etsy_dist1[i] <- mean(po.treat[treat == 1]) - mean(po.control[treat ==
        0])
}
plot(density(etsy_dist1), main = "Sharp null hypothesis dist. of Sales")
abline(v = ate, lwd = 2, col = "red")
```




```
e1.ttest <- t.test(e1[1:32, "price"], e1[33:94, "price"])
data.frame(SO_TR = mean(po.treat[treatment == 1]), SO_C1 = mean(po.control[treatment ==
    0]), p_val_s = sum(abs(etsy_dist1) >= abs(ate))/length(etsy_dist1),
    p_val_t = e1.ttest$p.value)
```

```
##      SO_TR    SO_C1 p_val_s    p_val_t
## 1 7.305645 10.17844  0.0169 0.03966834
```

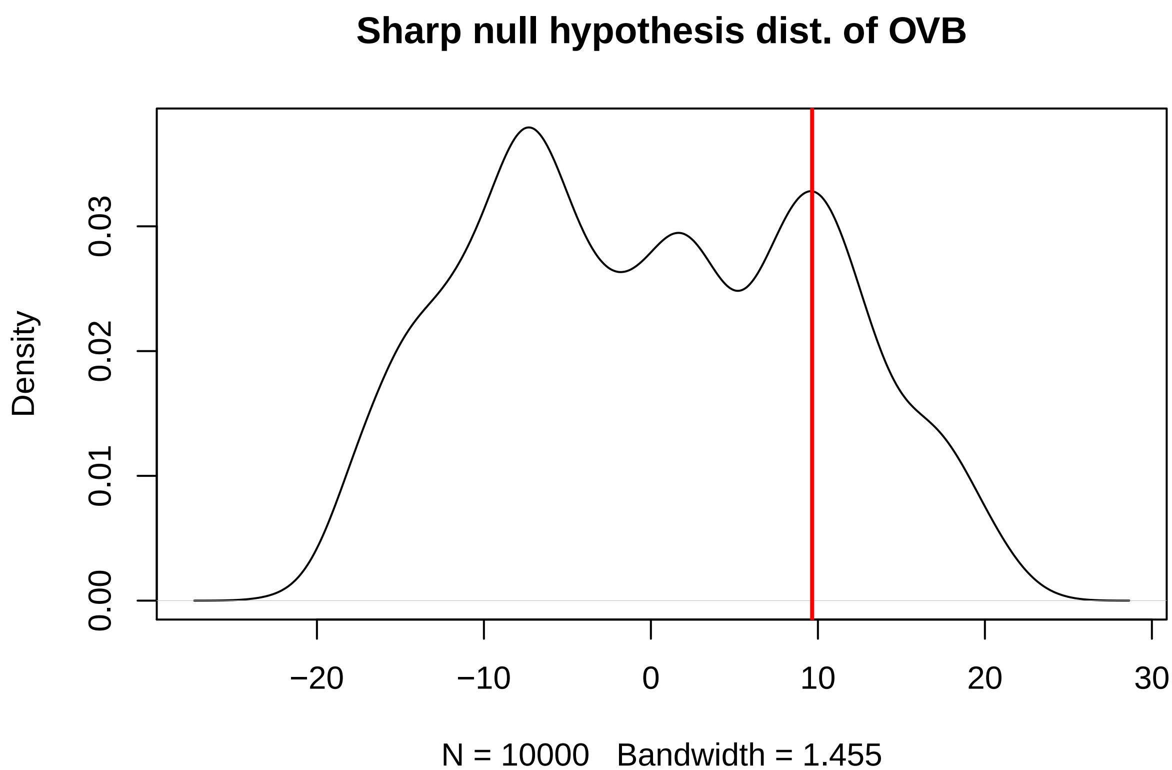
```
e2 <- data.frame(date = c(etsysales3[, 1], etsysales1[, 1]),
    price = c(etsysales3[, 2] * etsysales3[, 3], etsysales1[,
        2] * etsysales1[, 3]), quantity = c(etsysales3[, 2],
        etsysales1[, 2]))
po.control <- e2$price
po.treat <- po.control
treatment <- c(rep(x = 0, each = 18), rep(x = 1, each = 62))
ate1 <- mean(po.treat[treatment == 1]) - mean(po.control[treatment ==
    0])
etsy_dist2 <- vector(mode = "numeric", length = 10000)
for (i in 1:10000) {
    treat <- sample(x = treatment, size = 80)
    etsy_dist2[i] <- mean(po.treat[treat == 1]) - mean(po.control[treat ==
        0])
}
plot(density(etsy_dist2), main = "Sharp null hypothesis dist. of Sales")
abline(v = ate1, lwd = 2, col = "red")
```




```
e2.ttest <- t.test(e2[1:18, "price"], e1[19:80, "price"])
data.frame(SO_TR = mean(po.treat[treatment == 1]), SO_C2 = mean(po.control[treatment ==
    0]), p_val_s = sum(abs(etsy_dist2) >= abs(ate1))/length(etsy_dist2),
    p_val_t = e2.ttest$p.value)
```

```
##      SO_TR SO_C2 p_val_s   p_val_t
## 1 7.305645 9.045   0.172 0.1400535
```

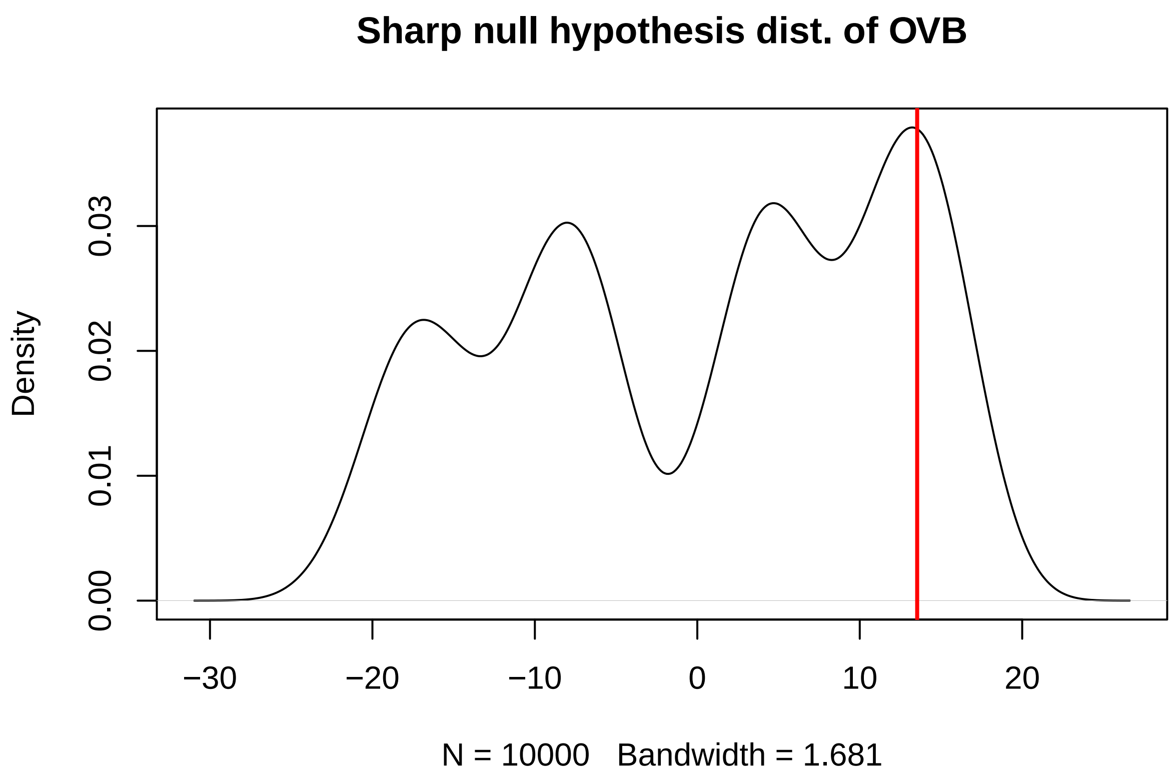
```
e3 <- aggregate(x = etsysales2[, 2] * etsysales2[, 3], by = list(etsysales2$Buyer),
    FUN = sum)
e4 <- aggregate(x = etsysales1[, 2] * etsysales1[, 3], by = list(etsysales1$Buyer),
    FUN = sum)
po.control <- c(e3$x, e4$x)
po.treat <- po.control
treatment <- c(rep(x = 0, each = 21), rep(x = 1, each = 18))
ate2 <- mean(po.treat[treatment == 1]) - mean(po.control[treatment ==
    0])
etsy_dist3 <- vector(mode = "numeric", length = 10000)
for (i in 1:10000) {
    treat <- sample(x = treatment, size = 39)
    etsy_dist3[i] <- mean(po.treat[treat == 1]) - mean(po.control[treat ==
        0])
}
plot(density(etsy_dist3), main = "Sharp null hypothesis dist. of OVB")
abline(v = ate2, lwd = 2, col = "red")
```




```
e3.ttest <- t.test(e3$x, e4$x)
data.frame(OVB_TR = mean(po.treat[treatment == 1]), OVB_C1 = mean(po.control[treatment ==
    0]), p_val_s = sum(abs(etsy_dist3) >= abs(ate2))/length(etsy_dist3),
    p_val_t = e3.ttest$p.value)
```

```
##     OVB_TR OVB_C1 p_val_s   p_val_t
## 1 25.16389  15.51  0.4177 0.3821768
```

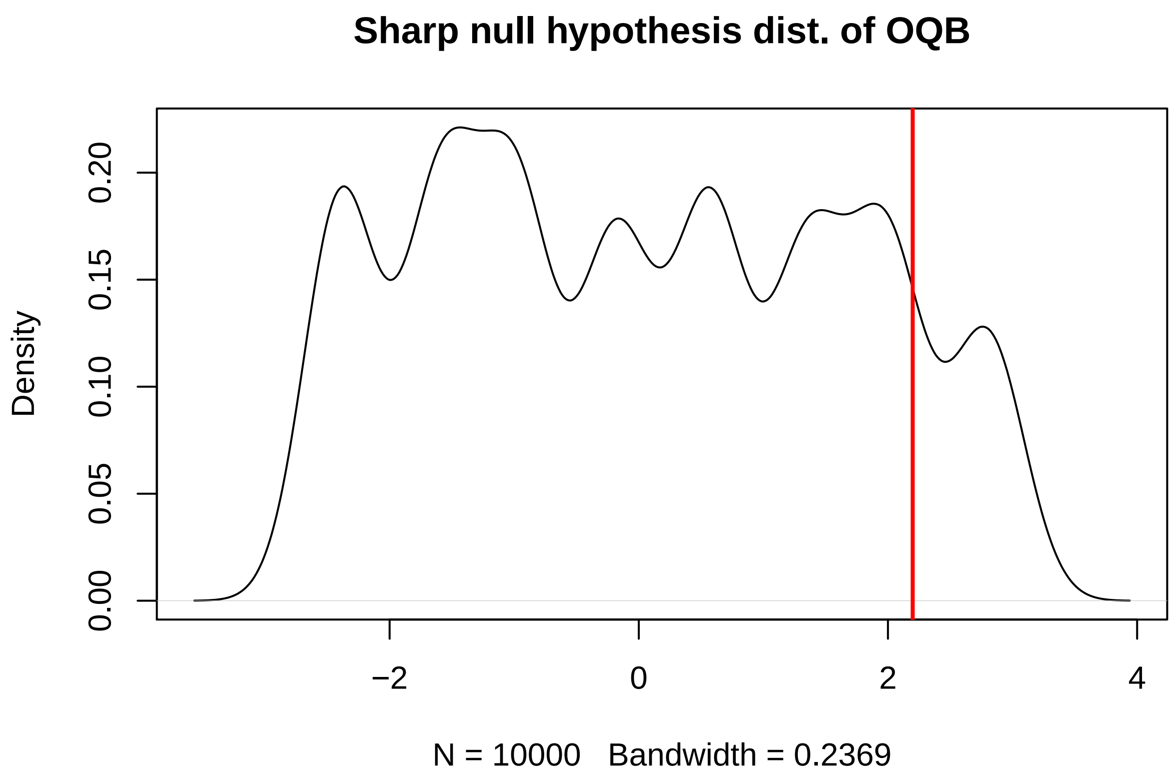
```
e5 <- aggregate(x = etsysales3[, 2] * etsysales3[, 3], by = list(etsysales3$Buyer),
    FUN = sum)
e6 <- aggregate(x = etsysales1[, 2] * etsysales1[, 3], by = list(etsysales1$Buyer),
    FUN = sum)
po.control <- c(e5$x, e6$x)
po.treat <- po.control
treatment <- c(rep(x = 0, each = 14), rep(x = 1, each = 18))
ate3 <- mean(po.treat[treatment == 1]) - mean(po.control[treatment ==
    0])
etsy_dist4 <- vector(mode = "numeric", length = 10000)
for (i in 1:10000) {
    treat <- sample(x = treatment, size = 32)
    etsy_dist4[i] <- mean(po.treat[treat == 1]) - mean(po.control[treat ==
        0])
}
plot(density(etsy_dist4), main = "Sharp null hypothesis dist. of OVB")
abline(v = ate3, lwd = 2, col = "red")
```




```
e4.ttest <- t.test(e5$x, e6$x)
data.frame(OVB_TR = mean(po.treat[treatment == 1]), OVB_C2 = mean(po.control[treatment ==
    0]), p_val_s = sum(abs(etsy_dist4) >= abs(ate3))/length(etsy_dist4),
    p_val_t = e4.ttest$p.value)
```

```
##     OVB_TR   OVB_C2 p_val_s   p_val_t
## 1 25.16389 11.62929  0.3237 0.2076161
```

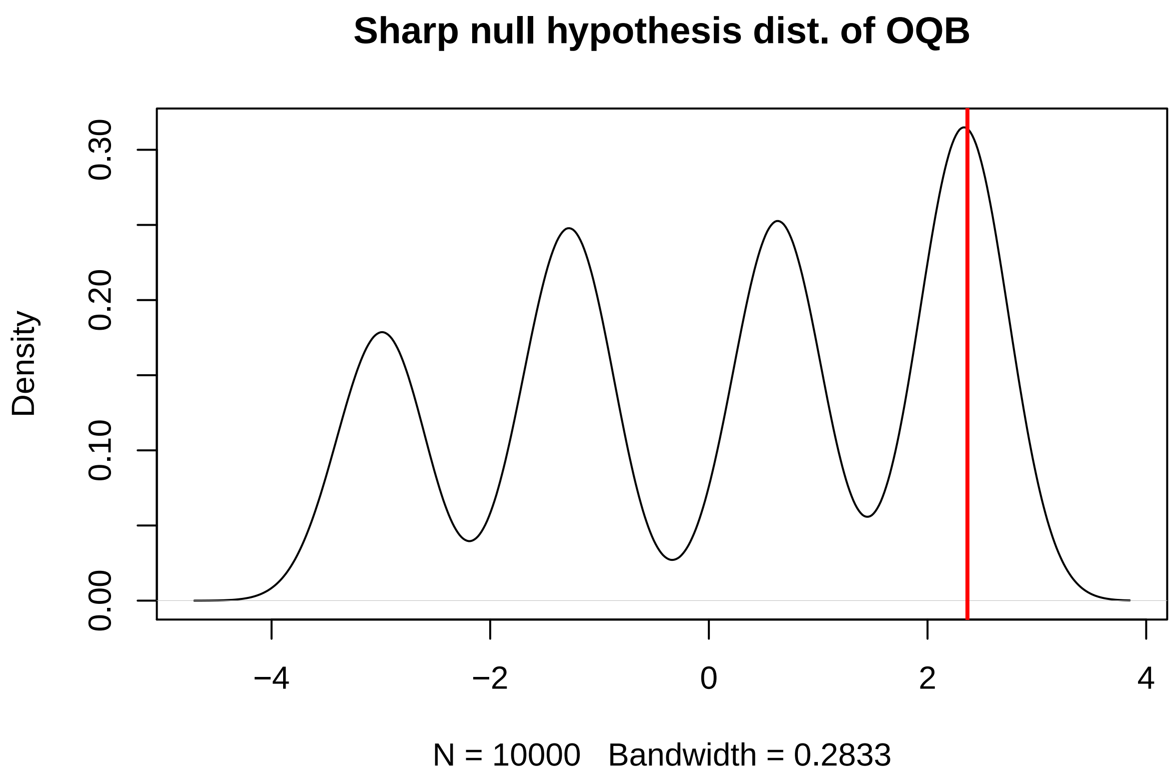
```
e7 <- aggregate(x = etsysales2[, 2], by = list(etsysales2$Buyer),
    FUN = sum)
e8 <- aggregate(x = etsysales1[, 2], by = list(etsysales1$Buyer),
    FUN = sum)
po.control <- c(e7$x, e8$x)
po.treat <- po.control
treatment <- c(rep(x = 0, each = 21), rep(x = 1, each = 18))
ate4 <- mean(po.treat[treatment == 1]) - mean(po.control[treatment ==
    0])
etsy_dist5 <- vector(mode = "numeric", length = 10000)
for (i in 1:10000) {
    treat <- sample(x = treatment, size = 39)
    etsy_dist5[i] <- mean(po.treat[treat == 1]) - mean(po.control[treat ==
        0])
}
plot(density(etsy_dist5), main = "Sharp null hypothesis dist. of OQB")
abline(v = ate4, lwd = 2, col = "red")
```




```r
e5.ttest <- t.test(e7$x, e8$x)
data.frame(OVB_TR = mean(po.treat[treatment == 1]), OVB_C2 = mean(po.control[treatment ==
    0]), p_val_s = sum(abs(etsy_dist5) >= abs(ate4))/length(etsy_dist5),
    p_val_t = e5.ttest$p.value)
```

```
##     OVB_TR  OVB_C2 p_val_s   p_val_t
## 1 3.722222 1.52381  0.2363 0.2295007
```

```r
e9 <- aggregate(x = etsysales3[, 2], by = list(etsysales3$Buyer),
    FUN = sum)
e10 <- aggregate(x = etsysales1[, 2], by = list(etsysales1$Buyer),
    FUN = sum)
po.control <- c(e9$x, e10$x)
po.treat <- po.control
treatment <- c(rep(x = 0, each = 14), rep(x = 1, each = 18))
ate5 <- mean(po.treat[treatment == 1]) - mean(po.control[treatment ==
    0])
etsy_dist6 <- vector(mode = "numeric", length = 10000)
for (i in 1:10000) {
    treat <- sample(x = treatment, size = 32)
    etsy_dist6[i] <- mean(po.treat[treat == 1]) - mean(po.control[treat ==
        0])
}
plot(density(etsy_dist6), main = "Sharp null hypothesis dist. of OQB")
abline(v = ate5, lwd = 2, col = "red")
```




```
e6.ttest <- t.test(e9$x, e10$x)
data.frame(OVB_TR = mean(po.treat[treatment == 1]), OVB_C2 = mean(po.control[treatment ==
    0]), p_val_s = sum(abs(etsy_dist6) >= abs(ate5))/length(etsy_dist6),
    p_val_t = e6.ttest$p.value)
```

```
##     OVB_TR   OVB_C2 p_val_s   p_val_t
## 1 3.722222 1.357143  0.3541 0.1925091
```

Analysis of age data over control and treatment weeks

```
data.age <- read.csv(file = "~/downloads/etsy_data/GA_age.csv")
data.age1 <- read.csv(file = "~/downloads/etsy_data/GA_age_C2.csv")
head(data.age)
```

```
##     Age  X0 X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14
## 1 25-34  49 28 27 34 27 24 27 40 36 27  49  36  34  32  35
## 2 18-24  44 43 47 32 26 14 31 39 23 32  34  17  27  38  27
## 3 35-44  19 12 13  0 16 22  0 11 14 13  17  18  14  11  13
## 4 45-54   0 12  0  0  0  0  0  0  0  0  14  15   0  11   0
## 5 Total 112 95 87 66 69 60 58 90 73 72 114  86  75  92  75
```

```
head(data.age1)
```

```
##     Age X0  X1 X2 X3 X4 X5 X6
## 1 25-34 35  39 29 21 39 25 29
## 2 18-24 27  39 30 24 28 27 34
## 3 35-44 13  17 13  0  0  0  0
## 4 45-54  0  10 15  0 14  0  0
## 5 Total 75 105 87 45 81 52 63
```

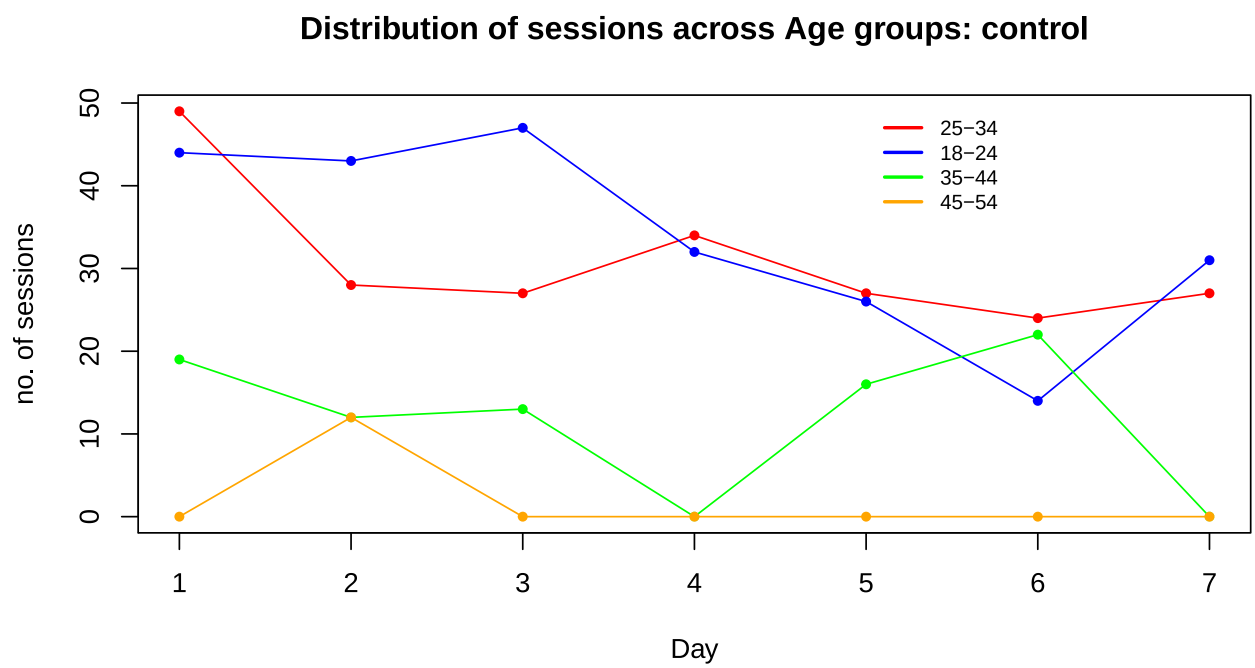
```r
rownames(data.age) <- data.age[, 1]
data.age <- data.age[, -c(1)]
rownames(data.age1) <- data.age1[, 1]
data.age1 <- data.age1[, -c(1)]
head(data.age)
```

```
##        X0 X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14
## 25-34  49 28 27 34 27 24 27 40 36 27  49  36  34  32  35
## 18-24  44 43 47 32 26 14 31 39 23 32  34  17  27  38  27
## 35-44  19 12 13  0 16 22  0 11 14 13  17  18  14  11  13
## 45-54   0 12  0  0  0  0  0  0  0  0  14  15   0  11   0
## Total 112 95 87 66 69 60 58 90 73 72 114  86  75  92  75
```

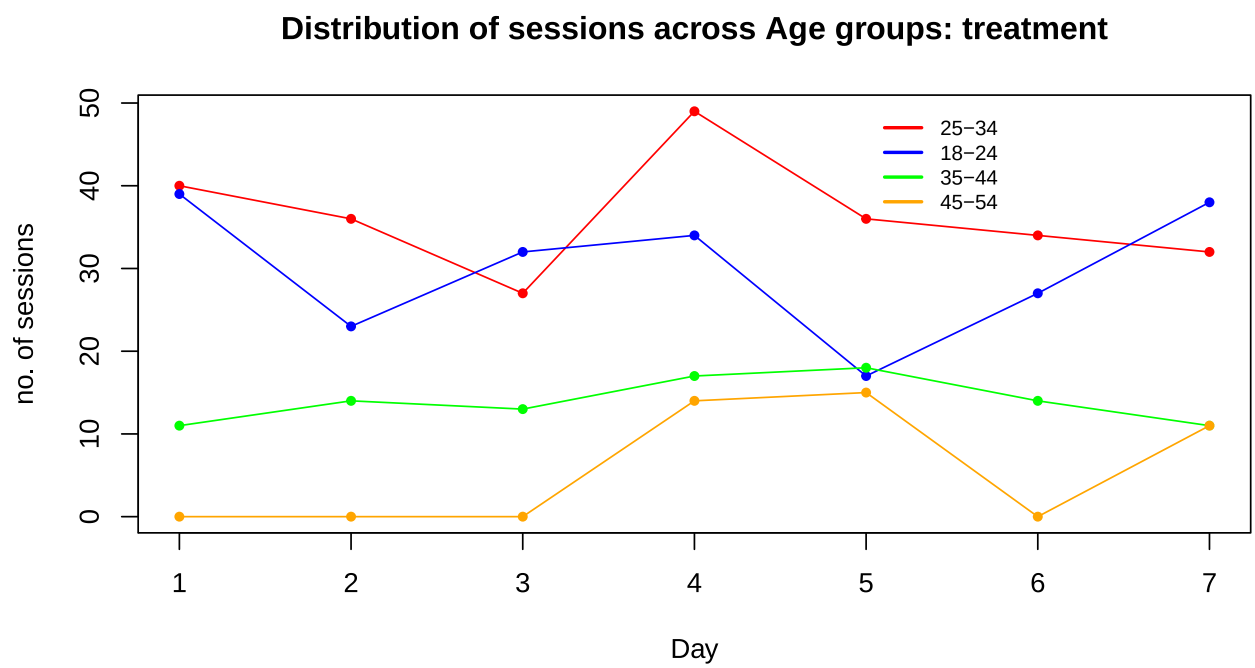
```r
head(data.age1)
```

```
##       X0  X1 X2 X3 X4 X5 X6
## 25-34 35  39 29 21 39 25 29
## 18-24 27  39 30 24 28 27 34
## 35-44 13  17 13  0  0  0  0
## 45-54  0  10 15  0 14  0  0
## Total 75 105 87 45 81 52 63
```

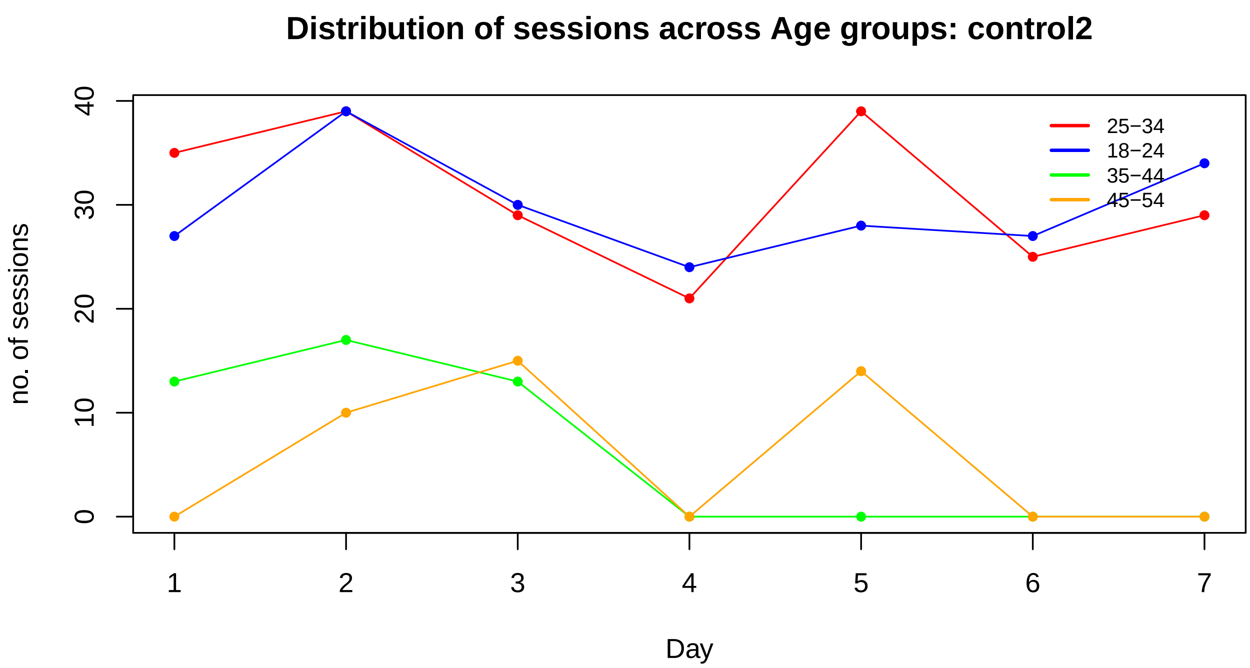
```r
data.age.control <- data.age[, c(1:7)]
data.age.treat <- data.age[, c(8:14)]
plot(c(1:7), data.age.control[1, ], type = "l", main = "Distribution of sessions across Age groups: con
    xlab = "Day", ylab = "no. of sessions", ylim = c(min(data.age.control[1:4,
        ]), max(data.age.control[1:4, ])), col = "red")
points(c(1:7), data.age.control[1, ], pch = 20, col = "red")
lines(c(1:7), data.age.control[2, ], type = "l", col = "blue")
points(c(1:7), data.age.control[2, ], pch = 20, col = "blue")
lines(c(1:7), data.age.control[3, ], type = "l", col = "green")
points(c(1:7), data.age.control[3, ], pch = 20, col = "green")
lines(c(1:7), data.age.control[4, ], type = "l", col = "orange")
points(c(1:7), data.age.control[4, ], pch = 20, col = "orange")
legend(x = 5, y = 50, legend = c("25-34", "18-24", "35-44", "45-54"),
    col = c("red", "blue", "green", "orange"), lwd = 2, bty = "n",
    cex = 0.75)
```




```
plot(c(1:7), data.age.treat[1, ], type = "l", main = "Distribution of sessions across Age groups: treatm
    xlab = "Day", ylab = "no. of sessions", ylim = c(min(data.age.treat[1:4,
        ]), max(data.age.treat[1:4, ])), col = "red")
points(c(1:7), data.age.treat[1, ], pch = 20, col = "red")
lines(c(1:7), data.age.treat[2, ], type = "l", col = "blue")
points(c(1:7), data.age.treat[2, ], pch = 20, col = "blue")
lines(c(1:7), data.age.treat[3, ], type = "l", col = "green")
points(c(1:7), data.age.treat[3, ], pch = 20, col = "green")
lines(c(1:7), data.age.treat[4, ], type = "l", col = "orange")
points(c(1:7), data.age.treat[4, ], pch = 20, col = "orange")
legend(x = 5, y = 50, legend = c("25-34", "18-24", "35-44", "45-54"),
    col = c("red", "blue", "green", "orange"), lwd = 2, bty = "n",
    cex = 0.75)
```




```r
plot(c(1:7), data.age1[1, ], type = "l", main = "Distribution of sessions across Age groups: control2",
    xlab = "Day", ylab = "no. of sessions", ylim = c(min(data.age1[1:4,
        ]), max(data.age1[1:4, ])), col = "red")
points(c(1:7), data.age1[1, ], pch = 20, col = "red")
lines(c(1:7), data.age1[2, ], type = "l", col = "blue")
points(c(1:7), data.age1[2, ], pch = 20, col = "blue")
lines(c(1:7), data.age1[3, ], type = "l", col = "green")
points(c(1:7), data.age1[3, ], pch = 20, col = "green")
lines(c(1:7), data.age1[4, ], type = "l", col = "orange")
points(c(1:7), data.age1[4, ], pch = 20, col = "orange")
legend(x = 6, y = 40, legend = c("25-34", "18-24", "35-44", "45-54"),
    col = c("red", "blue", "green", "orange"), lwd = 2, bty = "n",
    cex = 0.75)
```




```
data.age <- rbind(data.age[2, ], data.age[1, ], data.age[3:5,
    ])
data.age.control <- data.age[, c(1:7)]
data.age.treat <- data.age[, c(8:14)]
data.age1 <- rbind(data.age1[2, ], data.age1[1, ], data.age1[3:5,
    ])
rowSums(data.age.control)
```

```
## 18-24 25-34 35-44 45-54 Total
##   237   216    82    12   547
```

```
rowSums(data.age.treat)
```

```
## 18-24 25-34 35-44 45-54 Total
##   210   254    98    40   602
```

```
rowSums(data.age1)
```

```
## 18-24 25-34 35-44 45-54 Total
##   209   217    43    39   508
```

```
wilcox.test(as.numeric(data.age.control[1, ]), as.numeric(data.age.treat[1,
    ]))  ## Distribution among visitors in age group 18-24 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age.control[1, ]) and as.numeric(data.age.treat[1, ])
## W = 29.5, p-value = 0.5649
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age.control[1, ]), as.numeric(data.age.treat[1,
    ]))  ## Distribution among visitors in age group 18-24 yrs
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  as.numeric(data.age.control[1, ]) and as.numeric(data.age.treat[1, ])
## t = 0.716, df = 10.645, p-value = 0.4894
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.048114 15.762400
## sample estimates:
## mean of x mean of y
##  33.85714  30.00000
```

```
wilcox.test(as.numeric(data.age1[1, ]), as.numeric(data.age.treat[1,
    ]))  ## Distribution among visitors in age group 18-24 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age1[1, ]) and as.numeric(data.age.treat[1, ])
## W = 24, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age1[1, ]), as.numeric(data.age.treat[1,
    ]))  ## Distribution among visitors in age group 18-24 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age1[1, ]) and as.numeric(data.age.treat[1, ])
## t = -0.03959, df = 10.101, p-value = 0.9692
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.171990  7.886275
## sample estimates:
## mean of x mean of y
##  29.85714  30.00000
```

```
wilcox.test(as.numeric(data.age.control[2, ]), as.numeric(data.age.treat[2,
    ]))  ## Distribution among visitors in age group 25-34 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age.control[2, ]) and as.numeric(data.age.treat[2, ])
## W = 11.5, p-value = 0.1051
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age.control[2, ]), as.numeric(data.age.treat[2,
    ]))  ## Distribution among visitors in age group 25-34 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age.control[2, ]) and as.numeric(data.age.treat[2, ])
## t = -1.3072, df = 11.485, p-value = 0.2167
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.521759   3.664616
## sample estimates:
```

```
## mean of x mean of y
##  30.85714  36.28571
```

```r
wilcox.test(as.numeric(data.age1[2, ]), as.numeric(data.age.treat[2,
    ]))  ## Distribution among visitors in age group 25-34 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age1[2, ]) and as.numeric(data.age.treat[2, ])
## W = 15, p-value = 0.2486
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(as.numeric(data.age1[2, ]), as.numeric(data.age.treat[2,
    ]))  ## Distribution among visitors in age group 25-34 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age1[2, ]) and as.numeric(data.age.treat[2, ])
## t = -1.4305, df = 12, p-value = 0.1781
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.336456   2.765028
## sample estimates:
## mean of x mean of y
##  31.00000  36.28571
```

```r
wilcox.test(as.numeric(data.age.control[3, ]), as.numeric(data.age.treat[3,
    ]))  ## Distribution among visitors in age group 35-44 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age.control[3, ]) and as.numeric(data.age.treat[3, ])
## W = 23.5, p-value = 0.9488
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(as.numeric(data.age.control[3, ]), as.numeric(data.age.treat[3,
    ]))  ## Distribution among visitors in age group 35-44 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age.control[3, ]) and as.numeric(data.age.treat[3, ])
## t = -0.66417, df = 7.1536, p-value = 0.5274
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.388149   5.816721
## sample estimates:
## mean of x mean of y
##  11.71429  14.00000
```

```r
wilcox.test(as.numeric(data.age1[3, ]), as.numeric(data.age.treat[3,
    ]))  ## Distribution among visitors in age group 35-44 yrs
```

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

```
##
## data:  as.numeric(data.age1[3, ]) and as.numeric(data.age.treat[3, ])
## W = 10.5, p-value = 0.07873
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age1[3, ]), as.numeric(data.age.treat[3,
    ]))  ## Distribution among visitors in age group 35-44 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age1[3, ]) and as.numeric(data.age.treat[3, ])
## t = -2.5245, df = 7.434, p-value = 0.03768
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -15.1305983  -0.5836874
## sample estimates:
## mean of x mean of y
##  6.142857 14.000000
```

```
wilcox.test(as.numeric(data.age.control[4, ]), as.numeric(data.age.treat[4,
    ]))  ## Distribution among visitors in age group 45-54 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age.control[4, ]) and as.numeric(data.age.treat[4, ])
## W = 17, p-value = 0.2626
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age.control[4, ]), as.numeric(data.age.treat[4,
    ]))  ## Distribution among visitors in age group 45-54 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age.control[4, ]) and as.numeric(data.age.treat[4, ])
## t = -1.2403, df = 10.091, p-value = 0.2429
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.177212   3.177212
## sample estimates:
## mean of x mean of y
##  1.714286  5.714286
```

```
wilcox.test(as.numeric(data.age1[4, ]), as.numeric(data.age.treat[4,
    ]))  ## Distribution among visitors in age group 45-54 yrs
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age1[4, ]) and as.numeric(data.age.treat[4, ])
## W = 24, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age1[4, ]), as.numeric(data.age.treat[4,
    ]))  ## Distribution among visitors in age group 45-54 yrs
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age1[4, ]) and as.numeric(data.age.treat[4, ])
## t = -0.037268, df = 11.997, p-value = 0.9709
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.495041  8.209327
## sample estimates:
## mean of x mean of y
##  5.571429  5.714286
```

```
wilcox.test(as.numeric(data.age.control[5, ]), as.numeric(data.age.treat[5,
    ]))  ## Distribution among visitors in all age groups
```

```
##
##  Wilcoxon rank sum test
##
## data:  as.numeric(data.age.control[5, ]) and as.numeric(data.age.treat[5, ])
## W = 16, p-value = 0.3176
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age.control[5, ]), as.numeric(data.age.treat[5,
    ]))  ## Distribution among visitors in all age groups
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age.control[5, ]) and as.numeric(data.age.treat[5, ])
## t = -0.82637, df = 10.987, p-value = 0.4262
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -28.78704  13.07275
## sample estimates:
## mean of x mean of y
##  78.14286  86.00000
```

```
wilcox.test(as.numeric(data.age1[5, ]), as.numeric(data.age.treat[5,
    ]))  ## Distribution among visitors in all age groups
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.age1[5, ]) and as.numeric(data.age.treat[5, ])
## W = 15.5, p-value = 0.2769
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.age1[5, ]), as.numeric(data.age.treat[5,
    ]))  ## Distribution among visitors in all age groups
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.age1[5, ]) and as.numeric(data.age.treat[5, ])
## t = -1.3876, df = 10.835, p-value = 0.1931
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
## -34.768846   7.911703
## sample estimates:
## mean of x mean of y
##  72.57143  86.00000
```

Analysis of gender data over control and treatment weeks

```r
data.sex <- read.csv(file = "~/downloads/etsy_data/GA_gender.csv")
data.sex1 <- read.csv(file = "~/downloads/etsy_data/GA_gender_C2.csv")
head(data.sex)
```

```
##   gender  X0  X1  X2 X3 X4 X5 X6  X7 X8 X9 X10 X11 X12 X13 X14
## 1 female 110  92  81 83 72 64 72  96 65 73 109  86  87  92  70
## 2   male  26  14  24 14 18 16 13  11 20 13  20  13  17   0  22
## 3  total 136 106 105 97 90 80 85 107 85 86 129  99 104  92  92
```

```r
head(data.sex1)
```

```
##   gender X0  X1 X2 X3 X4 X5 X6
## 1 female 70 100 84 54 93 64 94
## 2   male 22  14 13 14  0 18  0
## 3  total 92 114 97 68 93 82 94
```

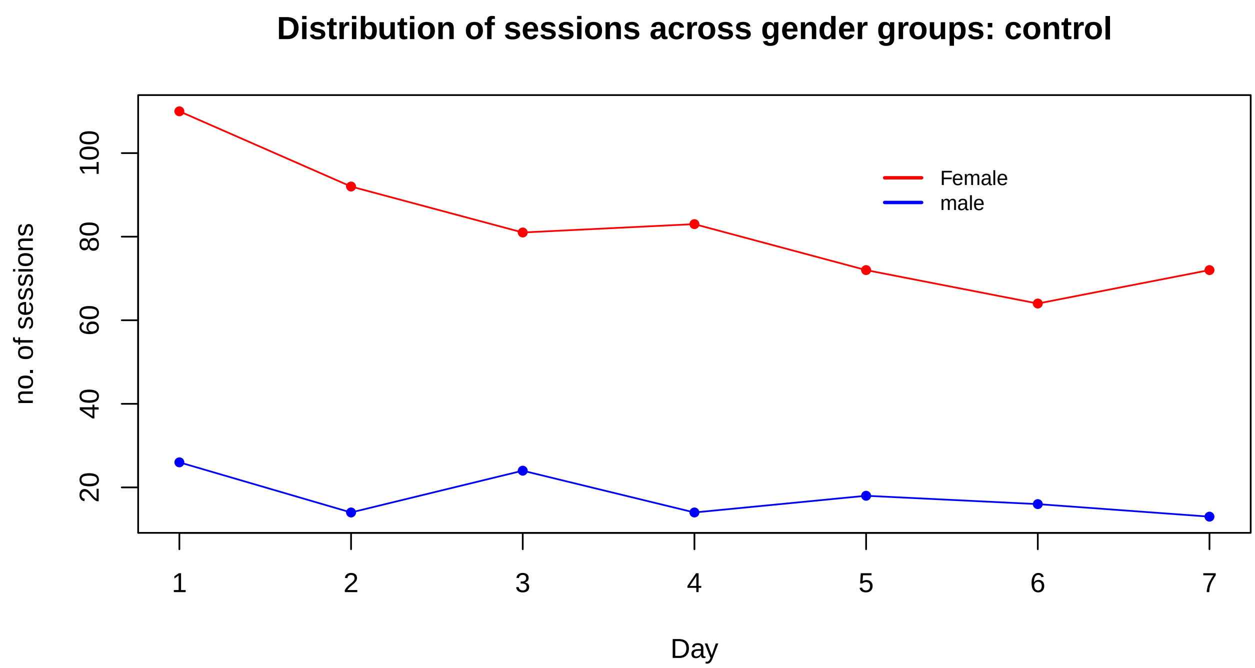
```r
rownames(data.sex) <- data.sex[, 1]
data.sex <- data.sex[, -c(1)]
rownames(data.sex1) <- data.sex1[, 1]
data.sex1 <- data.sex1[, -c(1)]
head(data.sex)
```

```
##         X0  X1  X2 X3 X4 X5 X6  X7 X8 X9 X10 X11 X12 X13 X14
## female 110  92  81 83 72 64 72  96 65 73 109  86  87  92  70
## male    26  14  24 14 18 16 13  11 20 13  20  13  17   0  22
## total  136 106 105 97 90 80 85 107 85 86 129  99 104  92  92
```

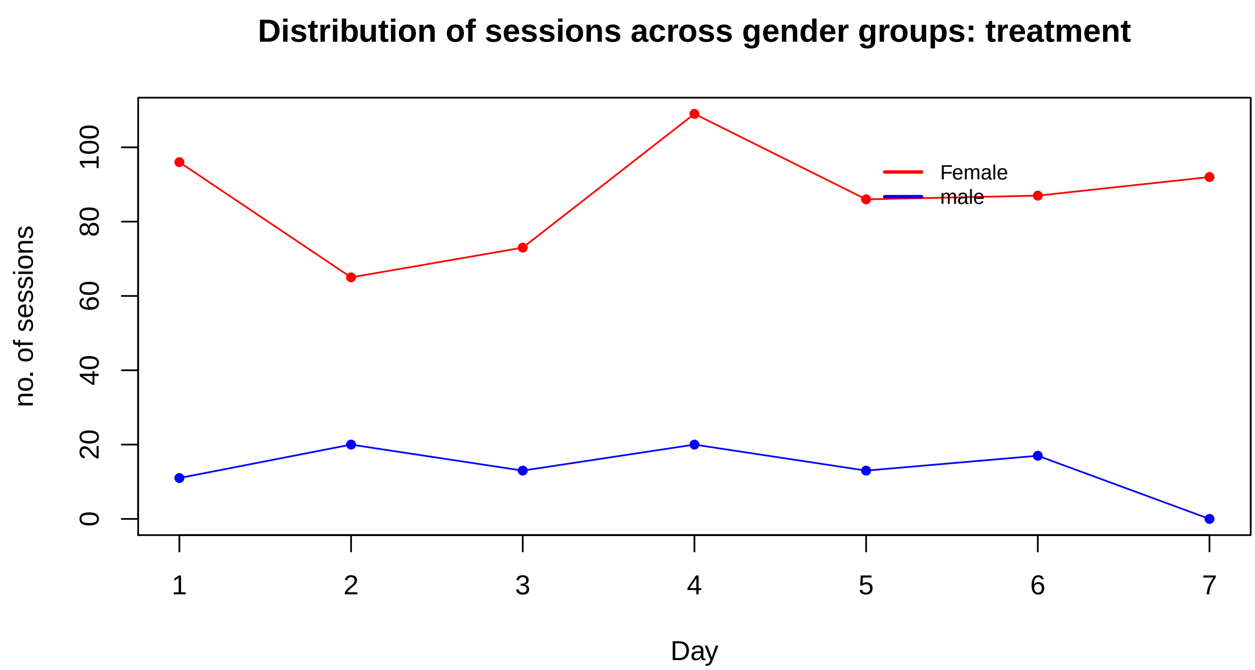
```r
head(data.sex1)
```

```
##        X0  X1 X2 X3 X4 X5 X6
## female 70 100 84 54 93 64 94
## male   22  14 13 14  0 18  0
## total  92 114 97 68 93 82 94
```

```r
data.sex.control <- data.sex[, c(1:7)]
data.sex.treat <- data.sex[, c(8:14)]
plot(c(1:7), data.sex.control[1, ], type = "l", main = "Distribution of sessions across gender groups: 
    xlab = "Day", ylab = "no. of sessions", ylim = c(min(data.sex.control[1:2,
        ]), max(data.sex.control[1:2, ])), col = "red")
points(c(1:7), data.sex.control[1, ], pch = 20, col = "red")
lines(c(1:7), data.sex.control[2, ], type = "l", col = "blue")
points(c(1:7), data.sex.control[2, ], pch = 20, col = "blue")
legend(x = 5, y = 100, legend = c("Female", "male"), col = c("red",
    "blue"), lwd = 2, bty = "n", cex = 0.75)
```

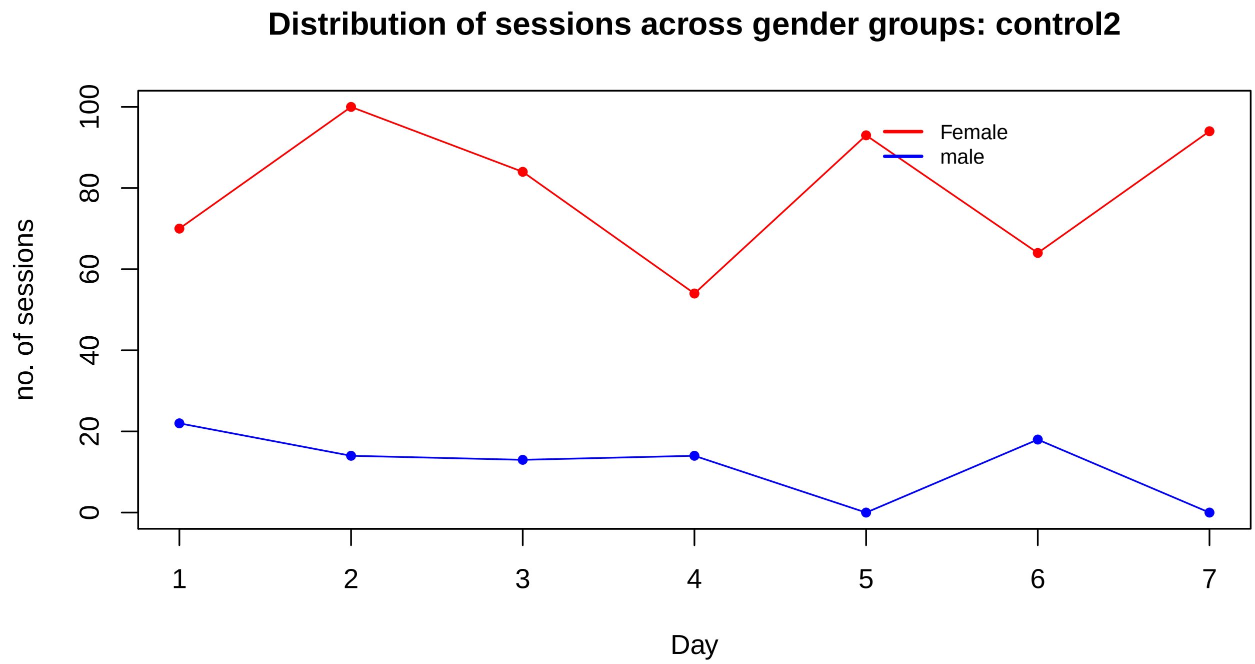



```
plot(c(1:7), data.sex.treat[1, ], type = "l", main = "Distribution of sessions across gender groups: tre
    xlab = "Day", ylab = "no. of sessions", ylim = c(min(data.sex.treat[1:2,
        ]), max(data.sex.treat[1:2, ])), col = "red")
points(c(1:7), data.sex.treat[1, ], pch = 20, col = "red")
lines(c(1:7), data.sex.treat[2, ], type = "l", col = "blue")
points(c(1:7), data.sex.treat[2, ], pch = 20, col = "blue")
legend(x = 5, y = 100, legend = c("Female", "male"), col = c("red",
    "blue"), lwd = 2, bty = "n", cex = 0.75)
```




```
plot(c(1:7), data.sex1[1, ], type = "l", main = "Distribution of sessions across gender groups: control2
    xlab = "Day", ylab = "no. of sessions", ylim = c(min(data.sex1[1:2,
        ]), max(data.sex1[1:2, ])), col = "red")
```

```r
points(c(1:7), data.sex1[1, ], pch = 20, col = "red")
lines(c(1:7), data.sex1[2, ], type = "l", col = "blue")
points(c(1:7), data.sex1[2, ], pch = 20, col = "blue")
legend(x = 5, y = 100, legend = c("Female", "male"), col = c("red",
    "blue"), lwd = 2, bty = "n", cex = 0.75)
```




```r
data.sex.control <- data.sex[, c(1:7)]
data.sex.treat <- data.sex[, c(8:14)]
wilcox.test(as.numeric(data.sex.control[1, ]), as.numeric(data.sex.treat[1,
    ]))  ## Distribution among females
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.sex.control[1, ]) and as.numeric(data.sex.treat[1, ])
## W = 17.5, p-value = 0.4052
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(as.numeric(data.sex.control[1, ]), as.numeric(data.sex.treat[1,
    ]))  ## Distribution among females
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.sex.control[1, ]) and as.numeric(data.sex.treat[1, ])
## t = -0.60795, df = 11.968, p-value = 0.5546
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -22.26967  12.55539
## sample estimates:
## mean of x mean of y
##  82.00000  86.85714
```

```r
wilcox.test(as.numeric(data.sex.control[2, ]), as.numeric(data.sex.treat[2,
    ]))  ## Distribution among males
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.sex.control[2, ]) and as.numeric(data.sex.treat[2, ])
## W = 34, p-value = 0.247
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(as.numeric(data.sex.control[2, ]), as.numeric(data.sex.treat[2,
    ]))  ## Distribution among males
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.sex.control[2, ]) and as.numeric(data.sex.treat[2, ])
## t = 1.3581, df = 11.13, p-value = 0.2013
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.738135 11.595277
## sample estimates:
## mean of x mean of y
##  17.85714  13.42857
```

```r
wilcox.test(as.numeric(data.sex.control[3, ]), as.numeric(data.sex.treat[3,
    ]))  ## Distribution among all visitors
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.sex.control[3, ]) and as.numeric(data.sex.treat[3, ])
## W = 22.5, p-value = 0.8478
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(as.numeric(data.sex1[1, ]), as.numeric(data.sex.treat[1,
    ]))  ## Distribution among females
```

```
##
##  Wilcoxon rank sum test
##
## data:  as.numeric(data.sex1[1, ]) and as.numeric(data.sex.treat[1, ])
## W = 19, p-value = 0.535
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(as.numeric(data.sex1[1, ]), as.numeric(data.sex.treat[1,
    ]))  ## Distribution among females
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.sex1[1, ]) and as.numeric(data.sex.treat[1, ])
## t = -0.81693, df = 11.64, p-value = 0.4304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -25.73365  11.73365
## sample estimates:
```

```
## mean of x mean of y
##  79.85714  86.85714
```

```
wilcox.test(as.numeric(data.sex1[2, ]), as.numeric(data.sex.treat[2,
    ]))  ## Distribution among males
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  as.numeric(data.sex1[2, ]) and as.numeric(data.sex.treat[2, ])
## W = 24, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(as.numeric(data.sex1[2, ]), as.numeric(data.sex.treat[2,
    ]))  ## Distribution among males
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(data.sex1[2, ]) and as.numeric(data.sex.treat[2, ])
## t = -0.44935, df = 11.523, p-value = 0.6615
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.903698   7.189412
## sample estimates:
## mean of x mean of y
##  11.57143  13.42857
```

Analysis of data langauge wise for treatment & control weeks

Since we are having 58 different countries/ categories we use Matched pair Wilcoxon test to see if the distribution is identical.

```
data.lang.control <- read.csv(file = "~/downloads/etsy_data/GA_lang_control.csv")
data.lang.treat <- read.csv(file = "~/downloads/etsy_data/GA_lang_treat.csv")
data.lang.C2 <- read.csv(file = "~/downloads/etsy_data/GA_lang_C2.csv")
lang <- intersect(data.lang.control$Language, data.lang.treat$Language)
data1 <- subset(data.lang.control, data.lang.control$Language %in%
    lang)
data2 <- subset(data.lang.treat, data.lang.treat$Language %in%
    lang)
data1 <- data1[order(data1$Language), 1:2]
data2 <- data2[order(data2$Language), 1:2]
data.lang <- cbind(data1, data2$Sessions)
names(data.lang) <- c("Langauge", "Sess.control", "Sess.treat")
head(data.lang)
```

```
##     Langauge Sess.control Sess.treat
## 51 (not set)            1          1
## 24        ar            5          5
## 41        bg            2          4
## 52        cs            1          4
## 35     cs-cz            3          2
## 53     da-dk            1          4
```

```r
wilcox.test(data.lang$Sess.control, data.lang$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.lang$Sess.control and data.lang$Sess.treat
## V = 602.5, p-value = 0.9242
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.lang$Sess.control, data.lang$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.lang$Sess.control and data.lang$Sess.treat
## t = -0.90231, df = 57, p-value = 0.3707
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.384866  1.660728
## sample estimates:
## mean of the differences
##               -1.362069
```

There are 57 languages common in the two weeks C2 and treatment weeks.

```r
lang1 <- intersect(data.lang.C2$Language, data.lang.treat$Language)
data1 <- subset(data.lang.C2, data.lang.C2$Language %in% lang1)
data2 <- subset(data.lang.treat, data.lang.treat$Language %in%
    lang1)
data1 <- data1[order(data1$Language), 1:2]
data2 <- data2[order(data2$Language), 1:2]
data.lang1 <- cbind(data1, data2$Sessions)
names(data.lang1) <- c("Langauge", "Sess.control", "Sess.treat")
head(data.lang1)
```

```
##    Langauge Sess.control Sess.treat
## 38       ar            2          5
## 39    ca-es            2          4
## 54    cs-cz            1          2
## 40    da-dk            2          4
## 7        de           29         42
## 28    de-at            5          9
```

```r
wilcox.test(data.lang1$Sess.control, data.lang1$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.lang1$Sess.control and data.lang1$Sess.treat
## V = 343.5, p-value = 0.04948
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.lang1$Sess.control, data.lang1$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
```

```
## data:  data.lang1$Sess.control and data.lang1$Sess.treat
## t = -1.7369, df = 56, p-value = 0.08791
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.3445003  0.3094126
## sample estimates:
## mean of the differences
##               -2.017544
```

Analysis of data country wise for treatment & control weeks

Since we are having 67 different countries/ categories we use Matched pair Wilconxon test to see if the distribution is identical.

```r
data.country.control <- read.csv(file = "~/downloads/etsy_data/GA_country_control.csv")
data.country.treat <- read.csv(file = "~/downloads/etsy_data/GA_country_treat.csv")
data.country.C2 <- read.csv(file = "~/downloads/etsy_data/GA_country_C2.csv")
country <- intersect(data.country.control$Country, data.country.treat$Country)
data1 <- subset(data.country.control, data.country.control$Country %in%
    country)
data2 <- subset(data.country.treat, data.country.treat$Country %in%
    country)
data1 <- data1[order(data1$Country), 1:2]
data2 <- data2[order(data2$Country), 1:2]
data.country <- cbind(data1, data2$Sessions)
names(data.country) <- c("Country", "Sess.control", "Sess.treat")
head(data.country)
```

```
##      Country Sess.control Sess.treat
## 27 Argentina            9          5
## 13 Australia           17         17
## 12   Austria           20         11
## 30   Belgium            8          9
## 51   Bolivia            2          3
## 5     Brazil           43         59
```

```r
wilcox.test(data.country$Sess.control, data.country$Sess.treat,
    paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.country$Sess.control and data.country$Sess.treat
## V = 869, p-value = 0.9066
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.country$Sess.control, data.country$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.country$Sess.control and data.country$Sess.treat
## t = -1.0864, df = 66, p-value = 0.2812
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -3.473033  1.025272
## sample estimates:
## mean of the differences
##               -1.223881
```

There are 63 common countries.

```
country1 <- intersect(data.country.C2$Country, data.country.treat$Country)
data1 <- subset(data.country.C2, data.country.C2$Country %in%
    country1)
data2 <- subset(data.country.treat, data.country.treat$Country %in%
    country1)
data1 <- data1[order(data1$Country), 1:2]
data2 <- data2[order(data2$Country), 1:2]
data.country1 <- cbind(data1, data2$Sessions)
names(data.country1) <- c("Country", "Sess.control", "Sess.treat")
head(data.country1)
```

```
##       Country Sess.control Sess.treat
## 26  Argentina            8          5
## 13  Australia           18         17
## 24    Austria            9         11
## 63 Bangladesh            1          1
## 30    Belgium            6          9
## 40    Bolivia            4          3
```

```
wilcox.test(data.country1$Sess.control, data.country1$Sess.treat,
    paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.country1$Sess.control and data.country1$Sess.treat
## V = 660, p-value = 0.2608
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(data.country1$Sess.control, data.country1$Sess.treat,
    paired = T)
```

```
##
##  Paired t-test
##
## data:  data.country1$Sess.control and data.country1$Sess.treat
## t = -1.5184, df = 62, p-value = 0.134
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.0814870  0.5576774
## sample estimates:
## mean of the differences
##               -1.761905
```

Analysis of data affinity wise for treatment & control weeks

We observe that the distributions across affinity are identical for the two weeks.There are 94 common afinity groups.

```r
data.aff.control <- read.csv(file = "~/downloads/etsy_data/GA_Affinity_control.csv")
data.aff.treat <- read.csv(file = "~/downloads/etsy_data/GA_Affinity_treat.csv")
data.aff.C2 <- read.csv(file = "~/downloads/etsy_data/GA_Affinity_C2.csv")
aff <- intersect(data.aff.control$Affinity.Category..reach.,
    data.aff.treat$Affinity.Category..reach.)
data1 <- subset(data.aff.control, data.aff.control$Affinity.Category..reach. %in%
    aff)
data2 <- subset(data.aff.treat, data.aff.treat$Affinity.Category..reach. %in%
    aff)
data1 <- data1[order(data1$Affinity.Category..reach.), 1:2]
data2 <- data2[order(data2$Affinity.Category..reach.), 1:2]
data.aff <- cbind(data1, data2$Sessions)
names(data.aff) <- c("Affinity", "Sess.control", "Sess.treat")
head(data.aff)
```

```
##                                                    Affinity Sess.control
## 19                                  Art & Theater Aficionados          252
## 38                                          Auto Enthusiasts          164
## 75                     Auto Enthusiasts/Motorcycle Enthusiasts           39
## 65 Auto Enthusiasts/Performance & Luxury Vehicle Enthusiasts           54
## 87                    Auto Enthusiasts/Truck & SUV Enthusiasts           29
## 36                                            Avid Investors          171
##    Sess.treat
## 19        227
## 38        194
## 75         32
## 65         54
## 87         24
## 36        144
```

```r
wilcox.test(data.aff$Sess.control, data.aff$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.aff$Sess.control and data.aff$Sess.treat
## V = 1665.5, p-value = 0.1684
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.aff$Sess.control, data.aff$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.aff$Sess.control and data.aff$Sess.treat
## t = -1.5184, df = 93, p-value = 0.1323
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.2851248  0.8383163
## sample estimates:
## mean of the differences
##               -2.723404
```

There are 90 common affinity groups.

```r
aff1 <- intersect(data.aff.C2$Affinity.Category..reach., data.aff.treat$Affinity.Category..reach.)
data1 <- subset(data.aff.C2, data.aff.C2$Affinity.Category..reach. %in%
    aff1)
data2 <- subset(data.aff.treat, data.aff.treat$Affinity.Category..reach. %in%
    aff1)
data1 <- data1[order(data1$Affinity.Category..reach.), 1:2]
data2 <- data2[order(data2$Affinity.Category..reach.), 1:2]
data.aff1 <- cbind(data1, data2$Sessions)
names(data.aff1) <- c("Affinity", "Sess.control", "Sess.treat")
head(data.aff1)
```

```
##                                                      Affinity Sess.control
## 27                              Art & Theater Aficionados          168
## 34                                       Auto Enthusiasts          141
## 79                  Auto Enthusiasts/Motorcycle Enthusiasts           24
## 71 Auto Enthusiasts/Performance & Luxury Vehicle Enthusiasts           31
## 84                 Auto Enthusiasts/Truck & SUV Enthusiasts           18
## 29                                         Avid Investors          159
##    Sess.treat
## 27        227
## 34        194
## 79         32
## 71         54
## 84         24
## 29        144
```

```r
wilcox.test(data.aff1$Sess.control, data.aff1$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.aff1$Sess.control and data.aff1$Sess.treat
## V = 115, p-value = 7.587e-15
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.aff1$Sess.control, data.aff1$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.aff1$Sess.control and data.aff1$Sess.treat
## t = -11.772, df = 89, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -38.68678 -27.51322
## sample estimates:
## mean of the differences
##                   -33.1
```

Analysis of data In-Market Segment wise for treatment & control weeks

We observe that the distributions across In-Market Segment are identical for the two weeks.There are 114 common IMS groups.

```r
data.ims.control <- read.csv(file = "~/downloads/etsy_data/GA_IMS_control.csv")
data.ims.treat <- read.csv(file = "~/downloads/etsy_data/GA_IMS_treat.csv")
data.ims.C2 <- read.csv(file = "~/downloads/etsy_data/GA_IMS_C2.csv")
ims <- intersect(data.ims.control$In.Market.Segment, data.ims.treat$In.Market.Segment)
data1 <- subset(data.ims.control, data.ims.control$In.Market.Segment %in%
    ims)
data2 <- subset(data.ims.treat, data.ims.treat$In.Market.Segment %in%
    ims)
data1 <- data1[order(data1$In.Market.Segment), 1:2]
data2 <- data2[order(data2$In.Market.Segment), 1:2]
data.ims <- cbind(data1, data2$Sessions)
names(data.ims) <- c("IMS", "Sess.control", "Sess.treat")
head(data.ims)
```

```
##                                                       IMS Sess.control
## 15                                  Apparel & Accessories          100
## 106                      Apparel & Accessories/Activewear           14
## 98                         Apparel & Accessories/Costumes           16
## 123          Apparel & Accessories/Formal Wear/Bridal Wear           12
## 50                         Apparel & Accessories/Handbags           43
## 28  Apparel & Accessories/Jewelry & Watches/Fine Jewelry           72
##     Sess.treat
## 15          93
## 106         16
## 98          14
## 123         13
## 50          38
## 28          82
```

```r
wilcox.test(data.ims$Sess.control, data.ims$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.ims$Sess.control and data.ims$Sess.treat
## V = 2507, p-value = 0.1038
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.ims$Sess.control, data.ims$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.ims$Sess.control and data.ims$Sess.treat
## t = -1.8129, df = 113, p-value = 0.0725
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.6899644  0.1636486
## sample estimates:
## mean of the differences
##               -1.763158
```

There are 114 common In-market segment groups in the control week 2 and treatment week.

```r
ims1 <- intersect(data.ims.C2$In.Market.Segment, data.ims.treat$In.Market.Segment)
data1 <- subset(data.ims.C2, data.ims.C2$In.Market.Segment %in%
```

```r
    ims1)
data2 <- subset(data.ims.treat, data.ims.treat$In.Market.Segment %in%
    ims1)
data1 <- data1[order(data1$In.Market.Segment), 1:2]
data2 <- data2[order(data2$In.Market.Segment), 1:2]
data.ims1 <- cbind(data1, data2$Sessions)
names(data.ims1) <- c("IMS", "Sess.control", "Sess.treat")
head(data.ims1)
```

```
##                                                       IMS Sess.control
## 16                                 Apparel & Accessories           82
## 83                       Apparel & Accessories/Activewear           18
## 102          Apparel & Accessories/Formal Wear/Bridal Wear           13
## 34                         Apparel & Accessories/Handbags           46
## 25  Apparel & Accessories/Jewelry & Watches/Fine Jewelry           64
## 68      Apparel & Accessories/Jewelry & Watches/Watches           25
##     Sess.treat
## 16          93
## 83          16
## 102         13
## 34          38
## 25          82
## 68          37
```

```r
wilcox.test(data.ims1$Sess.control, data.ims1$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.ims1$Sess.control and data.ims1$Sess.treat
## V = 975.5, p-value = 5.774e-10
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.ims1$Sess.control, data.ims1$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.ims1$Sess.control and data.ims1$Sess.treat
## t = -7.2252, df = 113, p-value = 6.274e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.126559  -5.768177
## sample estimates:
## mean of the differences
##               -7.947368
```

Analysis of data Other categories wise for treatment & control weeks

We observe that the distributions across other categories are identical for the two weeks.There are 97 common OC groups.

```r
data.OC.control <- read.csv(file = "~/downloads/etsy_data/GA_OC_control.csv")
data.OC.treat <- read.csv(file = "~/downloads/etsy_data/GA_OC_treat.csv")
data.OC.C2 <- read.csv(file = "~/downloads/etsy_data/GA_OC_C2.csv")
```

```r
OC <- intersect(data.OC.control$Other.Category, data.OC.treat$Other.Category)
data1 <- subset(data.OC.control, data.OC.control$Other.Category %in%
    OC)
data2 <- subset(data.OC.treat, data.OC.treat$Other.Category %in%
    OC)
data1 <- data1[order(data1$Other.Category), 1:2]
data2 <- data2[order(data2$Other.Category), 1:2]
data.OC <- cbind(data1, data2$Sessions)
names(data.OC) <- c("OC", "Sess.control", "Sess.treat")
head(data.OC)
```

```
##                                                                           OC
## 2                         Arts & Entertainment/Celebrities & Entertainment News
## 16                        Arts & Entertainment/Comics & Animation/Anime & Manga
## 58                               Arts & Entertainment/Comics & Animation/Comics
## 41 Arts & Entertainment/Events & Listings/Movie Listings & Theater Showtimes
## 67                  Arts & Entertainment/Fun & Trivia/Fun Tests & Silly Surveys
## 73                                                   Arts & Entertainment/Humor
##    Sess.control Sess.treat
## 2           185        176
## 16           54         44
## 58           20         23
## 41           29         21
## 67           18         27
## 73           17         16
```

```r
wilcox.test(data.OC$Sess.control, data.OC$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.OC$Sess.control and data.OC$Sess.treat
## V = 1809.5, p-value = 0.1997
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.OC$Sess.control, data.OC$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.OC$Sess.control and data.OC$Sess.treat
## t = -0.87688, df = 96, p-value = 0.3827
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.557123  0.990113
## sample estimates:
## mean of the differences
##              -0.7835052
```

For the second control week, we have 95 common other categories groups.

```r
OC1 <- intersect(data.OC.C2$Other.Category, data.OC.treat$Other.Category)
data1 <- subset(data.OC.C2, data.OC.C2$Other.Category %in% OC1)
data2 <- subset(data.OC.treat, data.OC.treat$Other.Category %in%
    OC1)
data1 <- data1[order(data1$Other.Category), 1:2]
```

```r
data2 <- data2[order(data2$Other.Category), 1:2]
data.OC1 <- cbind(data1, data2$Sessions)
names(data.OC1) <- c("OC", "Sess.control", "Sess.treat")
head(data.OC1)
```

```
##                                                                           OC
## 3                      Arts & Entertainment/Celebrities & Entertainment News
## 31                     Arts & Entertainment/Comics & Animation/Anime & Manga
## 94                            Arts & Entertainment/Comics & Animation/Comics
## 51 Arts & Entertainment/Events & Listings/Movie Listings & Theater Showtimes
## 52               Arts & Entertainment/Fun & Trivia/Fun Tests & Silly Surveys
## 63                                                Arts & Entertainment/Humor
##    Sess.control Sess.treat
## 3           135        176
## 31           37         44
## 94           12         23
## 51           25         21
## 52           25         27
## 63           20         16
```

```r
wilcox.test(data.OC1$Sess.control, data.OC1$Sess.treat, paired = T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data.OC1$Sess.control and data.OC1$Sess.treat
## V = 1209.5, p-value = 0.00184
## alternative hypothesis: true location shift is not equal to 0
```

```r
t.test(data.OC1$Sess.control, data.OC1$Sess.treat, paired = T)
```

```
##
##  Paired t-test
##
## data:  data.OC1$Sess.control and data.OC1$Sess.treat
## t = -3.4175, df = 94, p-value = 0.0009354
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.275478 -1.398206
## sample estimates:
## mean of the differences
##               -3.336842
```