T test

Aakanksha Baid

2/17/2020

## Step 1: Load the training

ctrtrain <- read.csv("tgTrain.csv")  
str(ctrtrain)

## 'data.frame': 300000 obs. of 18 variables:  
## $ No. : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ click : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ time : int 18 24 17 15 16 21 9 23 18 9 ...  
## $ app : int 2 1 1 1 1 1 8 1 1 2 ...  
## $ id : num 820025 552952 444907 273695 865888 ...  
## $ ad : int 2 3 6 2 3 3 1 2 3 3 ...  
## $ impup : int 1306 1237 2994 1505 684 1247 439 148 739 1556 ...  
## $ clup : int 1 2 4 4 9 5 2 8 20 0 ...  
## $ ctrup : num 0.000766 0.001617 0.001336 0.002658 0.013158 ...  
## $ impua : int 44 22 40 35 81 36 2 11 7 79 ...  
## $ clua : int 0 0 0 0 3 0 0 3 0 0 ...  
## $ ctrua : num 0 0 0 0 0.037 ...  
## $ imput : int 116 130 208 182 53 59 25 37 46 106 ...  
## $ clut : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ ctrut : num 0 0 0 0 0.0189 ...  
## $ imppat: int 27670 178805 7896 72170 70464 65929 71 189089 74855 5961 ...  
## $ clpat : int 21 2171 111 1232 1058 906 1 2606 1082 4 ...  
## $ ctrpat: num 0.000759 0.012142 0.014058 0.017071 0.015015 ...

## Step 2: Subset data

app1 <- ctrtrain[which(ctrtrain$app == 1),]  
app2 <- ctrtrain[which(ctrtrain$app == 2),]

## Step 3: Find the mean CTR for the two apps.

summary(app1)

## No. click time app   
## Min. : 2 Min. :0.00000 Min. : 1.00 Min. :1   
## 1st Qu.: 74606 1st Qu.:0.00000 1st Qu.:12.00 1st Qu.:1   
## Median :150176 Median :0.00000 Median :16.00 Median :1   
## Mean :149938 Mean :0.01216 Mean :15.49 Mean :1   
## 3rd Qu.:225228 3rd Qu.:0.00000 3rd Qu.:21.00 3rd Qu.:1   
## Max. :299999 Max. :1.00000 Max. :24.00 Max. :1   
## id ad impup clup   
## Min. : 28 Min. :1.000 Min. : 0 Min. : 0.000   
## 1st Qu.:250267 1st Qu.:1.000 1st Qu.: 273 1st Qu.: 1.000   
## Median :502670 Median :2.000 Median : 714 Median : 3.000   
## Mean :501349 Mean :2.316 Mean : 1095 Mean : 5.535   
## 3rd Qu.:752180 3rd Qu.:3.000 3rd Qu.: 1505 3rd Qu.: 7.000   
## Max. :999988 Max. :7.000 Max. :11218 Max. :144.000   
## ctrup impua clua ctrua   
## Min. :0.000000 Min. : 0.00 Min. : 0.0000 Min. :0.00000   
## 1st Qu.:0.001739 1st Qu.: 9.00 1st Qu.: 0.0000 1st Qu.:0.00000   
## Median :0.004686 Median : 25.00 Median : 0.0000 Median :0.00000   
## Mean :0.013043 Mean : 39.62 Mean : 0.3063 Mean :0.01499   
## 3rd Qu.:0.010345 3rd Qu.: 54.00 3rd Qu.: 0.0000 3rd Qu.:0.00000   
## Max. :1.000000 Max. :748.00 Max. :16.0000 Max. :1.00000   
## imput clut ctrut imppat   
## Min. : 0.00 Min. : 0.0000 Min. :0.00000 Min. : 0   
## 1st Qu.: 18.00 1st Qu.: 0.0000 1st Qu.:0.00000 1st Qu.: 28704   
## Median : 49.00 Median : 0.0000 Median :0.00000 Median : 65929   
## Mean : 77.92 Mean : 0.4337 Mean :0.01002 Mean : 70299   
## 3rd Qu.: 106.00 3rd Qu.: 0.0000 3rd Qu.:0.00000 3rd Qu.: 84039   
## Max. :1763.00 Max. :46.0000 Max. :1.00000 Max. :198959   
## clpat ctrpat   
## Min. : 0.0 Min. :0.00000   
## 1st Qu.: 359.0 1st Qu.:0.01214   
## Median : 994.0 Median :0.01383   
## Mean : 984.8 Mean :0.01350   
## 3rd Qu.:1283.0 3rd Qu.:0.01564   
## Max. :2656.0 Max. :0.04050

summary(app2)

## No. click time app   
## Min. : 1 Min. :0.0000000 Min. : 1.00 Min. :2   
## 1st Qu.: 75084 1st Qu.:0.0000000 1st Qu.:13.00 1st Qu.:2   
## Median :150163 Median :0.0000000 Median :16.00 Median :2   
## Mean :150064 Mean :0.0006029 Mean :15.98 Mean :2   
## 3rd Qu.:225034 3rd Qu.:0.0000000 3rd Qu.:21.00 3rd Qu.:2   
## Max. :300000 Max. :1.0000000 Max. :24.00 Max. :2   
## id ad impup clup   
## Min. : 13 Min. :1.000 Min. : 0 Min. : 0.0000   
## 1st Qu.:243670 1st Qu.:1.000 1st Qu.: 454 1st Qu.: 0.0000   
## Median :507184 Median :2.000 Median : 1112 Median : 0.0000   
## Mean :500465 Mean :2.367 Mean : 1719 Mean : 0.7927   
## 3rd Qu.:746835 3rd Qu.:3.000 3rd Qu.: 2129 3rd Qu.: 0.0000   
## Max. :999931 Max. :7.000 Max. :43253 Max. :211.0000   
## ctrup impua clua ctrua   
## Min. :0.0000000 Min. : 0.00 Min. : 0.0000 Min. :0.000000   
## 1st Qu.:0.0000000 1st Qu.: 13.00 1st Qu.: 0.0000 1st Qu.:0.000000   
## Median :0.0000000 Median : 33.00 Median : 0.0000 Median :0.000000   
## Mean :0.0005668 Mean : 53.45 Mean : 0.0575 Mean :0.001703   
## 3rd Qu.:0.0000000 3rd Qu.: 66.00 3rd Qu.: 0.0000 3rd Qu.:0.000000   
## Max. :0.2000000 Max. :1430.00 Max. :11.0000 Max. :1.000000   
## imput clut ctrut imppat   
## Min. : 0.0 Min. : 0.0000 Min. :0.000000 Min. : 0   
## 1st Qu.: 31.0 1st Qu.: 0.0000 1st Qu.:0.000000 1st Qu.:10727   
## Median : 75.0 Median : 0.0000 Median :0.000000 Median :20850   
## Mean : 118.2 Mean : 0.1106 Mean :0.001431 Mean :21962   
## 3rd Qu.: 150.0 3rd Qu.: 0.0000 3rd Qu.:0.000000 3rd Qu.:25883   
## Max. :3163.0 Max. :34.0000 Max. :0.666667 Max. :55495   
## clpat ctrpat   
## Min. : 0.00 Min. :0.0000000   
## 1st Qu.:11.00 1st Qu.:0.0007312   
## Median :20.00 Median :0.0009006   
## Mean :19.35 Mean :0.0008972   
## 3rd Qu.:25.00 3rd Qu.:0.0010552   
## Max. :60.00 Max. :0.0082645

From above, the mean CTR for app 1 or mean clicks observed is 0.01216 or 1.22%. The mean CTR for app 2 or mean clicks observed is 0.0006029 or 0.06%. The mean CTR of the two apps are different with app 1 CTR higher than that of app 2.

## Step 4: Run a t.test and draw inferences

t.test(app1$click,app2$click)

##   
## Welch Two Sample t-test  
##   
## data: app1$click and app2$click  
## t = 37.994, df = 178268, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.01096398 0.01215669  
## sample estimates:  
## mean of x mean of y   
## 0.0121632530 0.0006029181

Based on the results, as p-value < 0.05, there is a statistically significant difference between the CTR of app 1 and app 2. So we reject the null hypothesis that the two mean clicks from app 1 and app2 are equal.