# GameFun - AB testing analysis

*vincy hu*

*5/19/2020*

---

## Import the Data

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.1

## -- Attaching packages --------------------------------------------------------------------- tidyverse

## v ggplot2 3.2.1     v purrr   0.3.4
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.6.2

## Warning: package 'tidyr' was built under R version 3.6.1

## Warning: package 'readr' was built under R version 3.6.1

## Warning: package 'purrr' was built under R version 3.6.3

## Warning: package 'dplyr' was built under R version 3.6.1

## Warning: package 'stringr' was built under R version 3.6.1

## Warning: package 'forcats' was built under R version 3.6.1

## -- Conflicts ------------------------------------------------------------------------ tidyverse_confl
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
getwd()
```

```
## [1] "C:/Users/86136/Documents/GitHub/R-code-repository"
```

```
data <- read_excel('GameFun.xlsx')
```

## Question one

Before evaluating the effect of an experiment, it is important to make sure that the experiment was executed correctly. Check whether the test and control groups are probabilistically equivalent on their observables?

### 1a

More specific, compare the averages of the income, gender and gamer variables in the test and control groups. You should also report the % difference in the averages. Compute its statistical significance

```r
control <- filter(data,data$test == '0')
exp <- filter(data,data$test == '1')
#1a
var.test(control$income, exp$income, ratio = 1, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  control$income and exp$income
## F = 0.98973, num df = 11956, denom df = 28090, p-value = 0.5055
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.9603114 1.0202414
## sample estimates:
## ratio of variances
##          0.9897303
```

```r
#Variance test fails, so the variance in these two group are not equal
t.test(control$income, exp$income, alternative = "two.sided", mu = 0, paired = FALSE, var.equal = FALSE
```

```
##
##  Welch Two Sample t-test
##
## data:  control$income and exp$income
## t = 1.5238, df = 22667, p-value = 0.1276
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.06521165  0.52076184
## sample estimates:
## mean of x mean of y
##  55.16601  54.93824
```

```r
#The result shows that there is not significance to reject the null hypothesis. So the average of the i


var.test(control$gender, exp$gender, ratio = 1, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  control$gender and exp$gender
## F = 0.99925, num df = 11956, denom df = 28090, p-value = 0.963
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.9695498 1.0300563
## sample estimates:
## ratio of variances
##          0.9992518
```

```r
#Variance test fails, so the variance in these two group are not equal
t.test(control$gender, exp$gender, alternative = "two.sided", mu = 0, paired = FALSE, var.equal = FALSE
```

```
##
##  Welch Two Sample t-test
##
## data:  control$gender and exp$gender
```

```
## t = 0.11806, df = 22568, p-value = 0.906
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.00960814  0.01083979
## sample estimates:
## mean of x mean of y
## 0.6479050 0.6472892
```

```
#The result shows that there is not significance to reject the null hypothesis. So the average of the g

var.test(control$gamer, exp$gamer, ratio = 1, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  control$gamer and exp$gamer
## F = 0.99963, num df = 11956, denom df = 28090, p-value = 0.9826
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.969918 1.030448
## sample estimates:
## ratio of variances
##          0.9996312
```

```
#Variance test fails, so the variance in these two group are not equal
t.test(control$gamer, exp$gamer, alternative = "two.sided", mu = 0, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  control$gamer and exp$gamer
## t = 0.092, df = 22564, p-value = 0.9267
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.009986265  0.010969890
## sample estimates:
## mean of x mean of y
## 0.6018232 0.6013314
```

```
#The result shows that there is not significance to reject the null hypothesis. So the average of the g
```

## 1b

In this case, I would like to step back and evluate the business intuition for this experiment. This experiment wants to see the banners effect so it designs control and experiement test. To actually see the effectness of A/B testing, it requires to have the similar charterices people in the both control group and treatment group. So what is the baisc share charateircs of people who play games? I think to represent the whole picture of gamers, it is reasonable to inlcude income level, gender and the gamer So the above metric shows the average number in each category for both control and treatment group. It could mathmaticlly show that the control and treatment group match by comparing to the percentage of difference.

## 1c

I would hold the experiment for now. If we find out there is a large differnece, we need to investage why there is a such a big difference between two groups. We need to check our randomized theory and see if the total number of people is large enough. According to the central limit theory, when the number is large enough, it tends to behavine like a normal dist. Therefore, we need to see if the normality fails in our control and treatment group. If it fails, the reason could be we do not include enough people in the experiment

### 1d

According to Cameron, C. A., and P. K. Trivedi in the book **'Microeconometrics: Methods and applications'**, it is stated that the Bayesian information criterion(BIC) increases the penalty as sample size increases, whereas traditional hypothesis tests at a significance level such as 5% do not. Thus, when faced to 'big data' with large amount of observation, we can use BIC to compute the critical value that will increase with the sample size. For example, for nested models with q2 = q1 + 1 choosing the larger model on the basis of lower BIC is equivalent to using a two-sided t-test critical value of √ ln N, which equals 2.15, 3.03, and 3.72, respectively, for N = $10^2$ ,$10^4$ , and $10^6$

```
knitr::is_html_output(excludes = "markdown")
```

```
## [1] FALSE
```

## Question Two

```
#2a
#Control Group
all_control <- nrow(filter(control,control$purchase =='1'))/nrow(control)
#Exp Group
all_exp <- nrow(filter(exp,exp$purchase == '1'))/nrow(exp)
#ABS %
abs(all_control-all_exp)
```

```
## [1] 0.04060866
```

```
#2b
#Male
#Control Group
Male_control <- nrow(filter(control,control$gender =='1' & control$purchase == '1'))/nrow(filter(control
#Exp Group
Male_exp <- nrow(filter(exp,exp$gender == '1'& exp$purchase == '1'))/nrow(filter(exp,exp$gender == '1')
#ABS %
abs(Male_control-Male_exp)
```

```
## [1] 0.03739947
```

```
#FeMale
#Control Group
FeMale_control <- nrow(filter(control,control$gender =='0' & control$purchase == '1'))/nrow(filter(cont
#Exp Group
FeMale_exp <- nrow(filter(exp,exp$gender == '0'& exp$purchase == '1'))/nrow(filter(exp,exp$gender == '0
#ABS %
abs(FeMale_control-FeMale_exp)
```

```
## [1] 0.04650289
```

```
#2c
#Gamers
#Control Group
Gamer_control <- nrow(filter(control,control$purchase =='1' & control$gamer == '1'))/nrow(filter(contro
#Exp Group
Gamer_exp <- nrow(filter(exp,exp$purchase == '1'& exp$gamer == '1'))/nrow(filter(exp,exp$gamer == '1'))
```

```r
#ABS %
abs(Gamer_control-Gamer_exp)
```

```
## [1] 0.06905098
```

```r
#Non-Gamers
#Control Group
NoGamer_control <- nrow(filter(control,control$purchase =='1' & control$gamer == '0'))/nrow(filter(cont
#Exp Group
NoGamer_exp <- nrow(filter(exp,exp$purchase == '1'& exp$gamer == '0'))/nrow(filter(exp,exp$gamer == '0'
#ABS %
abs(NoGamer_control-NoGamer_exp)
```

```
## [1] 0.002294685
```

```r
#2d
#Male Gamer
#Control Group
Male_Gamer_control <- nrow(filter(control,control$gender =='1' & control$gamer == '1' & control$purchase
#Exp Group
Male_Gamer_exp <- nrow(filter(exp,exp$gender == '1'& exp$gamer == '1' & exp$purchase == '1'))/nrow(filte
#ABS %
abs(Male_Gamer_control-Male_Gamer_exp)
```

```
## [1] 0.06412899
```

```r
#Female Gamer
#Control Group
FeMale_Gamer_control <- nrow(filter(control,control$gender =='0' & control$gamer == '1' & control$purcha
#Exp Group
FeMale_Gamer_exp <- nrow(filter(exp,exp$gender == '1'& exp$gamer == '1' & exp$purchase == '1'))/nrow(fil
#ABS %
abs(FeMale_Gamer_control-FeMale_Gamer_exp)
```

```
## [1] 0.06936292
```

```r
#Male NoGamer
#Control Group
Male_NoGamer_control <- nrow(filter(control,control$gender =='1' & control$gamer == '0' & control$purcha
#Exp Group
Male_NoGamer_exp <- nrow(filter(exp,exp$gender == '1'& exp$gamer == '0' & exp$purchase == '1'))/nrow(fil
#ABS %
abs(Male_NoGamer_control-Male_NoGamer_exp)
```

```
## [1] 0.004843215
```

```r
#Female NoGamer
#Control Group
FeMale_NoGamer_control <- nrow(filter(control,control$gender =='0' & control$gamer == '0' & control$purc
#Exp Group
FeMale_NoGamer_exp <- nrow(filter(exp,exp$gender == '0'& exp$gamer == '0' & exp$purchase == '1'))/nrow(f
#ABS %
abs(FeMale_NoGamer_control-FeMale_NoGamer_exp)
```

```
## [1] 0.001760648
```

## Question Three

```
#3a
Control_Rev <- nrow(filter(control,control$purchase == '1'))*37.5
Exp_Rev <- nrow(filter(exp,exp$purchase == '1'))*12.5

#3b
Male_Control_Rev <- nrow(filter(control,control$purchase == '1' & control$gender == '1'))*37.5
Male_Exp_Rev <- nrow(filter(exp,exp$purchase == '1' & exp$gender == '1'))*12.5
```

## Question Four

Based on the evluation on Question two, in each segment, the absolute diff between the control and test group is less than 10%. There is not a significance number showing that this ads runs really well. Therefore, I wouldn't suggest to run the promotion again. Especially by considering the company could only earn 12.5 dollars on average for promoted customer.