# Stat 7350 Assignment 2

*Jingyu Wang 7701969*

*March 29, 2019*

## 1. Frame the problem

The idea is from the recent news which is about both Samsung and Huawei have released new cellphones and both of them are Android system. At the same time, I found that many friends around me are using iphones which have iOS system. Android and IOS are two major systems in the markets of mobile phones today. Assume that, I am a manager of one Canadian mobile phone company, and I want to know the impact of different genders, different age groups on the choice of mobile phone system.

Based on the previous sales record of company, we know what is the sales volume for each system of the mobile phones. Firstly, according to Bree R. study [1], the number of IOS user is greater than that of Android in North America. There are about 54.9% people are using the IOS system, and the rest of people choose Android. On the other hand, experience of different system leads people to have different perferences for the system. As the expression of Pern U. etc [2], there is significant difference of feeling for using some applications between IOS and Android system users. Thus, we need to design an experiment to observe the impact of different age groups, genders on the choice of cellphone system.

My motivation is that I can correctly recommend the right cellphone system to new customers to ensure that she or he will be willing to purchase our product and accept the service of our company.

## 2. State the testable question(s) and your hypothesse.

Question 1: Is there significant difference for number of both user based on difference range of age?

$H_0$: Number of both user based on difference range of age are equal. (no signidicant difference between both user based on difference range of age)

$H_1$: At least one age range for users amount are different with others.

Question 2: Is there significant difference for number of both user based on gender (male/female)?

$H_0$: Number of both user based on different genders are equal. (no signidicant difference between both user based on gender)

$H_1$: There is significant difference between the number of both users for different gender.

## 3. Project proposal

The questionnarie will be the method or collecting data in our case and question paper should be given to every person who come into our mobile company. TO ensure that people are willing to be an volunteer do the survey, we can use parts of money($700 of that $5000) to do prize draw. For example, 6 persons can

get a special gift from our company. The number of person who are chosen to get the prize can be adjusted depend on the number of complete survey.

Notice that, it is not a randomization procesure. This is the convenience sampling which chooses individuals who are easiest to reach. The disadvantage of this sampling method is that the sample usually can not represent entire population! We also consider that the survey can be done in the university, but another problem is coming out with the limitation of age range in the university. Thus, doing survey in the store will be a better choice for us.

The questionnaire should be done at same time for each day to reduce the variation of samples. For example, paper are given to people who come into our store from 10:00 am to 4:00 pm everyday.

In the questionnaire, it should include the following questions:

- Gender: Male/ Female
- Age range: 18-25/25-35/35-45/45-55/55+
- Prefered system: Android/IOS
- Salary: 0-30,000/30,000-50,000/50,000-70,000/70,000+ (Note that, salary is the number of annual pay for each person.)
- Email:

Firstly, we can consider the Randomized Complete Block Design (RCBD) for the experiment. Gender and age range are two treatments for this design, and for each treatment, there are two levels for gender and 5 levels for age range. Thus, a total of 10 treatment will be used in the experiment. Note that, there is no control treatment in this experiemnt, which is not suitable for this case.

There are two blocks in our case because we seperate and summarize our design in each two weeks. Thus, two block will be also considered as two replications in our design. Also, we expect there is not significant difference between each block. Assume that the month we do survey is a regular month(like March, May), then it is good to consider the procedure of experiment replicate twice. Otherwise, this method will give large bias of the results. For example, the number of cellphone sales and customers in the last two weeks in November will be much larger the beginning two weeks because of the black friday. Then, the blocks cannot be considers as replications under this situiation.

To find the experimental units, we calculate the percentage of IOS user for each treatment in every two weeks. For example, you should calculate the percentage of IOS user with male and 18-25 age as the number at the first row and first column in your data table. After doing all of above steps, you should have a data table like a $2 \times 10$ matrix with two blocks and 10 treatments.

In addition, there is one more factor we considered as covariate in our case, and also added it in the questionnaire such as annual salary. Salary is related to age of person, gender and the price choice of mobile phone, which will affect the percentage of users for both systems. It is not a confunding factors because of the definition. In the definition of confunding factor, it should be another variable which can affects all variables. Annual salary may changes the choice of cellphone and its system, however, it cannot cause the change of gender and age range. People who have higher salary may choose higher price and cellphone, at least they

have a greater probability of choosing more expensive mobile phones. To control this covariate, you need to find the relationship between gender, age and salary, and express annual salary by using other factors.

You can follow the following step to conduct this experiemnt.

1. Ensure the sample groups. Remider that, there should not be any missing values in this experiment to use RCBD, please check and make sure each treatment and block have their corresponding value.
2. Using the design method for this experiment. Please check you have everything prepared as above, such as blocks and treatments.
3. Analysis data by using Statistical method. Avova can be used for RCBD and check whether treatment and block effects are significant difference from one other. Also, all assumptions of AVOVA need to be checked such that the nomality of residuals with mean zero (Shapiro Test), the homogeneous of residual variance (Levene Test) and sample independency. The tests in the bracket are recommended method to use.

## 4. Data clooection and Statistical analysis plan

You can use excel to store the data and make filter at the top of column (is needed)to choose easier when doing analysis. We also consider to use Two-way ANOVA in our case, and the dataset have 8 rows and 4 columns. For each column, there are system(Andriod/IOS), rate(the percentage of user for each system), group(female/male) and week(first two week as 1, last two week as 2). Your dataset should look like the falmat as following. Also, all of system, group and week should be treates as factor, and groups and weeks are fixed effects.

```
library(readr)
data<-read.csv("C:/Users/summe/Downloads/Stat-7350-class/Assignment 2/A2.csv")
data
```

```
##   ï..system rate group week
## 1   Andriod   NA     F    1
## 2   Andriod   NA     F    2
## 3       IOS   NA     F    1
## 4       IOS   NA     F    2
## 5   Andriod   NA     M    1
## 6   Andriod   NA     M    2
## 7       IOS   NA     M    1
## 8       IOS   NA     M    2
```

Notice that, assumptions also need to be checked, which include the nomality of residuals with mean zero, the homogeneous of residual variance and sphericity for variance(Mauchly Test). The last assumption is the spectial one only for two-way ANOVA.

From this Two-way ANOVA, the results will give whether group and week are significant factor in our experiemnt. The expected reults are that group is significant and week is not significant.

Same Two-way ANOVA can be done to check whether age is the significant factor, and the only thing you

need to do is use age range as group. Then, there are 5 levels for age, and 20 rate number for this situiation. The dataset should have 20 rows and 4 columns with system, rate, group(age range) and week, and displays as below. The age range have 5 levels and set it as 1-5 to corresponding 18-25/25-35/35-45/45-55/55+ five age range.

```
read.csv("C:/Users/summe/Downloads/Stat-7350-class/Assignment 2/A2 age.csv")
```

```
##    ï..system rate group week
## 1    Andriod   NA     1    1
## 2    Andriod   NA     1    2
## 3        IOS   NA     1    1
## 4        IOS   NA     1    2
## 5    Andriod   NA     2    1
## 6    Andriod   NA     2    2
## 7        IOS   NA     2    1
## 8        IOS   NA     2    2
## 9    Andriod   NA     3    1
## 10   Andriod   NA     3    2
## 11       IOS   NA     3    1
## 12       IOS   NA     3    2
## 13   Andriod   NA     4    1
## 14   Andriod   NA     4    2
## 15       IOS   NA     4    1
## 16       IOS   NA     4    2
## 17   Andriod   NA     5    1
## 18   Andriod   NA     5    2
## 19       IOS   NA     5    1
## 20       IOS   NA     5    2
```

# References

[1] Bree R. 2018. Canada's prolific smartphone market skews to iOS: study. *Media in Canada*. http://mediaincanada.com/2018/01/23/canadas-prolific-smartphone-market-skews-to-ios-study/

[2] Perin Ü., Tugba T. and P. Erhan Eren. 2015. A Study on User Perception of Mobile Commerce for Android and iOS Device Users. *Springer*.