

Please initial each line before you start

XZ\_\_ The effort on this exam is entirely of my own.

XZ\_\_ I will not discuss any part of this exam with anyone from the class or out of the class (roommates, classmates, friends, acquaintances, etc..)

XZ\_\_ I will not seek any sort of assistance or advice from anyone other than the primary instructor.

Please sign below

I understand that if I violate the honor code or if there is any suspicion of violation, then I and others involved will be reported to the UC Davis Office of Judicial Affairs, which will then open a case for investigation. Until a decision is made, my grade will be on hold.

Xingxuan Zhang 11/19/2019\_\_\_\_\_

Sign and Date

Instructions:

- Please state your hypotheses formally (using symbols), except for nonparametric tests. Points will NOT be awarded for hypothesis not written formally (using symbols). For example, it is not sufficient to say that correlation exists or correlation does not exist. You need to write this formally using the appropriate population parameter symbol and the correct sign.
- Please submit your R code and a Word document stating your hypotheses and your interpretations.
- If you make assumptions, please state them clearly.

```
In [59]: install.packages('pwr2')
```

executed in 4.32s, finished 22:07:12 2019-11-21

```
In [12]: library(stats)
library(PMCMR)
library(reshape2)
library(reshape)
library(tidyverse)
library(PASWR2)
library(pwr2)
```

executed in 1.22s, finished 10:49:54 2019-11-27

## 1 (20 points) Question 1

As more and more companies use the Internet for advertising, click fraud has become a major concern. When a search engine, say, in this case, Google, places an ad for a company with its search results, the company pays a fee to Google each time someone clicks on the link. That's fine when it's a person who's interested in buying a product or service, but not so good when it's a computer program pretending to be a customer. An analysis of 1,200 clicks coming into a company's site during a week identified 175 of these clicks as fraudulent.

### 1.1 a. (5 points) Under what conditions does it make sense to treat these 1,200 clicks as a sample? What would be the population?

These 1,200 clicks must be randomly selected from the population.

The population would be the all clicks during the week, or if the week data can infer the entire time frame, the population would be the all clicks Goolde has.

## 1.2 b. (10 points) Show the 95% confidence interval for the population proportion of fraudulent clicks in a form suitable for sharing with a nontechnical audience.

```
In [4]: p_hat <-175/1200
        z<-qnorm(0.975,0,1)

        lower_limit<- p_hat-z*sqrt(p_hat*(1-p_hat)/1200)
        upper_limit<- p_hat+z*sqrt(p_hat*(1-p_hat)/1200)

        cat('The 95% confidence interval for the population proportion of fraudulent clicks is (',lower_l
```

executed in 31ms, finished 00:54:47 2019-11-20

The 95% confidence interval for the population proportion of fraudulent clicks is ( 0.1258643 , 0.1658024 ).

## 1.3 c. (5 points) If a company pays Google 4.50 for each click, give a confidence interval for the mean cost due to fraud per click.

```
In [6]: cat('The 95% confidence interval for the mean cost due to fraud per click is (',lower_limit*4.5,'
```

executed in 23ms, finished 00:56:34 2019-11-20

The 95% confidence interval for the mean cost due to fraud per click is ( 0.5663893 , 0.7461107 ).

## 2 (20 points) Question 2

Ten judges were asked to test the quality of four different brands of orange juice. The judges assigned scores using a 5-point scale where 1 = Bad, 2 = Poor, 3 = Average, 4 = Good, and 5 = Excellent. The results are shown on the CSV file Question 2. Run an appropriate statistical method to conclude at the 5% significance level if there are differences in sensory quality between the four brands of orange juice.

```
In [3]: setwd("D:/BAX441/Midterm Exam")
        Q2<-read.csv('Question2.csv')
        Q2
```

executed in 41ms, finished 17:25:50 2019-11-21

Judge	OJ.1	OJ.2	OJ.3	OJ.4
1	3	5	4	3
2	2	3	5	4
3	4	4	3	4
4	3	4	5	2
5	2	4	4	3
6	4	5	5	3
7	3	3	4	4
8	2	3	3	3
9	4	3	5	4
10	2	4	5	3

#1. Randomly selected sample - Satisfy

#2. Normality - Unsatisfy, because the data is ordinal data

### Nonparametric Test: Friedman Test (Random Block Design)

Test the Treatment Effect:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H_1$ : At least two means differ

where  $\mu$  is the mean score of the juice

Significant Level: 5%

```
In [7]: #####Prepare Data#####
df <- melt(Q2, id="Judge")
colnames(df)<-c('Judge','Brands','Rate')
block<-as.factor(df$Judge)
type<-as.factor(df$Brands)
rating<-df$Rate

#####Friedman Test (Nonparametric Test for Random Block Design)#####3
friedman.test(rating,type,block)
posthoc.friedman.nemenyi.test(rating,type,block)

#####Conclusion#####
print('At 5% significant level, there is enough evidence to say there are differences among these
```

executed in 41ms, finished 17:32:56 2019-11-21

Friedman rank sum test

data: rating, type and block  
Friedman chi-squared = 10.953, df = 3, p-value = 0.01198

Pairwise comparisons using Nemenyi multiple comparison test  
with q approximation for unreplicated blocked data

data: rating , type and block

	OJ.1	OJ.2	OJ.3
OJ.2	0.226	-	-
OJ.3	0.017	0.726	-
OJ.4	0.726	0.822	0.226

P value adjustment method: none

[1] "At 5% significant level, there is enough evidence to say there are differences among these four brands. With the host hoc test fro the Friedman Test, we can find that the Brand 1 is diffe rent from the brand 3."

## 3 (20 points) Question 3

Consider the data from the General Social Survey (2014). The dataset is available on the CSV file Question3. Can we infer from the data that jobs that are most secure (JOBLOSE: 1 =Very likely, 2 = Fairly likely, 3 = Not too likely, 4 = Not likely) are also the most satisfying (SATJOB: 1 = Very satisfied, 2 = Moderately satisfied, 3 = A little dissatisfied, 4 = Very dissatisfied)? SEX: 1 = MALE, 2 = FEMALE).

```
In [8]: setwd("D:/BAX441/Midterm Exam")
Q3<-read.csv('Question3.csv')
Q3
```

executed in 66ms, finished 17:33:35 2019-11-21

### Requirement Test

#1. Randomly selected sample - Satisfy

#2. Normality - Unsatisfy, because the data is ordinal data

### Nonparametric Test: Spearman Test (Correlationship)

$$H_0: \rho \geq 0$$

$$H_1: \rho < 0$$

where  $\rho$  is the correlationship between SATJOB and JOBLOSE

Significant Level: 5%

```
In [15]: #####Prepare Data#####
df<-Q3 %>% drop_na(c("JOBLOSE", "SATJOB"))

#####Spearman Test (Nonparametric Test for Correlation)#####
cor.test(x=df$JOBLOSE, y=df$SATJOB,method='spearman')

#####Conclusion#####
print('Yes, at 5% significant level, the joblose and jobsatisfy do have correlation. Since the rho
```

executed in 35ms, finished 19:20:51 2019-11-21

Warning message in cor.test.default(x = df\$JOBLOSE, y = df\$SATJOB, method = "spearman"):  
"Cannot compute exact p-value with ties"

Spearman's rank correlation rho

```
data: df$JOBLOSE and df$SATJOB
S = 215806555, p-value = 2.879e-11
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.2059142
```

[1] "Yes, at 5% significant level, the joblose and jobsatisfy do have correlation. Since the rho is -0.2059, we can say that the jobs that are more secure are also more satisfying."

## 4 (30 points) Question 4

Franchise Tax Board (FTB) in Sacramento regularly hires clerical employees to work on the tax-returns received. The HR group regularly interviews prospective clerical employees and tests their skill at data entry. The following table shows the number of errors made by 60 prospective clerks when entering a tax form with 80 numbers. Conduct a statistical test to determine if the number of errors follow a Poisson distribution.

# Errors	0	1	2	3	4 or more
# Clerks	12	20	9	14	5

### Requirement Test

#1.Simple Random Sample:satisfy

#2.Contingency table condition: The categories are mutually exclusive and the question made sure that these are 60 different users.

#3.Sample size condition: all frequency numbers are  $\geq 5$

```
In [31]: #####Prepare Data#####
errors<-c(0,1,2,3,4)
clerks<-c(12,20,9,14,5)
lambda<-sum(errors*clerks)/60

expect<-dpois(errors,lambda)
expect[length(expect)]<-(1 -ppois(4 -1, lambda))
expect
```

executed in 27ms, finished 19:57:59 2019-11-21

0.188875602837562 0.314792671395936 0.26232722616328 0.145737347868489 0.0882671517347324

### Categorical Data: Good-Fit Test

$H_0: p_0 = 0.189, p_1 = 0.315, p_2 = 0.262, p_3 = 0.146, p_{4+} = 0.088$

$H_1$ : At least one proportion differs from the expected Poisson distribution

```
In [32]: chisq.test(x=clerks,p=expect)
```

executed in 28ms, finished 19:58:04 2019-11-21

Chi-squared test for given probabilities

data: clerks  
X-squared = 6.1663, df = 4, p-value = 0.1871

```
In [33]: pchisq(6.1663,3,lower.tail=FALSE)
```

executed in 24ms, finished 19:58:38 2019-11-21

0.103793804333187

### Conclusion

At 5% significant level, the number of errors did not follow a poisson distribution.

## 5 (50 points) Question 5

Eastern Utility Power, a large electric power company in the Northeastern part of the United States, has just built a modern nuclear power plant. The plant discharges wastewater that is allowed to flow into the Atlantic Ocean. The Environmental Protection Agency (EPA) has ordered that the wastewater may not be excessively warm so that thermal pollution of the marine environment near the plant can be avoided. Because of this order, the wastewater is allowed to cool in specially constructed ponds and is then released into the ocean. The cooling system works properly if the mean temperature of the wastewater discharged is  $60^{\circ}$  F or cooler. Eastern Utility Power is required to monitor the temperature of the wastewater. A sample of 100 temperature readings will be obtained each day, and if the sample results cast a substantial amount of doubt on the hypothesis that the cooling system is working properly, then the plant must be shut down and appropriate actions must be taken to correct the problem.

**5.1 a. (5 points) Eastern Utility Power wishes to set up a hypothesis test so that the power plant will be shut down when the null hypothesis is rejected. Set up the null and alternative hypotheses that**

**should be used.**

$$H_0: \mu \leq 60^\circ \text{ F}$$

$$H_1: \mu > 60^\circ \text{ F}$$

where  $\mu$  is the average temperature of the wastewater discharged

Null hypothesis: the mean temperature of the wastewater discharged is  $60^\circ \text{ F}$  or cooler, which means the cooling system works.

Alternative hypothesis: the mean temperature of the wastewater discharged is higher than  $60^\circ$ , which means the cooling system cannot work well so the power plant will be shut down.

**5.2 b. (5 points) In the context of this situation, interpret making a Type I error; interpret making a Type II error.**

Type I error: The cooling system works and the power plant need not be shut down, but the test rejects this and indicates that the power plant should be shut down. Cost: the company lost its powerplant.

Type II error: The cooling system does not work well and the power plant should be shut down to protect the environment, but the test did not show this and allow the harmful power plant continue operating. Cost: the power plant harm the environment.

**5.3 c. (5 points) The EPA periodically conducts spot checks to determine whether the wastewater being discharged is too warm. Suppose the EPA has the power to impose very severe penalties (for example, very heavy fines) when the wastewater is excessively warm. Other things being equal, should Eastern Utility Power set the probability of a Type I error equal to 0.01 or 0.05?**

Since the power plant wants to avoid extreme high fines that charged by EPA, in the power plant daily test, the company should use 0.05 as the probability of a Type I error. By doing so, the power plant has larger probability to detect the abnormal wastewater temperature and do something to deal with the problem before the EPA's spot checks and extreme high fines.

**5.4 ? d. (5 points) Eastern Utility Power wishes to set up a hypothesis test and decides to use a level of significance  $\alpha = 0.05$ . Suppose a random sample of 100 temperature readings is obtained. If the sample mean of the 100 temperature readings is 60.482, run the hypothesis test and determine if the power plant should be shut down and the cooling system repaired. Assume that the population standard deviation equals 2.**

$$H_0: \mu \leq 60^\circ \text{ F}$$

$$H_1: \mu > 60^\circ \text{ F}$$

Significant Level: 5%

```
In [48]: mean_sample<-60.482
size_sample<-100
sd_pop<-2

#####Z-Test#####
z95<-qnorm(0.95,0,1)
upper_limit<-60+(z95*sd_pop/sqrt(size_sample))

cat('The sample mean is ',mean_sample, ', which is larger than 95% upper limit, ', upper_limit, '.

#####Conclusion#####
print('At 5% significant level, we have enough evidence to reject the null hypothesis. The power
```

executed in 34ms, finished 21:39:13 2019-11-21

The sample mean is 60.482 , which is larger than 95% upper limit, 60.32897 .  
 [1] "At 5% significant level, we have enough evidence to reject the null hypothesis. The power p  
 lant should be shut down and the cooling system repaired."

**5.5 e. (10 points) If the Eastern Utility Power sets  $\alpha = 0.025$ , calculate the probability of a Type II error for each of the following alternative values of  $\mu$ : 60.1, 60.2, 60.3, 60.4, 60.5, 60.6, 60.7, 60.8, 60.9, and 61.**

```
In [3]: alt_mean<-c(60.1,60.2,60.3,60.4,60.5,60.6,60.7,60.8,60.9,61)
alpha<-0.025
mean<-60
sd<-2

z95<-qnorm(1-alpha,mean,sd/sqrt(100),lower.tail=TRUE)
z95
type_II_error<-as.data.frame(pnorm(z95,alt_mean,sd/sqrt(100),lower.tail=TRUE))
df<-cbind(alt_mean,type_II_error)
colnames(df)<-c('alternatvie mean','type II error')
df
```

executed in 70ms, finished 19:21:16 2019-11-24

60.391992796908

alternatvie mean	type II error
60.1	0.927850014
60.2	0.831463329
60.3	0.677228964
60.4	0.484032207
60.5	0.294586098
60.6	0.149161584
60.7	0.061775787
60.8	0.020673369
60.9	0.005542053
61.0	0.001182749

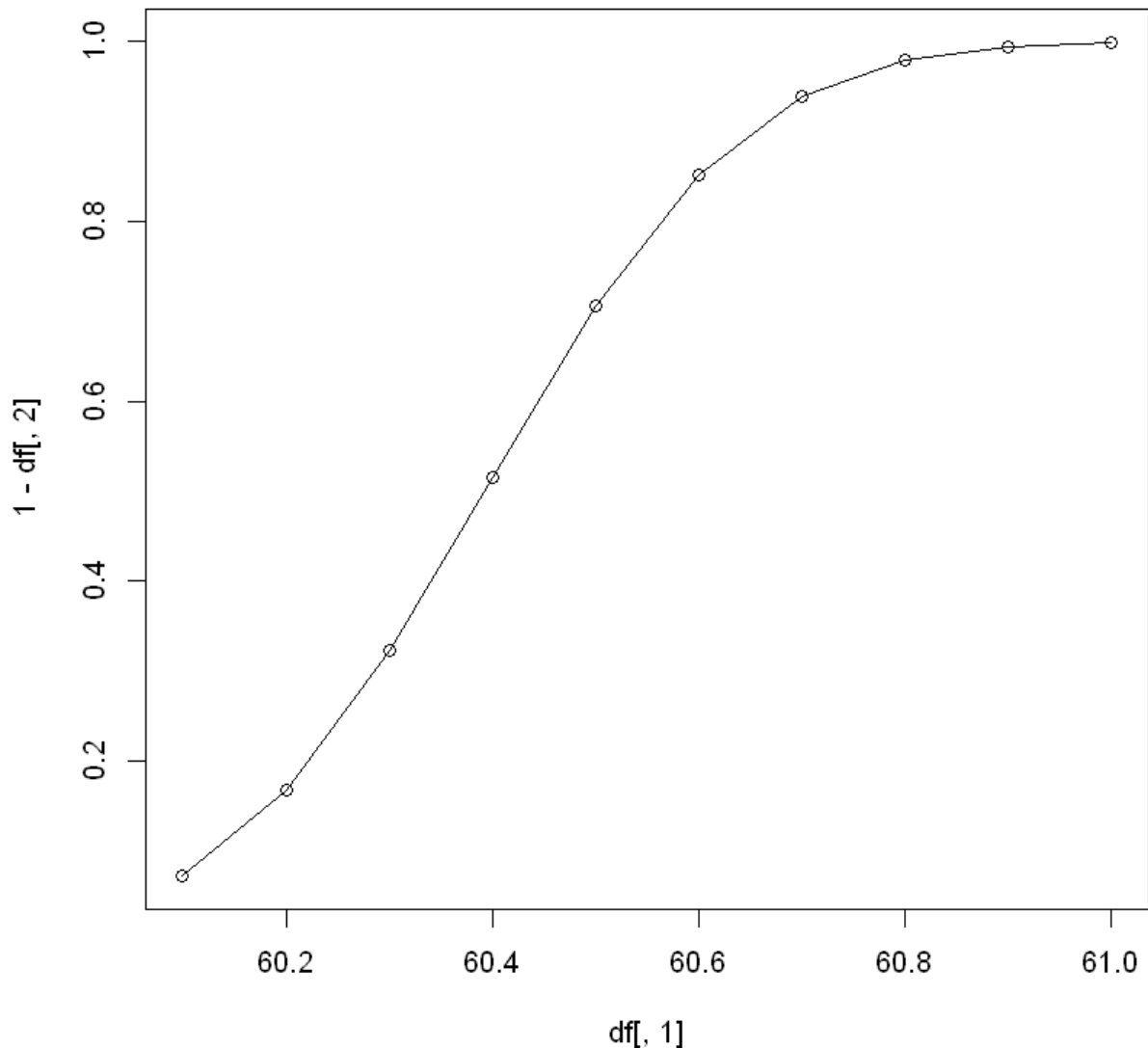
**5.6 ? f. (10 points) If we want the probability of making a Type II error when  $\mu = 60.5$  to be very small, is Eastern Utility Power's hypothesis test adequate? Explain why or why not. If not, and if we wish to maintain the value of  $\alpha = 0.025$ , what must be done?**

Eastern Utility Power's hypothesis test is not adequate because from the previous question, we know that if the  $\alpha = 0.025$ , the probability of making a type II error when  $\mu = 60.5$  is 29.46%, which is pretty high. If we wish to maintain the value of  $\alpha = 0.025$  and want the type II error to be very small, we must increase the sample size.

**5.7 g. (10 points) Create a power curve for the Eastern Utility Power. What happens to the power of the test as the alternative value of  $\mu$  moves away from 60?**

```
In [9]: plot(df[,1],1-df[,2],type='o')
```

executed in 88ms, finished 19:25:39 2019-11-24



The power of test increases when the alternative  $\mu$  moves away from the 60.

**6 (40 points) Question 6**



An HR department specialist of a leading hotel chain in the U.S. is exploring ways to improve retention among hotel staff. The problem is particularly acute among those who maintain rooms, work in hotel restaurant, and greet guests. Within this chain, among those who greet and register guests at the front desk, the annual percentage who quit is 36% . Among the employees who work the front desk, more than half are expected to quit during the next year. The HR specialist has estimated that the turnover rate costs \$20,000 per quitter, with the costs attributed to factors like:

- The time a supervisor spends to orient and train a new employee
- The effort to recruit and interview replacement workers
- The loss of efficiencies with a new employee rather than one who is more experienced and takes less time to complete tasks
- Administrative time both to add the new employee to the payroll and to remove the prior employee

To increase retention by lowering the quit rate, the specialist has formulated a benefits program targeted at employees who staff the front desk. The cost of offering these benefits averages \$2,000 per employee. The chain operates 225 hotels, each with 16 front-desk employees. As a test, the specialist has proposed extending improved benefits to 320 employees who work the front desk in 20 hotels.

**6.1 a. (5 points) Using the data given in the narrative above, answer why would it be important to test the effect of the employee benefits program before offering it to all front-desk employees at the hotel chain?**

Without the effect test of the employee benefits program, the company cannot know whether the program will decrease the churn rate. If the company offers this benefits to all employees at the hotel chain, the total cost would be \$7,200,000. If the benefits program does not work at all, the company still loses \$20,000 per quitter and also \$7,200,000 for the waste of the program. With the test that only offers 320 employees, the company's cost is only \$640,000.

**6.2 b. (5 points) If the benefits program is to be tested, how would you recommend choosing the hotels? How long will the test take to run? (There is no single best answer to this question; try your best to articulate your recommendations.)**

I would randomly choose the hotels simply. Or based on different revenue levels, I randomly choose some hotels to make sure that the hotel size would not generate bias to the test. Since the churn rate is calculated annually, I recommend that the test can take 1 year to run. However, to save the cost and effort, we can do a three-month test and compare to the data in the same period last year. However, if the churn rate's trend is not follow the same pattern every year, then we should consider the churn rate trend in our test.

**6.3 ? c. (5 points) An analyst proposed testing the null hypothesis  $H_0: p \leq 0.36$ , where  $p$  is the annual quit rate for employees who work the main desk if the new program is implemented. Is this the correct null hypothesis? If not, why not?**

No, this is not a correct null hypothesis because the test want to know whether this program can work or not. If we reject the null hypothesis, then the program works, so the  $H_0$  is  $p \geq 0.36$ .

**6.4 d. (5 points) Another analyst proposed the null hypothesis  $H_0: p \geq 0.36$ , where  $p$  is the annual quit rate for employees who work the main desk if the new program is implemented. Is this the correct**

## null hypothesis? What key issue does this hypothesis ignore?

This null hypothesis is half right. The key issue this hypothesis ignores is that it does not account the cost of the program, which is whether the savings generated by the program can cover the cost of it.

**6.5 e. (10 points) If the chosen null hypothesis is  $H_0: p \geq 0.30$ , what percentage of these 320 must not quit in order to reject  $H_0$  if  $\alpha = 0.05$ ?**

$$H_0: p \geq 0.3$$

$$H_1: p < 0.3$$

where  $p$  is the annual quit rate for employees

```
In [3]: z95<-qnorm(0.95,0,1)
cat('The percentage of the sample must not quit in order to reject null hypothesis if the signifi
```

executed in 55ms, finished 17:11:14 2019-11-23

The percentage of the sample must not quit in order to reject null hypothesis if the significant level is 0.05 is 74.21368 %.

**6.6 f. (10 points) Assume that the chosen null hypothesis is  $H_0: p \geq 0.30$ . Suppose that the actual quit rate among employees who receive these new benefits is 25%. What is the chance that the test of  $H_0$  will correctly reject  $H_0$ ?**

$$H_0: p \geq 0.3$$

$$H_1: p < 0.3$$

where  $p$  is the annual quit rate of the employees

The chance that the test of  $H_0$  will correctly reject  $H_0$  is  $1 - \beta$

Assume the significant level is 5%.

```
In [6]: z95<-qnorm(0.95,0,1)
critic_value<-0.3-z95*sqrt((0.3*0.7)/320)
z<-(critic_value-0.25)/sqrt((0.25*0.75)/320)
beta<-pnorm(z,mean=0, sd=1, lower.tail = FALSE)
cat('The chance that the test of H0 will correctly reject H0 is ', (1-beta)*100, '%.')
```

executed in 26ms, finished 10:27:12 2019-11-27

The chance that the test of  $H_0$  will correctly reject  $H_0$  is 62.73495 %.

## 7 (20 points) Question 7

The Survey of Consumer Finances (SCF) is conducted every 3 years to provide detailed information on the finances of U.S. households. The study is sponsored by the Federal Reserve Board in cooperation with the Department of the Treasury. Since 1992, data have been collected by the National Opinion Research Center (NORC) at the University of Chicago. The full dataset is very large, and the number of variables is extremely large as well. For this exercise, I have created a subset of the dataset. This subset contains data for the middle-class households (40% - 60% of the net worth). Consider the dataset on the CSV file Question7. You want to explore if there is sufficient evidence that middle-

class heads of households who work for someone else have higher incomes than heads of households who are self-employed. Use the appropriate statistical method and state why did you choose the selected method. Below is the definition of the OCCAT1 variable from your dataset: OCCAT1 Occupation categories for head of household:

1. Work for someone else,
2. Self-employed/partnership,
3. Retired/disabled + (student/homemaker/misc. not working and age 65 or older),
4. Other groups not working (mainly those under 65 and out of the workforce).

```
In [7]: setwd("D:/BAX441/Midterm Exam")
Q7<-read.csv('Question7.csv')
df<-Q7[Q7$OCCAT1==1|Q7$OCCAT1==2,]
OCCAT1<-as.factor(df$OCCAT1)
income<-df$INCOME
df
```

executed in 60ms, finished 10:30:57 2019-11-27

### Requirement Test

#1. Randomly selected sample - Satisfy

#2. Normality - Unsatisfy, because the p-value of the normality test is less than 0.05

#3. Variances are constant - Satisfy, because the p-value of the normality test is larger than 0.05

### Hypothesis

$$H_0: \mu_{work\ for\ someone} \leq \mu_{work\ for\ self}$$

$$H_1: \mu_{work\ for\ someone} > \mu_{work\ for\ self}$$

where  $\mu$  is the mean income

### Significant Level

Assume the significant level is 5%.

```

In [9]: #####Requirement Test#####
#1. Randomly selected sample - Satisfy
#2. Normality - Unsatisfy, because the p-value of the normality test is less than 0.05
model_ex7 <- lm( income~ OCCAT1)
resids_ex7 <- residuals(model_ex7)
preds_ex7 <- predict(model_ex7)

nortest::ad.test(resids_ex7)

#hist(resids_ex7)
#hist(df$INCOME)
#hist(df$OCCAT1)
#qqnorm(resids_ex7)
#qqline(resids_ex7)

#3. Variances are constant - Satisfy, because the p-value of the normality test is larger than 0.05
fligner.test( income~ OCCAT1)
car::leveneTest(income~ OCCAT1)
#####Wilcoxon Rank Sum Test #####

group1<-Q7[Q7$OCCAT1==1,]
group2<-Q7[Q7$OCCAT1==2,]
wilcox.test(income~ OCCAT1,alt='greater',paired=FALSE,exact=FALSE,conf.level=0.95)

plot(density(group1$INCOME),col='black')
lines(density(group2$INCOME),col = "red")
#####Conclusion#####

print('Yes, at 5% significant level, middle-class heads of households who work for someone else have higher incomes than heads of households who are self-employed')

```

executed in 189ms, finished 10:31:23 2019-11-27

#### Anderson-Darling normality test

data: resids\_ex7  
A = 17.136, p-value < 2.2e-16

#### Fligner-Killeen test of homogeneity of variances

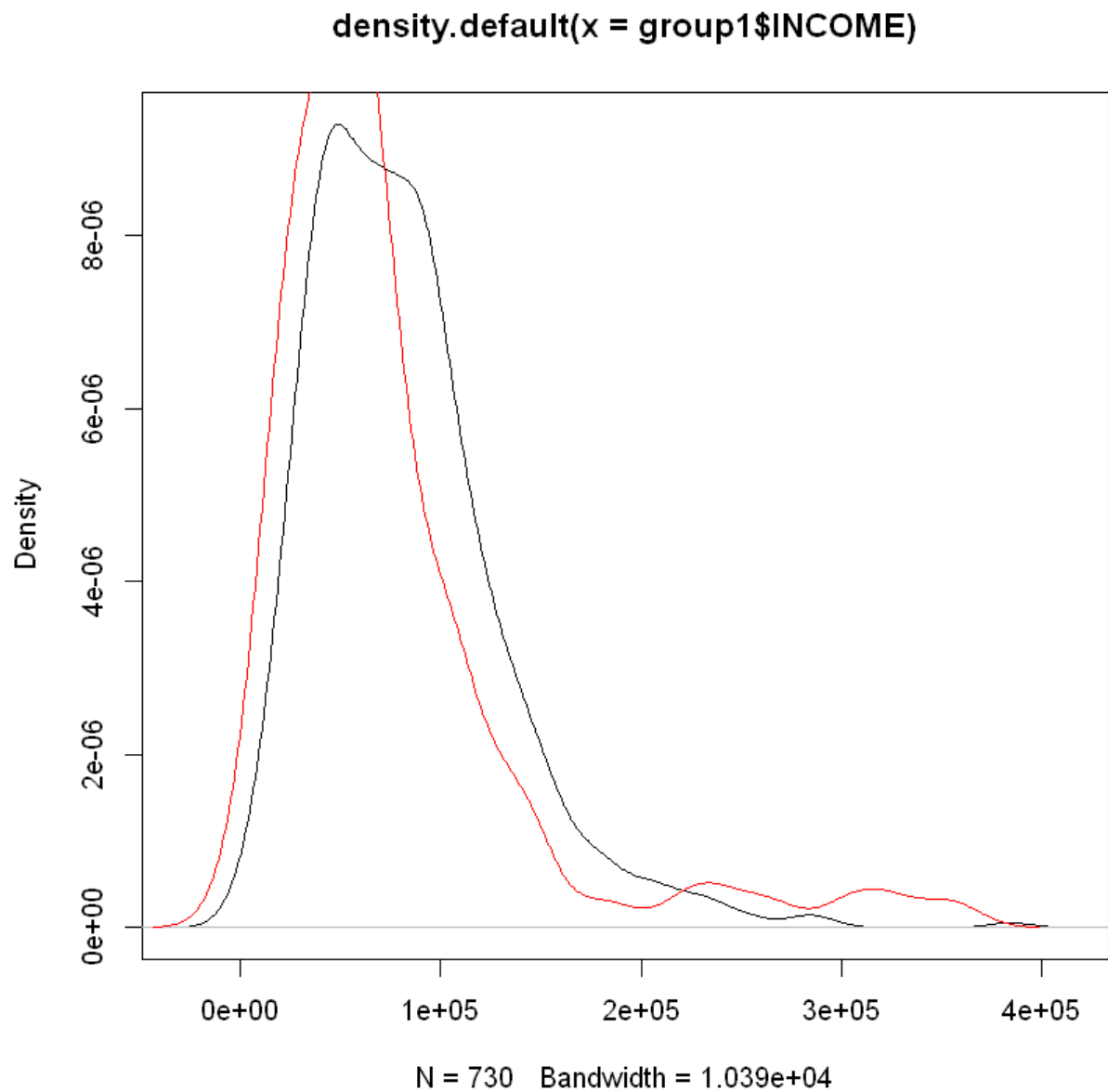
data: income by OCCAT1  
Fligner-Killeen:med chi-squared = 0.30022, df = 1, p-value = 0.5837

	Df	F value	Pr(>F)
group	1	0.7486684	0.3871494
	826	NA	NA

#### Wilcoxon rank sum test with continuity correction

data: income by OCCAT1  
W = 42590, p-value = 0.001079  
alternative hypothesis: true location shift is greater than 0

[1] "Yes, at 5% significant level, middle-class heads of households who work for someone else have higher incomes than heads of households who are self-employed"



8 (30 points) Question 8

Detergent manufacturers frequently make claims about the effectiveness of their products. A consumer-protection service decided to test the five best selling brands of detergent, where each manufacturer claims that its product produces the “whitest whites” in all water temperatures. The experiment was conducted in the following way. One hundred fifty white sheets were equally soiled. Thirty sheets were washed in each brand—10 with cold water, 10 with warm water, and 10 with hot water. After washing, the “whiteness” scores for each sheet were measured with laser equipment.

- Column 1: Water temperature code
- Column 2: Scores for detergent 1 (first 10 rows = cold water, middle 10 rows = warm, last 10 rows = hot)
- Column 3: Scores for detergent 2 (same format as column 2)
- Column 4: Scores for detergent 3 (same format as column 2)
- Column 5: Scores for detergent 4 (same format as column 2)
- Column 6: Scores for detergent 5 (same format as column 2)

Perform a statistic analysis using a 5% significance level to determine whether there is sufficient statistical evidence to infer that there are differences in whiteness scores between the five detergents, differences in whiteness scores between the three water temperatures, or interaction between detergents and temperatures. Explain the interaction, if you observe any. The dataset is available on the CSV file Question8.

```
In [13]: setwd("D:/BAX441/Midterm Exam")
Q8<-read.csv('Question8.csv')
df <- melt(Q8, id="Temperature")
colnames(df)<-c('Temperature','Brands','Rate')
temp<-as.factor(df$Temperature)
brands<-as.factor(df$Brands)
rates<-df$Rate
temp_brand<-df%>% unite(col='temp_brand',Temperature,Brands,sep='-', remove=FALSE)
temp_brand
```

executed in 63ms, finished 10:50:02 2019-11-27

**Hypothesis** Do interaction test first

$H_0$ : water temperatures and detergents do not have interaction to affect the whiteness scores

$H_1$ : water temperatures and detergents do have interaction to affect the whiteness scores

**Significant Level** 5%

In [15]:

```
#####2 Factor ANOVA #####
ANOVA8 <- aov(rates ~ temp + brands+brands*temp)
summary(ANOVA8)

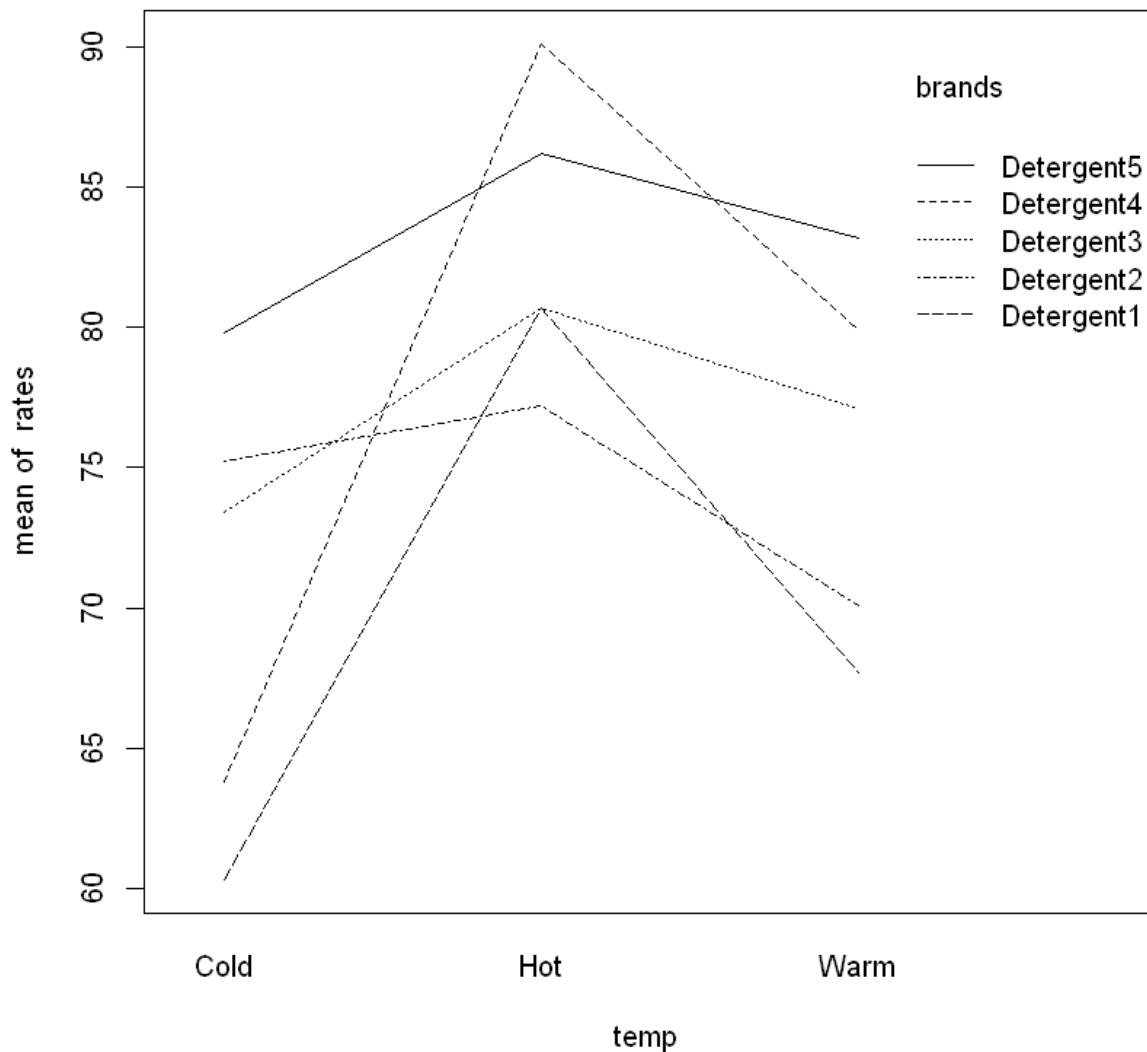
interaction.plot(temp, brands, rates)
interaction.plot(brands, temp, rates)
#####Conclusion#####
#At 5% significant Level, there is an evidence of a strong interaction between the two factors: w
#The effect that water temperature has on the rates is not independent of the brands.
#From the interaction plot and the test results, there is difference in whiteness score among dif
#temperature. Hot water always does the best job, while warm water is better than cold water exce
#There are also some differences among the brands, but the whiteness scores of different brands a
#on the water temperature.
```

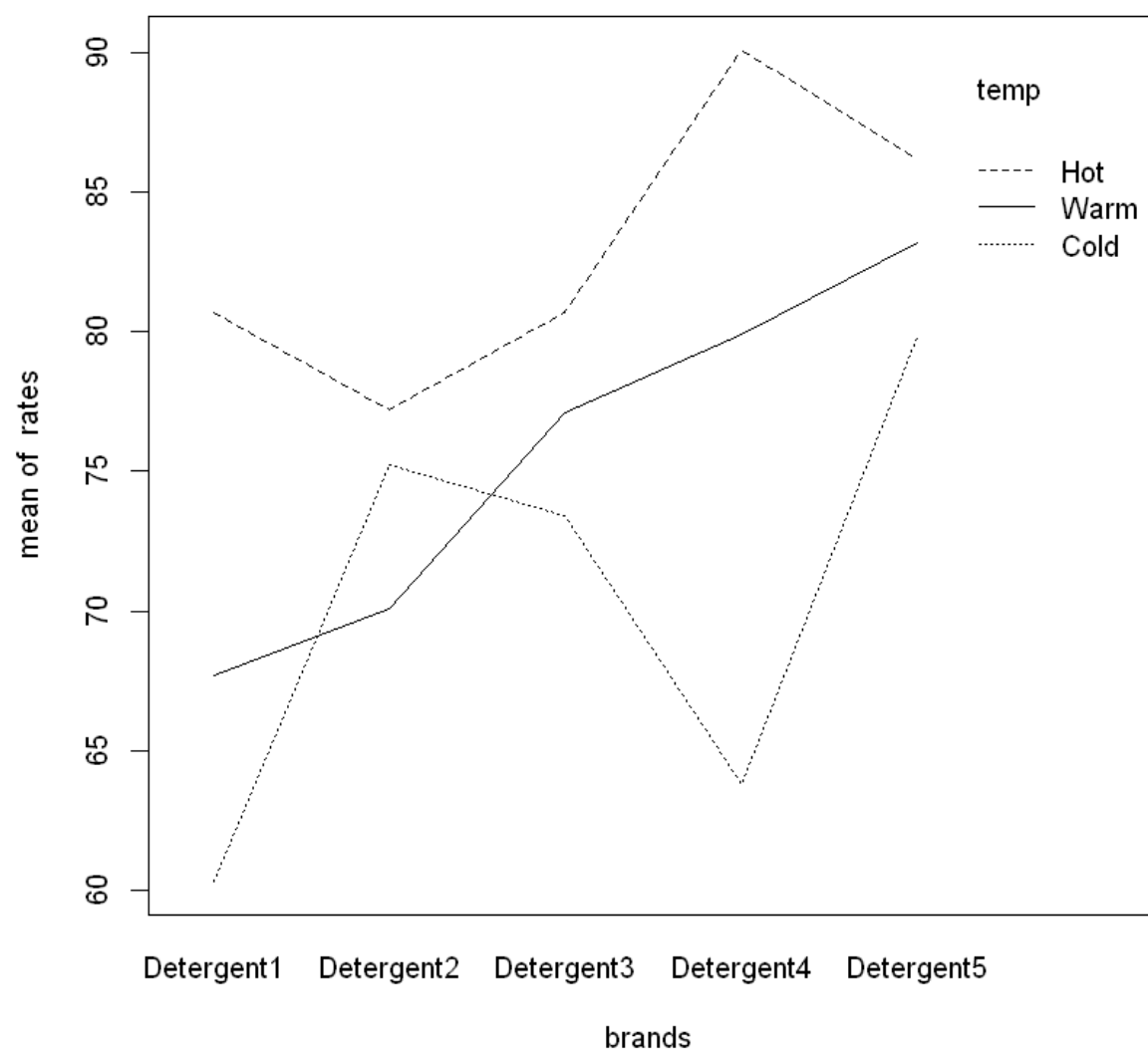
executed in 113ms, finished 10:51:56 2019-11-27

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
temp	2	3937	1968.5	17.824	1.35e-07 ***
brands	4	2967	741.9	6.717	5.83e-05 ***
temp:brands	8	2452	306.5	2.775	0.00714 **
Residuals	135	14910	110.4		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1







## 9 (20 points) Question 9

Suppose you are gambling on a roulette wheel. Each time the wheel is spun, the result is one of the outcomes 0, 1, and so on through 36. Of these outcomes, 18 are red, 18 are black, and 1 is green. On each spin you bet \$5 that a red outcome will occur and \$1 that the green outcome will occur. If red occurs, you win a net \$4. (You win \$10 from red and nothing from green.) If green occurs, you win a net \$24. (You win \$30 from green and nothing from red.) If black occurs, you lose everything you bet for a loss of \$6.

**9.1 a. (10 points) Use random numbers to generate 20 plays from this strategy. Each play should indicate the net amount won or lost. Then, based on these 20 outcomes, calculate a 95% confidence interval for the total net amount won or lost from 1000 plays of the game. Would you conclude that this strategy is a winning one for you?**

```
In [18]: spin<-seq(0,36,1)
outcome<-c(rep(list(4),18),rep(list(-6),18),24)
outcome
```

executed in 32ms, finished 11:00:37 2019-11-27

...

```
In [20]: set.seed(10)
samples <- sample(outcome, 20, replace=TRUE)
sample_sum<-sum(as.numeric(samples))
sample_mean<-mean(as.numeric(samples))
sample_sd<-sd(as.numeric(samples))
t95<-qt(0.975,19)

lower_limit<-(sample_mean-t95*sample_sd/sqrt(20))*1000
upper_limit<-(sample_mean+t95*sample_sd/sqrt(20))*1000
cat('The confidence interval for the total net amount won or lost form 1000 games is (',lower_lim
print('No, I would not conclude that this strategy is a winning one for me.')
```

executed in 33ms, finished 11:02:20 2019-11-27

The confidence interval for the total net amount won or lost form 1000 games is ( -2352.356 , 2352.356 ).[1] "No, I would not conclude that this strategy is a winning one for me."

**9.2 b. (10 points) Repeat part a, but with slightly changed rules. Now your betting strategy is the same, but if red occurs, your net gain is 5 (you win 11 from red, nothing from green). Comment on whether this slight change makes much of a difference in the mean total from 1000 bets.**

```
In [16]: spin<-seq(0,36,1)
outcome<-c(rep(list(5),18),rep(list(-6),18),24)
```

executed in 22ms, finished 10:58:06 2019-11-27

```
In [17]: set.seed(10)
samples <- sample(outcome, 20, replace=TRUE)
sample_sum<-sum(as.numeric(samples))
sample_mean<-mean(as.numeric(samples))
sample_sd<-sd(as.numeric(samples))
t95<-qt(0.975,19)

lower_limit<-(sample_mean-t95*sample_sd/sqrt(20))*1000
upper_limit<-(sample_mean+t95*sample_sd/sqrt(20))*1000
cat('The confidence interval for the mean total form 1000 games is (',lower_limit,', ',upper_limit
print('Yes, this slight change makes much of a difference in the mean total from 1000 bets. The l
```

executed in 36ms, finished 10:58:07 2019-11-27

The confidence interval for the mean total form 1000 games is ( -1987.592 , 3187.592 ).[1] "Yes, this slight change makes much of a difference in the mean total from 1000 bets. The lower limit and the upper limit become higher."

## 10 \* (30 points) Question 10

A/B testing has become a standard technique used to design Web pages for e-commerce. The success of A/B testing comes from taking advantage of randomization.

In this question, owners of a startup fashion Web site believe that getting a shopper to add an item into a shopping cart is a key step leading to purchases. They want to design their site to attract customers to place an item in a shopping cart. Staff have proposed two designs that we will name Design A and Design B. These designs take different approaches to enticing shoppers to add an item into a shopping cart. Customers visiting a Web site are randomly directed to a Web page that is based on one of the two proposed designs.

Based on prior experience, managers are most interested in the population of customers who visit the site during weekday evenings. Hence, the A/B test occurred during a 30-minute period on a Wednesday evening. Develop a 95% confidence interval for the difference between two proportions of visitors who add to shopping cart using designs A and

B. If there's no meaningful difference between the designs in the use of shopping carts, then management will let the software engineers choose the one that is easier to implement. What should the management decide? Dataset is provided on CSV file Question10.

```
In [33]: setwd("D:/BAX441/Midterm Exam")
Q10<-read.csv('Question10.csv')
Q10$CART<-ifelse(Q10[,3] == "No",0,1)
df_A<-Q10[Q10[,2]=="A",]
df_B<-Q10[Q10[,2]=="B",]
p_A<-sum(df_A[,4])/length(df_A[,4])
p_B<-sum(df_B[,4])/length(df_B[,4])
```

executed in 28ms, finished 13:39:46 2019-11-27

**H0:**  $p_A - p_B = 0$

**H1:**  $p_A - p_B \neq 0$

```
In [34]: prop.test(x=c(sum(df_A[,4]),sum(df_B[,4])), n=c(length(df_A[,4]),length(df_B[,4])),alternative = "two.sided",
print('At 5% significant level, there is difference between A and B, and B is better than A. The management should use design B.'
```

executed in 27ms, finished 13:39:56 2019-11-27

2-sample test for equality of proportions with continuity correction

```
data: c(sum(df_A[, 4]), sum(df_B[, 4])) out of c(length(df_A[, 4]), length(df_B[, 4]))
X-squared = 15.864, df = 1, p-value = 6.805e-05
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.04922122 -0.01643597
sample estimates:
 prop 1      prop 2 
0.02534751 0.05817610
```

```
[1] "At 5% significant level, there is difference between A and B, and B is better than A. The management should use design B."
```

## 11 (20 points) Question 11

Vendors work hard to put their products in front of consumers, whether choosing the right site for advertising on the Web or getting the best location in a crowded supermarket. Is the effort worthwhile? How could you tell?

To find an answer, vendors often collect their own data from experiments like the one considered in this question .

These data describe an experiment conducted by the maker of a popular brand of snack food (tortilla chips). Stores charge a premium for placing products on the middle shelf, near eye level for most shoppers. The bottom shelf is cheaper, and the top shelf “rents” for the least cost. The chipmaker wants to learn if shelf placement matters: Are customers who buy chips more likely to purchase the brand occupying the coveted middle shelf?

To obtain data, the chipmaker ran an experiment. Employees randomly changed the location of products on the shelves every few hours in a cooperating market. Each day, over a period of two weeks (14 days), agents randomly moved the packages from shelf to shelf. Check-out scanner receipts recorded results for customers who bought any of the three brands in the test (the chipmaker’s product, Competitor 1, Competitor 2). The data for each purchase identify the brand and the shelf location (top, middle, or bottom) and are shown on the CSV file Question11.

```
In [37]: setwd("D:/BAX441/Midterm Exam")
Q11<-read.csv('Question11.csv')

location<-as.factor(Q11[,1])
brand<-as.factor(Q11[,2])
sale<-Q11[,3]
Q11

test<-c(203,263,189)
competitor_1<-c(148,132,165)
competitor_2<-c(325,278,215)
df<-cbind(test,competitor_1,competitor_2)
rownames(df)<-c('top','middle','botton')

df
```

executed in 50ms, finished 14:03:32 2019-11-27

Shelf.Location	Brand	Sales
Top	Test	203
Top	Competitor 1	148
Top	Competitor 2	325
Middle	Test	263
Middle	Competitor 1	132
Middle	Competitor 2	278
Bottom	Test	189
Bottom	Competitor 1	165
Bottom	Competitor 2	215

	test	competitor_1	competitor_2
top	203	148	325
middle	263	132	278
botton	189	165	215

**11.1 a. (2 points) It would have been easier to record sales with the products in a fixed location for, say, one week, then move them, rather than run an experiment. What's the advantage of performing an experiment?**

The experiment allows us to record the same product's sales at different locations, and compare it with other two products. If we only record sales with the products in a fixed location for one week and then move them, the factor, date, may impact the results. If there is a holiday week, the sales may be impacted by the holiday. The experiment can provide us a more accurate result by eliminating or reducing some potential factors that may impact the outcome.

**11.2 b. (2 points) How can the chipmaker use the results of its experiment when negotiating shelf location with retail stores?**

If the middle shelf does have extra impact on sales, the chipmaker would accept the premium rental fee when the extra sales can cover the premium. However, if there is no difference among three shelf, the chipmaker can choose the cheapest one or show the result to the store and ask them eliminate the premium.

### 11.3 c. (10 points) Run an appropriate statistical test. Are the required conditions satisfied? Clearly state which required conditions did you test for.

#### Requirement Test

- #1. Randomly selected sample
- #2. Categories defined in the table are mutually exclusive.
- #3. Have enough data in every cell. Expected counts > 5. ☐

#### Categorical Test for Independency

Test the Treatment Effect:

$H_0$ : Shelf location and sales for the producers are independent

$H_1$ : Shelf location and sales for the producers are dependent

```
In [40]: chisq.test(df)
chisq.test(df)$expected

#####Conclusion#####
print('Yes, at 5% significant level, the shelf location and sales for the producers are dependent

executed in 39ms, finished 14:20:09 2019-11-27
```

Pearson's Chi-squared test

data: df  
X-squared = 28.876, df = 4, p-value = 8.284e-06

	test	competitor_1	competitor_2
top	230.8551	156.8405	288.3045
middle	229.8306	156.1444	287.0250
botton	194.3144	132.0151	242.6705

```
[1] "Yes, at 5% significant level, the shelf location and sales for the producers are dependent."
```

### 11.4 d. (3 points) What problem might arise if the experiment lasted, say, 2 months?

People who are loyalty of a specific brand may be confused and buy less or non from the store. The experiment that last too long and changes shelf too frequent may impact store's revenue and the retention of customers.

### 11.5 e. (3 points) Identify any key limitations of the experiment and its results.

The experiment did not consider other potential factors that impact the sales, such as whether the brand did the promotion. The result of this limitation let the company draw wrong conclusion about the shelf location and lost the opportunity to earn more money by placing product at the middle of the shelf.

The experiment results show the shelg locaiton and sales of the producers are dependent, but cannot say the middle shelf is the best place for products. From the expection table, we can find that there is not very large difference between top and middle shelf.

In [ ]: