Department of Information and Communication Engineering (ICE)

Pabna University of Science and Technology

Faculty of Engineering and Technology

B.Sc. (Engineering) 2nd Year 2nd Semester Examination-2019

Course Code: Stat-2202   Course Title: Sampling Distribution and Hypothesis Testing Sessional

Topics: Lab Question and Solution by R programming language.

Name: Saidul Islam

Roll: 180614

Ex. No.	Name of Experiments
01	Find out the point estimate of the population mean and interval estimate of the population mean. Where 30 students quiz test marks is (2,4,3,23,25,27,28,13,15,16,20,14,35,33,32,21,35,40,42,22,33,13,17,20,25,29,27,40,38,31), total marks 50. Here polulaion size N=30 and sample size n=10. also illustrate the sample size determination, sampling distribution for mean and check the unbiasness of the population mean.
02	Two dice rolled, S is the sum of both faces, Find the expectation of S, E(s) and variance of S, V(s). Plot the distribution of S and dice D.
03	A herd of 1500 steer was fed a special high protein gain for a month. A random sample of 29 was weighted and had gained an average of 6.7 pounds. If the sd of weight gain for the entire herd is 7.1. Test the hypothsis at 5% level of significance that the average weight gain per steer for the month was more than 5 pounds. Also comments on the test using the p-value. Create the confidence interval.
04	In order to find out whether children with chronic diarrhea have the same average hemoglobin level(Hb) that is normally seen in healthy children in the same area, a random sample of 10 children with chronic diarrhea are selected, and their Hb levels <g/dl) are obtained as follows: 12.3, 11.4, 14.2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2 Do the data provide sufficient evidence to indicate that the mean Hb level for children with chronic diarrhea is less than the normal value of 14.6 (g/dl)? Test at 0.01 level of significance. Draw a boxplot and normal plot for this data and comments.
05	In order to find out whether children with chronic diarrhea have the same average hemoglobin level(Hb) that is normally seen in healthy children in the same area, a random sample of 10 children with chronic diarrhea are selected, and their Hb levels <g/dl) are obtained as follows: 12.3, 11.4, 14,2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2 another random sample of 12 children with chronic diarrhea are 11.1, 17.2, 13.4, 15.2, 14.1, 13.0, 12.5, 11.5, 12.7, 14.5, 15.3, 14.0. Is there any difference in the mean Hb label between the two groups of children???

| 06 | Test the hypothesis that the mean systolic blood pressure of healthy subjects (status-0) and subject with hypertension(status-1) are equal, have do= 0. The dataset contains n1= 25 subjects with status-0 and n2= 30 with status-1.

Status-0: (120, 115, 94, 118, 111, 102, 102, 131, 104, 107, 115, 139, 115, 113, 114, 105, 115, 134, 109, 109, 93, 118, 109, 106, 125).

Status-1: (150, 142, 119, 127, 141, 149, 144, 142, 149, 161, 143, 140 , 148, 149, 141, 146, 159, 152, 135, 134, 161, 130, 125, 141, 148 ,153, 145, 137, 147, 169). |

| 07 | The 126 people have some doing smoking and some do not smoke. Some of this type of data are tabulated is given below: |

Disses / Smoking	Heart disses	Not heart disses	Total
YES	55	16	71
NO	23	32	55
Total	78	48	N=126

Is there any association between smoking and heart disses for the given data.???

| 08 | There are two COVID-19 testing booths, we test some people and their recorded data is below, where the numbers of people of booth-1 are 11 and the numbers of people of booth-2 are 10:

Booth-1: positive, positive, negative, positive, negative, negative, positive, positive, positive, negative, positive.

Both-2: negative, negative, negative, positive, positive, negative, positive, negative, negative, negative.

Is there any relation between two both??? |

Solution of the problems:

Lab Problem: 01

Question:
Find out the point estimate of the population mean and interval estimate of the population mean. Where 30 students quiz test marks is (2,4,3,23,25,27,28,13,15,16,20,14,35,33,32,21,35,40,42,22,33,13,17,20,25,29,27,40,38,31), total marks 50. Here polulaion size N=30 and sample size n=10.
also illustrate the sample size determination, sampling distribution for mean and check the unbiasness of the population mean.

R source code:

```
#Problem-01: Point estimation and Interval estimation
#Where 30 students quiz test marks, total marks 50
#Find out the point estimate of population mean and
#interval estimate of population mean,, Population size N=30 and Sample size n=10.

data<-c(2,4,3,23,25,27,28,13,15,16,20,14,35,33,32,21,35,40,42,22,33,13,17,20,25,29,27,40,38,31)
length(data)


#for cheak normality
qqnorm(data)
qqline(data)

set.seed(125)
x<-sample(data,10,replace=TRUE)
x

y<-mean(x)
y     #point estimate for mean = 21.8

sigma= sd(data)
sigma

#Interval estimate
qnorm(0.025,0,1) # -1.96

#lower class interval
l= y-(((1.96)*(sigma))/sqrt(10))
l

#Upper class interval
u= y+(((1.96)*(sigma))/sqrt(10))
u
#95% confidance interval for population mean is (14.97, 28.63)


#sample size determination
# There are two ways
#  1). We know that the range is four times of standerd deviation(signam).
#     so we will get the sigma value if the range divided by 4.
#  2).
     n= ((2*1.96*sigma)/2)^2
```

```
    n  # Probable Sample size will be  465.8536 ~ 466

pnorm(-1.96,0,1)  #0.025


#SAMPLING DISTIBUTION FOR MEAN
choose(30,10)    # 30045015

set.seed(125)
a<-rep(0,3004)
for(j in 1:3004){
a[j]<-mean(sample(data,10,replace=TRUE))}
mean(a)  #Expected value E(x ber)= 24.12693

mean(data)  #population mean
bais=mean(a)-mean(data)
bais #bias is 0.02693076 that is almost zero, so Sampling mean or (x ber) is an unbiased estimator of
population mean µ.
hist(a)
qqnorm(a)
qqline(a)
#Comment: Sampling mean is an unbiased estimator of population mean.




###############################################################################################
######          The extra part which is no needed for this code          ############
###############################################################################################

curve(dnorm(x), xlim=c(-3.5, 3.5), ylab="density", main="Standard Normal Distribution")
dnorm(x=0)
curve(pnorm(x), xlim=c(-3.5, 3.5), ylab="probability", main="Standard Normal Cumulative Distribution")

ami<-function(x){
(1/(sqrt(2*pi)))*exp(-0.5*x^2)
}

ami(-1.96)   #dnorm(-1.96)=ami(-1.96)

integrate(ami, lower=-Inf, upper=3)

################Chi-square distrubution#########
curve(dchisq(x,df=3), xlim=c(0,10), ylim=c(0,1), col="red", main="PDF of chi-square distribution")

###### t distribution#####
curve(dt(x,df=3), xlim=c(-4,4), col="red", main="PDF of t distribution")
b<-rt(100,3)
b
hist(b)
```

Lab problem:02

Question:
Two dice rolled, S is the sum of both faces, Find the expectation of S, E(s) and variance of S, V(s). Plot the distribution of S and dice D.

R source code:

```
#Problem-02: Two dice rolled, S is the sum of both face, Find the E(s) and V(s)

#Vector of outcomes
s<- 2:12

#vector of probabilites
ps<- c(1:6, 5:1) / 36

#Expectation of s
es<- sum(s* ps)
es

# Variance of s
esq<- sum((s^2) * ps)
vs<- esq-es^2
vs


# Divide the plotting area into one row with two columns
par(mfrow = c(1,2))

#plot the distribution of s
barplot(ps,
        ylim=c(0, 0.2),
        xlab= "S",
        ylab="Probabilites",
        col ="red",
        space= 0,
        main= "Sum of two dice rolls")


#plot the distribution of D
probability <- rep(1/6, 6)
names(probability) <- 1:6
barplot(probability,
        ylim=c(0, 0.2),
        xlab= "D",
        ylab="Probabilites",
        col ="green",
        space= 0,
        main= "Outcomes of a sngle dice rolls")




################## Home Task  ##############
#Two dice rolled, S is the sum of both face that is gratter than 2.
#  i.e sum of both faces > 2.
#Find the E(s) and V(s)
```

```r
#Vector of outcomes for (sum of both face)>2.
s1<- 3: 12
s1

#vector of probabilites
ps1<- c(2:6, 5:1) / 36
ps1

sum(ps1)  # total probability

#Expectation of s
es1<- sum(ps1* s1)
es1


# Variance of s
esq1<- sum((s1^2) * ps1)
vs1<- esq1-es1^2
vs1


# Divide the plotting area into one row with two columns
par(mfrow = c(1,2))

#plot the distribution of s
barplot(ps1,
        ylim=c(0, 0.2),
        xlab= "S",
        ylab="Probabilites",
        col ="yellow",
        space= 0,
        main= "Sum of two dice rolls")


#plot the distribution of D
probability <- rep(1/6, 6)
names(probability) <- 1:6
barplot(probability,
        ylim=c(0, 0.2),
        xlab= "D",
        ylab="Probabilites",
        col ="green",
        space= 0,
        main= "Outcomes of a sngle dice rolls")
```

Lab problem:03

Question:
A herd of 1500 steer was fed a special high protein gain for a month. A random sample of 29 was weighted and had gained an average of 6.7 pounds. If the sd of weight gain for the entire herd is 7.1. Test the hypothsis at 5% level of significance that the average weight gain per steer for the month was more than 5 pounds. Also comments on the test using the p-value. Create the confidence interval.

R source code:

```
##Problem-03: A herd of 1500 steer was feed a speacial high protein gain for month.
# A random sample of 29 was weighted and had gain an average of 6.7 pounds. If the
# sd of weight gain for the entire herd is 7.1. Test the hypothsis at 5% level of
# significance that the average weight gain per steer fop the month was more than
# 5 pounds. Also comments on the test using p-value.

#Ho: mue equal 5
#H1: mue greater than 5
x_ber <- 6.7
mue <- 5
sd <- 7.1
n <- 29
alpha= 0.05

z_tab <- qnorm(0.05,lower.tail=FALSE)
z_tab  #1.644

z_cal <- (x.ber-mue)/(sd/sqrt(n))
z_cal  #1.289
#Comments: Ho is accepted, Since z.cal<z.tab

####Using p-value ####
p_value <- pnorm(z.cal,lower.tail=FALSE)
p_value  # 0.098
#Comments: Since p-value= 0.098 > 0.05, so Ho is accepted
#This p-value also indicate that if we test the hipothesis with maximum 9.8%
#level of signicance, Ho was also accepted. But if we test above 9.8% then
#Ho was rejected.


####If it was two tail test ####
# i.e. Ho equal 5
#and H1: mue not equal 5

alpha= 0.05
z_tab1<- qnorm(0.025)
z_tab1 #-1.96
z_tab2<- qnorm(0.975)
z_tab2 #1.96

#Also Ho is accepted, since z_cal=1.289 fall between
# -1.96 to 1.96

#Using p-value##
p_value2<- 2*pnorm(z_cal, lower.tail=FALSE)
p_value2 # 0.19
```

#Here, p_value=0.19>0.05, so Ho is accepted

#######Now make the confidance interval#####
# 95% confidance interval for mue(Population mean)
CI<-c(x_ber+z_tab1*sd/sqrt(n), x_ber+z_tab2*sd/sqrt(n))
CI # 4.1159 to 9.2840
#we are 95% sure, we have confidance that the average
weight gain is between 4.1159 to 9.2840 due to applying high protien.

Lab problem:04

Question:
In order to find out whether children with chronic diarrhea have the same average hemoglobin level(Hb) that is normally seen in healthy children in the same area, a random sample of 10 children with chronic diarrhea are selected, and their Hb levels <g/dl) are obtained as follows: 12.3, 11.4, 14.2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2
Do the data provide sufficient evidence to indicate that the mean Hb level for children with chronic diarrhea is less than the normal value of 14.6 (g/dl)? Test at 0.01 level of significance. Draw a boxplot and normal plot for this data and comments.

R source code:
```
###Problem-04: In order to find out whether children with
#choronic diarrhea have the same average hemoglobin
#level(Hb) that is normally seen in healthy children in the same area
#, a random sample of 10 children with chonic diarrhea are selected and there
# Hb levels <g/dl) are obtained as follows:
# 12.3, 11.4, 14.2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2

#Do the data provide suffitient evidance to indicate that the
#mean Hb level for children with chorinc diarrhea is
#less than of the normal value of 14.6 (g/dl)?
#Test at 0.01 level of significance.
#Draw a boxplot and normal plot for this data and comments.


#Ho: mue equal 14.6
# and H1: mue less than 14.6
data<-c(12.3,11.4,14.2,15.3,14.8,13.8,11.1,15.1,15.8,13.2)
n<-length(data)
n
x_ber<- mean(data)
x_ber

sample_sd<- sd(data)
sample_sd

mue<- 14.6

t_tab<- qt(0.01, n-1)
t_tab #-2.821

t_cal<- ((x_ber-mue)/(sample_sd/sqrt(n)))
```

```
t_cal  #-1.71

#Comments: since t_cal>t_tab, so Ho is accepted.


#Using p-value########
p_value<- pt(t_cal, n-1)
p_value  #0.059
#Comments: since p_value=0.059>0.01, so Ho is accepted.



###Using function#####
t.test(data, mu=14.6, conf.level=0.99, alternative="less")


boxplot(data,ylab="Hb lebel", col="red")

qqnorm(data, main="Normal Q-Q plot of Hb lebel")
qqline(data)
```

Lab problem:05

Question:
In order to find out whether children with chronic diarrhea have the same average hemoglobin level(Hb) that is normally seen in healthy children in the same area, a random sample of 10 children with chronic diarrhea are selected, and their Hb levels <g/dl) are obtained as follows: 12.3, 11.4, 14,2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2
another random sample of 12 children with chronic diarrhea are 11.1, 17.2, 13.4, 15.2, 14.1, 13.0, 12.5, 11.5, 12.7, 14.5, 15.3, 14.0.
Is there any difference in the mean Hb label between the two groups of children???

R source code:
```
###problem-05: In order to find out whether children with
#choronic diarrhea have the same average hemoglobin
#level(Hb) that is normally seen in healthy children in the same area
#, a random sample of 10 children with chonic diarrhea are selected and there
# Hb levels <g/dl) are obtained as follows:
# 12.3, 11.4, 14,2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2

#another random sample of 12 children with chonic diarrhea are
#  11.1, 17.2, 13.4, 15.2, 14.1, 13.0, 12.5, 11.5, 12.7, 14.5, 15.3, 14.0
# is there any differance between in mean  Hb label between the two group of children???



##############Solution##############
#Ho: mue1=mue2
#H1: mue not equal mue2

level<- 0.05
alpha<- 0.05/2
alpha  #0.025
```

```
data1<- c(12.3, 11.4, 14.2, 15.3, 14.8, 13.8, 11.1,15.1,15.8,13.2)
n1<- length(data1)
n1
s1<- sd(data1)
s1
x_ber1<- mean(data1)
x_ber1


data2<- c(11.1, 17.2, 13.4, 15.2, 14.1, 13.0, 12.5, 11.5, 12.7, 14.5, 15.3, 14.0)
n2<- length(data2)
n2
s2<- sd(data2)
s2
x_ber2<- mean(data2)
x_ber2


t_tab1<- qt(alpha, n1+n2-2)
t_tab1 #-2.085
t_tab2<- qt(1-alpha, n1+n2-2)
t_tab2 #2.085


######## Checking the variance equal or not ##########
boxplot(list(sample_1=data1, sample_2=data2), col="red")
#If the middle line of both boxplot are very colse then is indicate equal variances.
ratio_sd<- s1/s2
ratio_sd  # 0.961985 is close to 1, that's why we can say that they have equal variances.




sp<- sqrt((((n1-1)*s1^2)+((n2-1)*s2^2))/(n1+n2-2))
sp

t_cal<- (x_ber1 - x_ber2)/sqrt(sp^2*((1/n1)+(1/n2)))
t_cal #  -0.01150547

#Comments: Ho is accepted, since t_cal= -0.0137 fall between the t_tab value of -2.085 to 2.085


######using p-value###
p_value<- 2*pt(t_cal, n1+n2-2)
p_value  # 0.9891

#Comments: since p_value= 0.9891> 0.05, so Ho is accepted.
######### 95% Confidance Interval #########
CI<-c((x_ber1-x_ber2)+(t_tab1*sp*sqrt((1/n1)+(1/n2))), (x_ber1-x_ber2)+t_tab2*sp*sqrt((1/n1)+(1/n2)))
CI  # -1.519183  1.502516
# -1.519183   Cofidance interval 95% lower
# 1.502516   Cofidance interval 95% upper
```

```
##########  Using t.test function ##########
?t.test
t.test(data1, data2, alternative="two.sided", mu=0, paired= FALSE, conf.level=0.95)




########################  For Another Alternative hypothesis #############
# Ho: mue1=mue2
#H1: mue1>mue2

t_cal  #  -0.0137

?qt
t_tabx<- qt(0.05, n1+n2-2, lower.tail=FALSE)
t_tabx  #1.724718

#Comments: Ho is accepted, since t_cal<t_cal.



############ Using P-value ##########
p_valuex<- pt(t_cal, n1+n2-2)
p_valuex  # 0.4945988

#Comments: Ho is accepted, since p-valuex= 0.4945988> 0.05
```

Lab problem:06

Test the hypothesis that the mean systolic blood pressure of healthy subjects (status-0) and subject with hypertension(status-1) are equal, have do= 0. The dataset contains n1= 25 subjects with   status-0 and n2= 30 with status-1.

Status-0: (120, 115, 94, 118, 111, 102, 102, 131, 104, 107, 115, 139, 115, 113, 114, 105, 115, 134, 109, 109, 93, 118, 109, 106, 125).

Status-1: (150, 142, 119, 127, 141, 149, 144, 142, 149, 161, 143, 140 , 148, 149, 141, 146, 159, 152, 135, 134, 161, 130, 125, 141, 148 ,153, 145, 137, 147, 169).

R source code:
```
# Problem-6: Test the hypothesis that the mean systolic blod pressure
of healdy subject(status-0) and subject with hypertension(status-1)
are equal, have do= 0. The dataset contains n1= 25 subject with
status-0 nad n2= 30 with status-1.

(120, 115, 94, 118, 111, 102, 102, 131, 104, 107, 115, 139, 115, 113, 114,
105, 115, 134, 109, 109, 93, 118, 109, 106, 125)

(150, 142, 119, 127, 141, 149, 144, 142, 149, 161, 143, 140 ,
148, 149, 141, 146, 159, 152, 135, 134, 161, 130, 125, 141, 148 ,153,
145, 137, 147, 169)
```

```
##############Solution##############
Ho: mue1=mue2
H1: mue1 not equal mue2

data1<- c(120, 115, 94, 118, 111, 102, 102, 131, 104, 107, 115, 139, 115, 113, 114,
105, 115, 134, 109, 109, 93, 118, 109, 106, 125)
n1<- length(data1)
n1
s1<- sd(data1)
s1
x_ber1<- mean(data1)
x_ber1


data2<- c(150, 142, 119, 127, 141, 149, 144, 142, 149, 161, 143, 140 ,
148, 149, 141, 146, 159, 152, 135, 134, 161, 130, 125, 141, 148 ,153,
145, 137, 147, 169)
n2<- length(data2)
n2
s2<- sd(data2)
s2
x_ber2<- mean(data2)
x_ber2



###########Checking the normality ########
par(mfrow=c(1,2))
qqnorm(data1)
qqline(data1)

qqnorm(data2)
qqline(data2)


############Checking the variance equal or not ###########
boxplot(list(status_0=data1, status_1=data2), col="red")
ratio_sd<- s1/s2
ratio_sd #  1.018009 is close to 1, that's why we can say that they have equal variances.

alpha<- 0.05

t_tab1<- qt(alpha/2, n1+n2-2)
t_tab1  # -2.005746
t_tab2<- qt(1-alpha/2, n1+n2-2)
t_tab2  #2.005746




sp<- sqrt(((n1-1)*s1^2+(n1-1)*s2^2)/(n1+n2-2))
sp

t_cal<- (x_ber1 - x_ber2)/sqrt(sp^2*((1/n1)+(1/n2)))
```

t_cal #-10.9903

#Comments: t_cal= -10.9903 is not fall in between the Critical value -2.005746 to 2.005746, so the Null hypothsis(Ho) is rejected.

###########using p-value##############
p_value<- 2*pt(t_cal, n1+n2-2)
p_value #2.793985e-15 that is < 0.05, so Ho is rejected.

########### using t.test function ############
t.test(data1, data2, alternative="two.sided", mu=0, paired= FALSE, conf.level=0.95)

Lab problem:07

Question:

The 126 people have some doing smoking and some do not smoke. Some of this type of data are tabulated is given below:

Disses ⟍ Smoking	Heart disses	Not heart disses	Total
YES	55	16	71
NO	23	32	55
Total	78	48	N=126

Is there any association between smoking and heart disses for the given data.???

R source code:

Problem-7: Look the folder: E:\2'2\Sampling and hypothisis testing\Lab\Lab with R\Lab 07.docx

################## Solution ##############

Ho: There is no association with smoking and heart disses.
H1: There is a association with smoking and heart disses.

m<- matrix(c(55, 16, 23, 32), ncol=2, byrow=TRUE, dimnames=list(c("yes", "no"), c("disses", "not_disses")))
m

c1<- sum(m[,1])
c1

```r
c2<- sum(m[,2])
c2

r1<- sum(m[1,])
r1

r2<- sum(m[2,])
r2

n<- sum(m)
n


E11<- (c1*r1)/n
E11

E21<- (c1*r2)/n
E21

E12<- (c2*r1)/n
E12

E22<- (c2*r2)/n
E22


chi_cal<- (((m[1]-E11)^2)/E11) + (((m[2]-E21)^2)/E21) + ((m[3]-E12)^2/E12) + ((m[4]-E22)^2/E22)
chi_cal  #16.69906

#df=(r-1)(c-1) , this is the fourmula to find the df.

chi_tab<- qchisq(0.05, df=1, lower.tail=FALSE)
chi_tab  # 3.841459

#Comments: chi_cal= 16.69906 > chi_tab= 3.841459 , so Ho is rejected.
# i.e smoking is associated with heart disses.


######  P-value ##########
p_value<- pchisq(chi_square, df=1, lower.tail=FALSE)
p_value  #  4.38026e-05

#Coments: p-value< 0.05, so Ho is rejected.




############################### Using chisq.test function #############
chisq.test(m)

# Here p-value =  9.56e-05 < 0.05, so Ho is rejected.
#i.e smoking is associated with heart disses.
```

Lab problem:08

There are two COVID-19 testing booths, we test some people and their recorded data is below, where the numbers of people of booth-1 are 11 and the numbers of people of booth-2 are 10:

Booth-1: positive, positive, negative, positive, negative, negative, positive, positive, positive, negative, positive.

Both-2: negative, negative, negative, positive, positive, negative, positive, negative, negative, negative.

Is there any relation between two both???

R source code:
#Probelem-08: There are two COVID-19 testing booths, we test some people
and their recorded data is below, where the numbers of people of booth-1 is 11 and
the numbers of people of booth-2 is 10:

Booth-1: positive,positive, negative,positive, negative, negative,positive,
positive,positive, negative, positive

Both-2: negative, negative, negative, positive, positive,negative, positive,
negative, negative, negative

is there any relation between two both???????

```
################### Solution  #############
Ho: There is no relation between booth-1 and booth-2.
H1: There is relation between booth-1 and booth-2.


booth_1<- c("positive","positive", "negative","positive", "negative", "negative",
"positive","positive","positive", "negative", "positive")

#booth_1<- c(1, 1, 0, 1,0, 0, 1, 1, 1, 0, 1)
#table(booth_1)



booth_2<- c("negative", "negative", "negative", "positive", "positive",
"negative", "positive", "negative", "negative", "negative")

#booth_2<- c(0, 0, 0, 1, 1, 0, 1, 0, 0, 0)
#table(booth_2)



x_table1<- table(booth_1)
x_table1

x_table2<- table(booth_2)
x_table2
```

```r
m<- matrix(c(4, 7, 7, 3), ncol=2, byrow=TRUE, dimnames=list(c("Booth-1", "Booth-2"), c("negative",
"positive")))
m

c1<- sum(m[,1])
c1

c2<- sum(m[,2])
c2

r1<- sum(m[1,])
r1

r2<- sum(m[2,])
r2

n<- sum(m)
n


E11<- (c1*r1)/n
E11

E21<- (c1*r2)/n
E21

E12<- (c2*r1)/n
E12

E22<- (c2*r2)/n
E22  # 4.761905 < 5, so we need to do "Yates" continuity correction.


chi_yates<- (((abs(m[1]-E11)-0.5)^2)/E11) + (((abs(m[2]-E21)-0.5)^2)/E21) + ((abs(m[3]-E12)-0.5)^2/E12)
+ ((abs(m[4]-E22)-0.5)^2/E22)
chi_yates  #1.218781

#df=(r-1)(c-1) , this is the fourmula to find the df.

chi_tab<- qchisq(0.05, df=1, lower.tail=FALSE)
chi_tab  # 3.841459
#Comments: chi_yates = 1.218781 < chi_tab= 3.841459 , so Ho is accepted.
#i.e. There is no relation between booth-1 and booth-2.

j

######  P-value ##########
p_value<- pchisq(chi_yates, df=1, lower.tail=FALSE)
p_value  # 0.2696

#Coments: p-value = 0.2696 > 0.05, so Ho is accepted.
```

```
############################## Using chisq.test function #############
chisq.test(m)

# Here  p-value = 0.2696 > 0.05, so Ho is accepted.
#i.e There is no relation between booth-1 and booth-2.
```