NAME

Shaikh Mohammad Shaad Shabbir Ali

REG. NO.

21BCE1542

EXPERIMENT 06

A manufacturer of pins knows that 2% of his products are defective. If he sells pins in boxes of 20 and find the number of boxes containing

(i) at least 2 defective

(ii) exactly 2 defective

(iii) at most 2 defective

pins in a consignment of 1000 boxes

(iv) plot the distribution

#(v) E(x)

#(vi) variance of X?

```
# number of trials
m = 20
m

#probability of success
ps = 0.02
# poisson parameter
lambda = m*ps
lambda

# (i) at least 2 defective
p1 = sum(dpois(2:m, lambda))
p1

# (i) number if boxes containing at least two defectives
round(1000*p1)

# exactly rwo defectives
p2 = dpois(2, lambda)
```

p2

(ii) number of boxes containing exactly 2 defectives

round(1000*p2)

at most 2 defectives

p3 = sum(dpois(0:2, lambda))

p3

(iii) number of boxes containing at most two defectives

round(1000*p3)

(iv) plot the distribution

x1 = 0:m

px1 = dpois(x1, lambda)

px1

plot(x1, px1, type="h", xlab="values of x", ylab = "probability distribution of x", main="Poisson distribution")

#(v) E(x)

Ex1 = weighted.mean(x1, px1)

Ex1

#(vi) variance of x

varx1 = weighted.mean(x1*x1, px1)-(weighted.mean(x1, px1))^2

varx1

#Normal distribution

A company finds that the time taken by ine of its engineers to complete or repair job has a normal distribution

with mean 20 minutes and SD 5 minutes. State what proportion of the jobs take

```
# (i) Less than 15 minutes
# (ii) More than 25 minutes
# (iii) Between 15 and 25
minutes # (iv) Plot the
distribution
# (v) Table the distribution

# Generating the data
x = seq(0, 40)
x

# find the density function of x
y = dnorm(x, mean=20, sd = 5)
y

#plot the normal distribution curve
plot(x,y,type='l')
#Proportion of jobs take less than 15 minutes
p1 = pnorm(15,mean=20, sd = 5)
p1

x2 = seq(0,15)
x2

y2 = dnorm(x2, mean=20, sd = 5)
y2

polygon(c(0, x2, 15), c(0,y2,0), col='yellow')

#Proportion of jobs that take more than 25 minutes
p2 = pnorm(40, mean=20, sd=5)- pnorm(25,mean=20,sd=5)
p2
```

```
x1 = seq(25,40)
x1

y1 = dnorm(x1, mean=20, sd=5)
y1

polygon(c(25,x1,40), c(0, y1, 0), col='red')

#Proportion of jobs take between 15 and 25
p3 = pnorm(25, mean=20, sd=5) - pnorm(15, mean=20, sd=5)
p3

x3 = seq(15,25)
x3

y3 = dnorm(x3, mean=20, sd=5)
y3

polygon(c(15,x3,25), c(0, y3, 0), col='green')

#Probabilty distribution
data.frame(p1, p2, p3)
```

OUTPUT

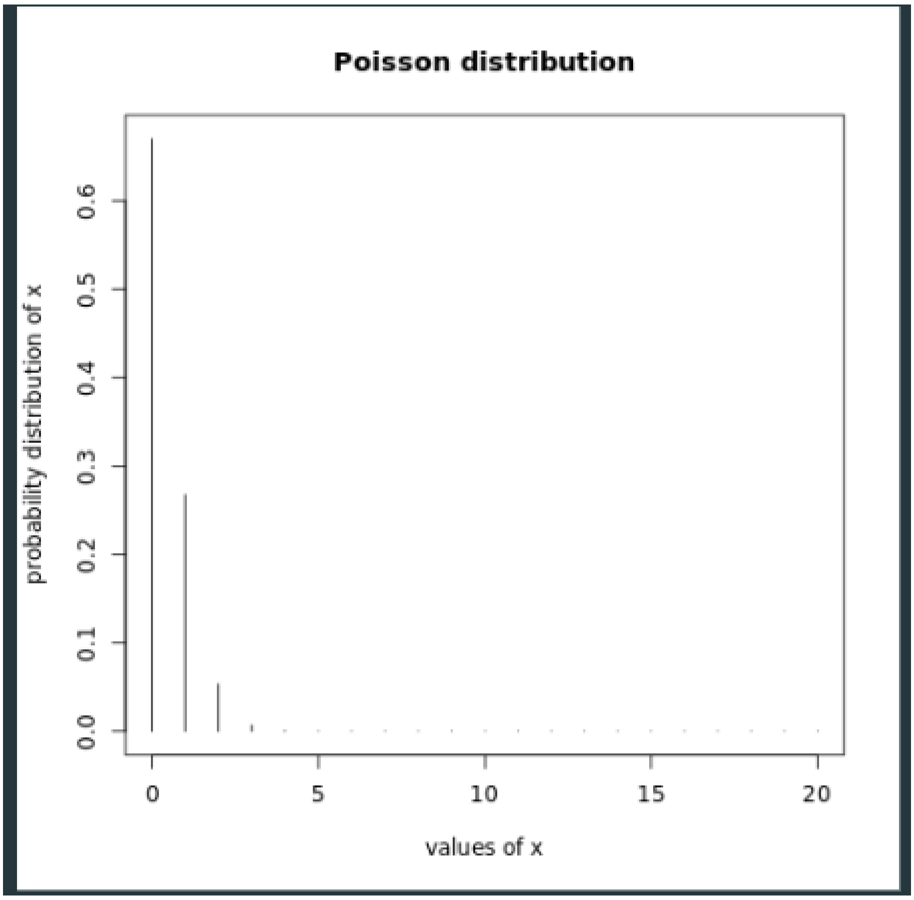
Output

```
[1] 20
[1] 0.4
[1] 0.06155194
[1] 62
[1] 0.0536256
[1] 54
[1] 0.9920737
[1] 992
 [1] 6.703200e-01 2.681280e-01 5.362560e-02 7.150080e-03 7.150080e-04
 [6] 5.720064e-05 3.813376e-06 2.179072e-07 1.089536e-08 4.842383e-10
[11] 1.936953e-11 7.043466e-13 2.347822e-14 7.224067e-16 2.064019e-17
[16] 5.504051e-19 1.376013e-20 3.237677e-22 7.194838e-24 1.514703e-25
[21] 3.029406e-27
[1] 0.4
[1] 0.4
 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
[26] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
 [1] 2.676605e-05 5.838939e-05 1.223804e-04 2.464438e-04 4.768176e-04
 [6] 8.863697e-04 1.583090e-03 2.716594e-03 4.478906e-03 7.094919e-03
[11] 1.079819e-02 1.579003e-02 2.218417e-02 2.994549e-02 3.883721e-02
[16] 4.839414e-02 5.793831e-02 6.664492e-02 7.365403e-02 7.820854e-02
[21] 7.978846e-02 7.820854e-02 7.365403e-02 6.664492e-02 5.793831e-02
[26] 4.839414e-02 3.883721e-02 2.994549e-02 2.218417e-02 1.579003e-02
[31] 1.079819e-02 7.094919e-03 4.478906e-03 2.716594e-03 1.583090e-03
[36] 8.863697e-04 4.768176e-04 2.464438e-04 1.223804e-04 5.838939e-05
[41] 2.676605e-05
[1] 0.1586553
 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
 [1] 2.676605e-05 5.838939e-05 1.223804e-04 2.464438e-04 4.768176e-04
 [6] 8.863697e-04 1.583090e-03 2.716594e-03 4.478906e-03 7.094919e-03
[11] 1.079819e-02 1.579003e-02 2.218417e-02 2.994549e-02 3.883721e-02
```

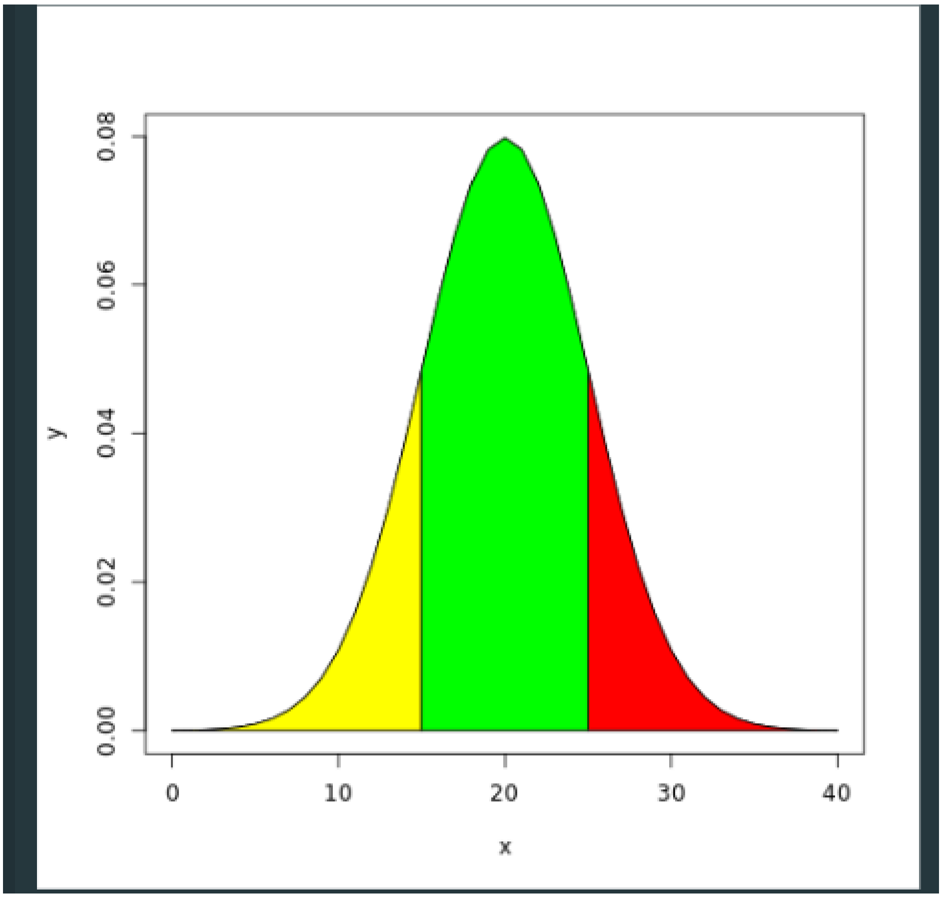
```
 [6] 8.863697e-04 1.583090e-03 2.716594e-03 4.478906e-03 7.094919e-03
[11] 1.079819e-02 1.579003e-02 2.218417e-02 2.994549e-02 3.883721e-02
[16] 4.839414e-02
[1] 0.1586236
 [1] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
 [1] 4.839414e-02 3.883721e-02 2.994549e-02 2.218417e-02 1.579003e-02
 [6] 1.079819e-02 7.094919e-03 4.478906e-03 2.716594e-03 1.583090e-03
[11] 8.863697e-04 4.768176e-04 2.464438e-04 1.223804e-04 5.838939e-05
[16] 2.676605e-05
[1] 0.6826895
 [1] 15 16 17 18 19 20 21 22 23 24 25
 [1] 0.04839414 0.05793831 0.06664492 0.07365403 0.07820854 0.07978846
 [7] 0.07820854 0.07365403 0.06664492 0.05793831 0.04839414
         p1        p2        p3
1 0.1586553 0.1586236 0.6826895

[Execution complete with exit code 0]
```

FIGURES



Poisson distribution
probability distribution of x
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
15
20
values of x



y
0.00
0.02
0.04
0.06
0.08
0
10
20
30
40
x

EXPERIMENT 7

#Aim - to understand the testing of hypothesis for large sample tests using R functions

#Suppose the mean weight of King Penguins found in an Antarctic colony last year was 15.4 KG. In a sample of 35 penguins same time this year in the same colony, the mean penguins

#weight is 14.6 kg. Assume the population standard deviation is 2.5 kg. At 0.05 significance level, can we reject the null hypothesis that the mean penguin weight does not differ from last yerar?

#xbar = 14.6

#mu = 15.4

#alpha = 0.05

#sigma = 2.5

#n = 35

#input the sample mean

xbar=14.6

xbar

#Input the population mean

mu0=15.4

mu0

#Input the standard deviation

sigma = 2.5

sigma

```
#Input the sample size
n = 35

#Test statistic
z = (xbar - mu0)/(sigma/sqrt(n)) z

#Level of significance
alpha = 0.05
alpha

#Two tailed test critical value
zhalfalpha = qnorm(1 -
(alpha/2)) zhalfalpha

c(-zhalfalpha, zhalfalpha)

#To find p-value
pval = 2*pnorm(z)
pval

#conclusion
if (pval > alpha) {
  print("Accept null
hypothesis") } else {


  print("Reject null hypothesis")
```

}

#Problem-2

#z = (p - P)/(root(PQ/n))

#The fatality rate of typhoid patients is believed to be 17.26%. In a certain year 640 patients siffereing from typhoid were treated in a metropolitan hospital and only 63 patients died.

#Can you consider the efficient? Level of significance is 5%

#Input the data

#Size of the sample

n = 640

n

#Sample proportion

Sprop = 63/n

Sprop

#Population proportion

Pprop=0.1726

Pprop

#Probability of failure

Q = 1 - Pprop

Q

```
#Test statistic
z = (Sprop-Pprop)/sqrt(Pprop*Q/n)
z

#critical value
E = qnorm(.975)
E

#Critical region
c(-E, E)

#Confidence interval
Sprop+c(-E,E)*sqrt(Pprop*(1-Pprop)/
n)

#Conclusion
if (z>-E && z < E) {
  print("Hospital is not efficient")
} else {

  print("Hospital is efficient")
}
```

OUTPUT

```
[1] 14.6
[1] 15.4
[1] 2.5
[1] -1.893146
[1] 0.05
[1] 1.959964
[1] -1.959964  1.959964
[1] 0.05833852
[1] "Accept null hypothesis"
```

```
[1] 640
[1] 0.0984375
[1] 0.1726
[1] 0.8274
[1] -4.964736
[1] 1.959964
[1] -1.959964  1.959964
[1] 0.06915985 0.12771515
[1] "Hospital is efficient"
```

EXPERIMENT8

#Testing of hypothesis

 # In a random sample of size 500, them mean is found to be 20. In another independent sample

# of size 400, the mean is 15. Could the samples have been drawn from the same population with SD 4?

#input the sample mean

xbar = 20

xbar

ybar=15

ybar

input the standard deviation

sigma = 4

input the sample size

n1 = 500

n1

```
n2 = 400

n2

# Test statistic

z = (xbar-ybar)/(sigma*sqrt((1/n1)+(1/n2)))

z

#Level of significance

alpha = 0.05

alpha

# Two tailed critical value

zalpha = qnorm(1 - (alpha/2))

zalpha



#conclusion

if (z <= zalpha)

{

  print("Accept Null Hypothesis")

}else

{

  print("Reject Null Hypothesis")

}

#---------------------------------------------------------
```

```
# Testing of hypothesis - Two sample proportion test

# z = (p1 - p2)/sqrt(PQ(1/n1 + 1/n2))

# In a large city A, 20% of a random sample of 900 school boys had a slight

# physical defect. In another large city B, 18.5% of a random sample of 1600 boys had the same defect.

# Is the difference between the proportions significant


# Input the sample proportions

p1 = 0.20

p1


p2 = 0.185

p2


# Input the sample sizes

n1 = 900

n1

n2 = 1600

n2


#Population proportion

P = (n1*p1 + n2*p2)/(n1+n2)

P

Q = 1 - P

Q


#Test statistic
```

z = (p1-p2)/sqrt(P*Q*((1/n1)+(1/

n2))) z

alpha=0.05

alpha

#Two-tailed critical value

zalpha = qnorm(1 - (alpha/2)) zalpha

conclusion

if (z<=zalpha) {print("Accept Null Hypothesis")} else {print("Reject Null Hypothesis")}

OUTPUT

```
Output
[1] 20
[1] 15
[1] 500
[1] 400
[1] 18.6339
[1] 0.05
[1] 1.959964
[1] "Reject Null Hypothesis"
[1] 0.2
[1] 0.185
[1] 900
[1] 1600
[1] 0.1904
[1] 0.8096
[1] 0.9169249
[1] 0.05
[1] 1.959964
[1] "Accept Null Hypothesis"

[Execution complete with exit code 0]
```

EXPERIMENT 9

#T-Test

 # t = (xbar - u)/(sigma/sqrt(n))

 # Two independent samples of sizes 8 and 7 are given contained the following values

 # Sample 1 - 19 17 15 21 16 18 16 14

 # Sample 2 - 15 14 15 19 15 18 16 20

 # Is the difference between the sample means significant

 # Problem 1

```
#input the data
sample1=c(19, 17, 15, 21, 16, 18, 16, 14)
sample1

sample2 = c(15, 14, 15, 19, 15, 18, 16, 20)
sample2

# output using t-distribution
t = t.test(sample1, sample2)
t

#test-statistic
cv = t$statistic
cv

#critical value
tv = qt(0.975, 14)
tv
```

```
#conclusion

if (cv <= tv) {print("Accept Ho")} else {print("Reject Ho")}


# Problem 2

# The following data relate to the marks obtained by 10 students in two test, one held at the beginning

# of a year and the other at the end of the year after intensive coaching. Do the data indicate that the

# students have got benefited by coaching

# Test-1 19 17 15 21 16 18 16 14 19 20

# Test-2 15 14 15 19 15 18 16 20 22 19


#input the data

test1 = c(19, 17, 15, 21, 16, 18, 16, 14, 19, 20)

test1


test2 = c(15, 14, 15, 19, 15, 18, 16, 20, 22, 19)

test2


# output using t-distribution

t2 = t.test(test1, test2)

t2


#test-statistic

cv2 = t2$statistic

cv2
```

```
#critical value
tv2 = qt(0.975, 18)
tv2



#conclusion
if (cv2 <= tv2) {print("Accept Ho")} else {print("Reject Ho")}

#-----------------------------------------------------
#F Test
# Problem: 3 (F-Test)
#Two independent samples of sizes 8 and 7 contained the following values:
# Sample 1: 19 17 15 21 16 18 16 14
# Sample 2: 15 14 15 19 15 18 16 20

sample1 = c(19, 17, 15, 21, 16, 18, 16, 14)
sample1

sample2 = c(15, 14, 15, 19, 15, 18, 16, 20)
sample2

#output using t-distribution
f = var.test(sample1, sample2)
f

#F Test to compare two variances
```

```
#     Test     statistic
cv    =    f$statistic
cv

tv = qf(0.95, 7, 7)
tv

#conclusion
if (cv <= tv) {print("Accept Ho")} else {print("Reject Ho")}
```

OUTPUT

Output

```
[1] 19 17 15 21 16 18 16 14
[1] 15 14 15 19 15 18 16 20

        Welch Two Sample t-test

data:  sample1 and sample2
t = 0.44721, df = 13.989, p-value = 0.6616
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.898128  2.898128
sample estimates:
mean of x mean of y
     17.0      16.5

        t
0.4472136
[1] 2.144787
[1] "Accept Ho"
 [1] 19 17 15 21 16 18 16 14 19 20
 [1] 15 14 15 19 15 18 16 20 22 19

        Welch Two Sample t-test

data:  test1 and test2
t = 0.18042, df = 17.555, p-value = 0.8589
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.133217  2.533217
sample estimates:
```

```
sample estimates:
mean of x mean of y
     17.5      17.3

        t
0.1804154
[1] 2.100922
[1] "Accept Ho"
[1] 19 17 15 21 16 18 16 14
[1] 15 14 15 19 15 18 16 20

        F test to compare two variances

data:  sample1 and sample2
F = 1.0588, num df = 7, denom df = 7, p-value = 0.9418
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2119805 5.2887274
sample estimates:
ratio of variances
          1.058824

       F
1.058824
[1] 3.787044
[1] "Accept Ho"
```

EXPERIMENT 10

 # Problem-1

#Five coins are tossed 256 times. The number of heads observed by binomial distribution is given by below.

 # Examine if the coins are unbiased by employing chi-square goodness of fit.

 # No of heads 0 1 2 3 4 5

 # Frequency 5 35 75 84 45 12

number of coins

n = 5

n

level of significance

alpha = 0.05

```
alpha

N = 256
N

P = 0.5
P

x = c(0:n);x

obf = c(5, 35, 75, 84, 45, 12)
obf

exf = (dbinom(x, n, P)*256)
exf

sum(obf)

sum(exf)

chisq <- sum((obf-exf)^2/exf)
chisq
cv = chisq;cv

#Critical value using chisq-distribution
tv = qchisq(1-alpha, n);tv
```

```
#Hypothesis conclusion
if (cv <= tv) {print("Accept H0/Fit is good")} else {print("Reject H0/Fit is not good.")}

# Problem 2
# From the following information state whether the condition of the child is associated with the condition of the house
# Condition of the child    Condition of the house clean    Condition of the house dirty
#      Clean                     69                          51
#      Fairly Clean                 81                          20
#      Dirty                    35                          44

# Input the data
data <- matrix(c(69, 51, 81, 20, 35, 44), ncol=2, byrow=T)
data

l = length(data);l

#Output by chisq distribution
cv = chisq.test(data)
cv

#p-value
cv = cv$p.value
cv
```

alpha = 0.05

#Hypothesis

conclusion

if (cv > alpha) {print("Attributes are independent")} else {print("Attributes are not independent")}

OUTPUT

```
Output

[1] 5
[1] 0.05
[1] 256
[1] 0.5
[1] 0 1 2 3 4 5
[1]  5 35 75 84 45 12
[1]  8 40 80 80 40  8
[1] 256
[1] 256
[1] 4.8875
[1] 4.8875
[1] 11.0705
[1] "Accept H0/Fit is good"
     [,1] [,2]
[1,]   69   51
[2,]   81   20
[3,]   35   44
[1] 6

	Pearson's Chi-squared test

data:  data
X-squared = 25.629, df = 2, p-value = 2.721e-06

[1] 2.72114e-06
[1] "Attributes are not independent"
```

EXPERIMENT 11

Completely randomized design

```
# Problem: A car rental agency, which uses 5 different brands of tyres in the process of deciding the brand of tyre

# to purchase as standard equipment for its fleet, finds that each of  5 tyres of each brand

# last the following number of kilometers

# A B C D E

# 36 46 35 45 41

# 37 39 42 36 39

# 42 35 37 39 37

# 38 37 43 35 35

# 47 43 38 32 38

# Test the hypothesis that the five brands have almost the same average life

# One way ANOVA

# Test of tyres

A = c(36, 37, 42, 38, 47)

B = c(46, 39, 35, 37,43)

C = c(35,42,37,43,38)

D = c(45, 36, 39, 35, 32)

E = c(41,39,37,35,38)

group <- data.frame(cbind(A,B,C,D,E))

group


summary(group)

#stack vector from data frame

stgr <- stack(group);stgr


#Completely randomized design
```

```
crd <- aov (values~ind, data=stgr)

#ANOVA table

summary(crd)

# THere is no difference in the average lofe of

tyres # Visualization of data


boxplot(group, ylab="Average life of tyres in kilometers", main="Brands of Tyres")


# Two way ANOVA

# Randomized block design

# THe follwing table gives monthly sales in thousand rupees of a certain firm in the 3 states by its four salesmen


# States      Salesmen

#          I   II   III   IV

#   A      6   5    3    8

#   B      8   9    6    5

#   C      10  7    8    7


# Setup the analysis of variance table and test whether there is any significant difference

#(i) between the salesmen

# (ii) between sales in the states


#Monthly sales of the states

StateA = c(6,5,3,8)

StateA
```

```
StateB = c(8, 9, 6, 5)
StateB

StateC = c(10, 7, 8, 7)
StateC

#Frame the data set
Group <- data.frame(cbind(StateA, StateB, StateC))
Group

Sales = c(t(as.matrix(Group))); Sales

f = c("State A", "State B", "State C")
f
g = c("Salesman 1", "Salesman 2", "Salesman 3", "Salesman 4")
g
# Number of columns
k = ncol(Group)
k
# Number of rows
n = nrow(Group)
n

#Generate factor levels of states
States = gl(k, 1, n*k, factor(f))
```

States

Generate factor levels of Salesmen

Salesmen = gl(n, k, n*k, factor(g))

Salesmen

#ANOVA Table

anova = aov(Sales~States +

Salesmen) summary(anova)

Reject the states

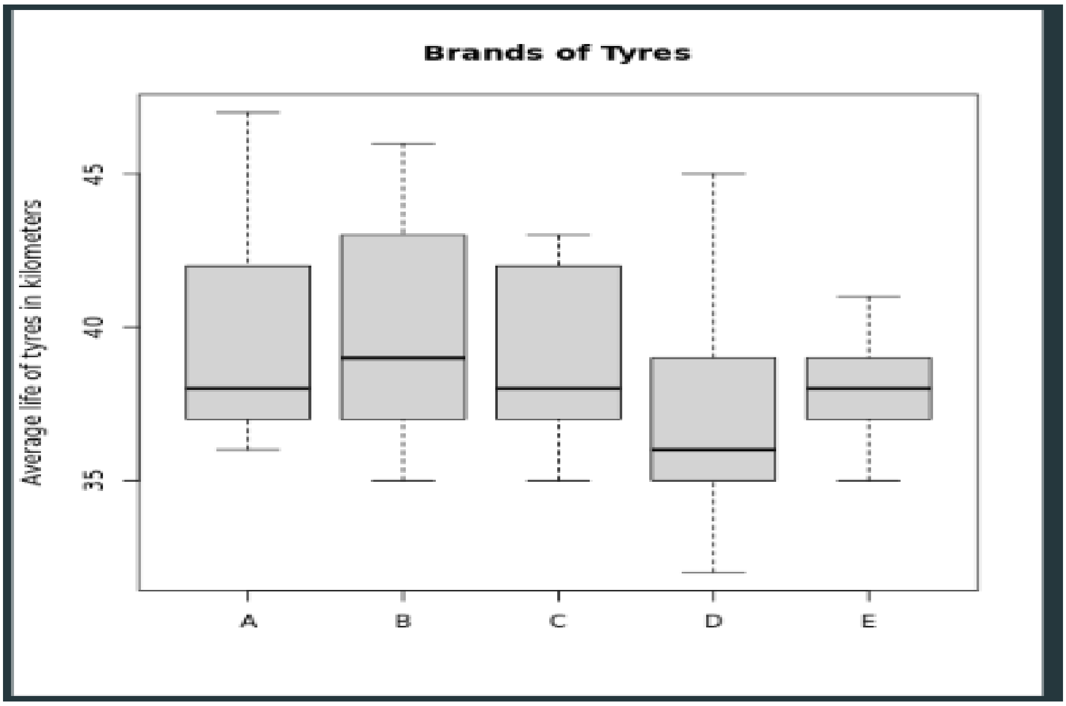
Reject the salesmen

OUTPUT

Output

```
   A  B  C  D  E
1 36 46 35 45 41
2 37 39 42 36 39
3 42 35 37 39 37
4 38 37 43 35 35
5 47 43 38 32 38
       A            B            C            D              E
 Min.   :36   Min.   :35   Min.   :35   Min.   :32.0   Min.   :35
 1st Qu.:37   1st Qu.:37   1st Qu.:37   1st Qu.:35.0   1st Qu.:37
 Median :38   Median :39   Median :38   Median :36.0   Median :38
 Mean   :40   Mean   :40   Mean   :39   Mean   :37.4   Mean   :38
 3rd Qu.:42   3rd Qu.:43   3rd Qu.:42   3rd Qu.:39.0   3rd Qu.:39
 Max.   :47   Max.   :46   Max.   :43   Max.   :45.0   Max.   :41
   values ind
1      36   A
2      37   A
3      42   A
4      38   A
5      47   A
6      46   B
7      39   B
8      35   B
9      37   B
10     43   B
11     35   C
12     42   C
13     37   C
14     43   C
15     38   C
16     45   D
17     36   D
18     39   D
19     35   D
20     32   D
21     41   E
22     39   E
23     37   E
24     35   E
25     38   E
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
ind          4   27.4    6.86   0.422  0.791
Residuals   20  325.2   16.26
[1] 6 5 3 8
[1] 8 9 6 5
[1] 10  7  8  7
  StateA StateB StateC
1      6      8     10
2      5      9      7
3      3      6      8
4      8      5      7
 [1]  6  8 10  5  9  7  3  6  8  8  5  7
[1] "State A" "State B" "State C"
[1] "Salesman 1" "Salesman 2" "Salesman 3" "Salesman 4"
[1] 3
[1] 4
 [1] State A State B State C State A State B State C State A State B State C
[10] State A State B State C
Levels: State A State B State C
 [1] Salesman 1 Salesman 1 Salesman 1 Salesman 2 Salesman 2 Salesman 2
 [7] Salesman 3 Salesman 3 Salesman 3 Salesman 4 Salesman 4 Salesman 4
Levels: Salesman 1 Salesman 2 Salesman 3 Salesman 4
            Df Sum Sq Mean Sq F value Pr(>F)
States       2 12.667   6.333   1.839  0.238
Salesmen     3  8.333   2.778   0.806  0.535
Residuals    6 20.667   3.444
```



```
#no. of trials
m=20
m

#probability of success
ps=0.02

#poisson parameter
lambda=m*ps
lambda

#atleast 2 defectives
p1=sum(dpois(2:m,lambda))
p1

#(i) no. of boxes containing atleast 2 defectives
p2=dpois(2,lambda)
p2

#(ii) no. of boxes containing exactly 2 defectives
round(1000*p2)

#atmost 2 defectives
p3=sum(dpois(0:2,lambda))
p3

#(iv)plot the distribution
x1=0:m
px1=dpois(x1,lambda)
plot(x1,px1,type="h",xlab="values of x",ylab="probability ofx",main="poisson
distribution")

#(v)E(x)
Ex1=weighted.mean(x1,px1)
Ex1

#(vi) variance of x
varx1=weighted.mean(x1*x1,px1)-(weighted.mean(x1,px1))^2
varx1

#SAMPLE PROBLEM:
#A company finds the time taken by one of its engineers to complete or repair
job has a normal distribution with mean 20 minutes and S.D 5 minutes. state
what proportion of job takes:
#(i) less than 15 minutes (ii) more than 25 minutes (iii) between 15 & 25
minutes (iv)plot the distribution (v) table the distribution
```

```r
#generating the data x
x= seq(0,40)
x

#find the density function of x
y=dnorm(x,mean=20,sd=5)
y

#plot the normal distribution curve
plot(x,y,type='l')

#preparation of jobs take less than 15 minutes
p1=pnorm(15,mean=20,sd=5)
p1

x2=seq(0,15)
x2

y2=dnorm(x2,mean=20,sd=5)
y2
polygon(c(0,x2,15),c(0,y2,0),col='yellow')

#preparation of jobs take more than 25 mins
p2=pnorm(40,mean=20,sd=5)-pnorm(25,mean=20,sd=5)
p2

x1=seq(25,40)
x1
y1=dnorm(x1,mean=20,sd=5)
y1
polygon(c(25,x1,40),c(0,y1,0),col='red')

#proportion of jobs taken b/w 15 and 25 minutes
p3=pnorm(25,mean=20,sd=5)-pnorm(15,mean=20,sd=5)
p3

x3=seq(15,25)
x3

y3=dnorm(x3,mean=20,sd=5)
y3

polygon(c(15,x3,25),c(0,y3,0),col='green')

#probability distribution
data.frame(p1,p2,p3)

#EXPERIMENT 7
```

```r
#Input the sample mean
xbar=14.6
xbar

#input the population mean

mu0=15.4
mu0

#input standard deviation
sigma=2.5
sigma

#input the sample size
n=35
n

#test statistics
z=(xbar-mu0)/(sigma/sqrt(n))
z

#level of significance
alpha=0.05
alpha

#two-tailed critical value
zhalfalpha=qnorm(1-(alpha/2))
zhalfalpha

c(-zhalfalpha,zhalfalpha)

#to find p-value
pval=2*pnorm(z)
pval

#conclusion
if(pval>alpha){print("accept null hypothesis")} else{print("reject null
hypothesis")}

#Testing Hypothesis -Large sample proportion test

#input the data
#size of the sample

n=640
n
```

```
#sample proportion
Sprop=63/n
Sprop

#population proportion
Pprop=0.1726
Pprop

#probability of failure
Q=1-Pprop
Q

#test statistics
z=(Sprop-Pprop)/sqrt(Pprop*Q/n)
z

E=qnorm(.975)

#critical region
c(-E,E)

#confidence interval
Sprop+c(-E,E)*sqrt(Pprop*(1-Pprop)/n)

#conclusion
if(z>-E&&z<E){print("hospital is not efficient")} else{print("hospital is
efficient")}

#EXPERIMENT 8
#QUESTION 1
P1=0.20
P1
p2=0.185
p2
n1=900
n1
n2=1600
n2
P=(n1*p1+n2*p2)/(n1+n2)
P
Q=1-P
Q
z=(p1-p2)/sqrt(P*Q*((1/n1)+(1/n2)))
z
alpha=0.05
alpha
#two tailed test value
zalpha=qnorm(1-(alpha/2))
```

```
zalpha
#conclusion:testing of hypothesis fpr large sample tests using R functions has been explored and concluded
if(z<=zalpha){print("accept null hypothesis")} else{print("reject null hypothesis")}

#QUESTION 2
sample1=c(19,17,15,21,16,18,16,14)
sample1
sample2=c(15,14,15,19,15,18,16,20)
sample2

#output using t-distribution
t=t.test(sample1,sample2)
t

#test-statistics
cv=t$statistic
cv

#critical value
tv=qt(0.975,14)
tv

#conclusion
if(cv<=tv){print("accept H0")} else{print("reject H0")}

#QUESTION 3

#variance test or F-test
sample1=c(19,17,15,21,16,18,16,14)
sample1
sample2=c(15,14,15,19,15,18,16,20)
sample2
#output using t- distribution
f=var.test(sample1,sample2)
f
#test statistics
tv=qf(0.95,7,7)
tv
#conclusion
if(cv<=tv){print("accept H0")} else{print("reject H0")}

#QUESTION 4
#Testing of hypothesis
# In a random sample of size 500, them mean is found to be 20. In another independent sample
```

```
# of size 400, the mean is 15. Could the samples have been drawn from the same
population with SD 4?

#input the sample mean
xbar = 20
xbar

ybar=15
ybar

# input the standard deviation
sigma = 4

# input the sample size
n1 = 500
n1
n2 = 400
n2

# Test statistic
z = (xbar-ybar)/(sigma*sqrt((1/n1)+(1/n2)))
z

#Level of significance
alpha = 0.05
alpha

# Two tailed critical value
zalpha = qnorm(1 - (alpha/2))
zalpha


#conclusion
if (z <= zalpha)
{
  print("Accept Null Hypothesis")
}else
{
  print("Reject Null Hypothesis")
}


#EXPERIMENT 9
#T-Test
# t = (xbar - u)/(sigma/sqrt(n))
# Two independent samples of sizes 8 and 7 are given contained the following
values
# Sample 1 - 19 17 15 21 16 18 16 14
# Sample 2 - 15 14 15 19 15 18 16 20
```

```r
# Is the difference between the sample means significant
# Problem 1
#input the data
sample1=c(19, 17, 15, 21, 16, 18, 16, 14)
sample1

sample2 = c(15, 14, 15, 19, 15, 18, 16, 20)
sample2

# output using t-distribution
t = t.test(sample1, sample2)
t

#test-statistic
cv = t$statistic
cv

#critical value
tv = qt(0.975, 14)
tv

#conclusion
if (cv <= tv) {print("Accept Ho")} else {print("Reject Ho")}

# Problem 2
# The following data relate to the marks obtained by 10 students in two test, one held at the beginning
# of a year and the other at the end of the year after intensive coaching. Do the data indicate that the
# students have got benefited by coaching
# Test-1 19 17 15 21 16 18 16 14 19 20
# Test-2 15 14 15 19 15 18 16 20 22 19

#input the data
test1 = c(19, 17, 15, 21, 16, 18, 16, 14, 19, 20)
test1

test2 = c(15, 14, 15, 19, 15, 18, 16, 20, 22, 19)
test2

# output using t-distribution
t2 = t.test(test1, test2)
t2

#test-statistic
cv2 = t2$statistic
cv2
```

```
#critical value
tv2 = qt(0.975, 18)
tv2

#conclusion
if (cv2 <= tv2) {print("Accept Ho")} else {print("Reject Ho")}

#------------------------------------------------------
#F Test
# Problem: 3 (F-Test)
#Two independent samples of sizes 8 and 7 contained the following values:
# Sample 1: 19 17 15 21 16 18 16 14
# Sample 2: 15 14 15 19 15 18 16 20

sample1 = c(19, 17, 15, 21, 16, 18, 16, 14)
sample1

sample2 = c(15, 14, 15, 19, 15, 18, 16, 20)
sample2

#output using t-distribution
f = var.test(sample1, sample2)
f

#F Test to compare two variances
# Test statistic
cv = f$statistic
cv

tv = qf(0.95, 7, 7)
tv

#conclusion
if (cv <= tv) {print("Accept Ho")} else {print("Reject Ho")}

#EXPERIMENT 10
# Problem-1
#Five coins are tossed 256 times. The number of heads observed by binomial
distribution is given by below.
# Examine if the coins are unbiased by employing chi-square goodness of fit.
# No of heads 0 1 2 3 4 5
# Frequency 5 35 75 84 45 12

# number of coins
n = 5
n

# level of significance
```

```r
alpha = 0.05
alpha

N = 256
N

P = 0.5
P

x = c(0:n);x

obf = c(5, 35, 75, 84, 45, 12)
obf

exf = (dbinom(x, n, P)*256)
exf

sum(obf)

sum(exf)

chisq <- sum((obf-exf)^2/exf)
chisq
cv = chisq;cv

#Critical value using chisq-distribution
tv = qchisq(1-alpha, n);tv

#Hypothesis conclusion
if (cv <= tv) {print("Accept H0/Fit is good")} else {print("Reject H0/Fit is
not good.")}

# Problem 2
# From the following information state whether the condition of the child is
associated with the condition of the house
# Condition of the child    Condition of the house clean    Condition of the
house dirty
#       Clean                           69                               51
#       Fairly Clean                    81                               20
#       Dirty                           35                               44

# Input the data
data <- matrix(c(69, 51, 81, 20, 35, 44), ncol=2, byrow=T)
data

l = length(data);l

#Output by chisq distribution
```

```r
cv = chisq.test(data)
cv

#p-value
cv = cv$p.value
cv
alpha = 0.05
#Hypothesis conclusion
if (cv > alpha) {print("Attributes are independent")} else {print("Attributes
are not independent")}


#EXPERIMENT 11

# Completely random mixed design

# Problem: A car rental agency, which uses 5 different brands of tyres in the
process of deciding the brand of tyre
# to purchase as standard equipment for its fleet, finds that each of  5 tyres
of each brand
# last the following number of kilometers
# A B C D E
# 36 46 35 45 41
# 37 39 42 36 39
# 42 35 37 39 37
# 38 37 43 35 35
# 47 43 38 32 38
# Test the hypothesis that the five brands have almost the same average life
# One way ANOVA
# Test of tyres
A = c(36, 37, 42, 38, 47)
B = c(46, 39, 35, 37,43)
C = c(35,42,37,43,38)
D = c(45, 36, 39, 35, 32)
E = c(41,39,37,35,38)
group <- data.frame(cbind(A,B,C,D,E))
group

summary(group)
#stack vector from data frame
stgr <- stack(group);stgr

#Completely randomized design
crd <- aov (values~ind, data=stgr)
#ANOVA table
summary(crd)
# THere is no difference in the average lofe of tyres
# Visualization of data
```

```r
boxplot(group, ylab="Average life of tyres in kilometers", main="Brands of Tyres")

# Two way ANOVA
# Randomized block design
# THe follwing table gives monthly sales in thousand rupees of a certain firm in the 3 states by its four salesmen

# States        Salesmen
#           I   II    III   IV
#   A       6   5      3    8
#   B       8   9      6    5
#   C       10  7      8    7

# Setup the analysis of variance table and test whether there is any significant difference
#(i) between the salesmen
# (ii) between sales in the states

#Monthly sales of the states
StateA = c(6,5,3,8)
StateA

StateB = c(8, 9, 6, 5)
StateB

StateC = c(10, 7, 8, 7)
StateC

#Frame the data set
Group <- data.frame(cbind(StateA, StateB, StateC))
Group

Sales = c(t(as.matrix(Group))); Sales

f = c("State A", "State B", "State C")
f
g = c("Salesman 1", "Salesman 2", "Salesman 3", "Salesman 4")
g
# Number of columns
k = ncol(Group)
k
# Number of rows
n = nrow(Group)
n

#Generate factor levels of states
States = gl(k, 1, n*k, factor(f))
```

```r
States

# Generate factor levels of Salesmen
Salesmen = gl(n, k, n*k, factor(g))
Salesmen

#ANOVA Table
anova = aov(Sales~States + Salesmen)
summary(anova)
# Reject the states
# Reject the salesmen
```
