R Textbook Companion for Probability and Statistics for Engineers by Richard L. Scheaffer, Madhuri S. Mulekar, James T. McClave[1]

Created by
Shikha Vyas
B.E.
Information Technology
Institute of Engineering and Technology, DAVV, Indore
Cross-Checked by
R TBC Team

May 26, 2020

[1]Funded by a grant from the National Mission on Education through ICT - http://spoken-tutorial.org/NMEICT-Intro. This Textbook Companion and R codes written in it can be downloaded from the "Textbook Companion Project" section at the website - https://r.fossee.in.

Book Description

Title: Probability and Statistics for Engineers

Author: Richard L. Scheaffer, Madhuri S. Mulekar, James T. McClave

Publisher: Cengage Learning, USA

Edition: 5

Year: 2011

ISBN: 978-0-534-40302-7

R numbering policy used in this document and the relation to the above book.

Exa Example (Solved example)

Eqn Equation (Particular equation of the above book)

For example, Exa 3.51 means solved example 3.51 of this book. Sec 2.3 means an R code whose theory is explained in Section 2.3 of the book.

Contents

List of R Codes

Chapter 1

Data Collection and Exploring Univariate Distributions

R code Exa 1.1 Auto paint shop relative freq

```
complain = c(32,17,5,4,3,3,1,1)
n= sum(complain)
rf=complain/n
cat("Relative frequency :", rf)

a =function(x){
sum(  rf[1:x])
}

cat("CRF:")
for (i in 1:8) {
cat( a(i),"   ")
}
```

R code Exa 1.2 Air quality relative frequency

```
LA_days= c(155,138,36,5)
O_days= c(233,39,1,1)

n1=sum(LA_days)
rf1 =LA_days/n1
n2=sum(O_days)
rf2=O_days/n2
cat("Relative frequency for LA:", rf1)
cat("Relative frequency for Oriando :",rf2)


a =function(x){
  sum(  rf1[1:x])
}
cat("Cumilative frequency for LA: ")
for (i in 1:4) {
  cat( a(i),"   ")
}

b =function(x){
  sum( rf2[1:x])
}
cat("Cumilative frequency for Ontario: " )
for (i in 1:4) {
  cat( b(i),"   ")
}
```

R code Exa 1.3 histograms

```
source =c("Fuel","Industrial","Transport","Misc")
CO1990 =c(5.510,5.582,76.635,11.122)
CO2000 =c(4.500,7.521,76.383,20.806)
data1 <- data.frame(CO1990, source)
data2  <- data1[order(data1[,1],decreasing=TRUE),]
barplot(data2[,1],names.arg=data2[,2],ylim = c
```

```
   (0,100),space = 0)

data3 <- data.frame(CO2000,source)
data4  <- data3[order(data3[,1],decreasing=TRUE),]
 barplot(data4[,1],names.arg=data4[,2],ylim = c
    (0,100),space = 0)

VOC1990 =c(1.005,10.000,8.988,1.059)
VOC2000 =c(1.206,8.033,8.396,2.710)

data5 <- data.frame(VOC1990, source)
data6  <- data5[order(data5[,1],decreasing=TRUE),]
barplot(data6[,1],names.arg=data6[,2],ylim = c
   (0,100),space = 0)

data7 <- data.frame(VOC2000,source)
data8  <- data7[order(data7[,1],decreasing=TRUE),]
barplot(data8[,1],names.arg=data8[,2],ylim = c
   (0,100),space = 0)

#Graph for VOC compounds plotted in book is
   incorrect
```

R code Exa 1.4 Dotplot

```
#Number of AQI exceedences for 1990, 1998, and 2006

year1 =c(42,0,5,0,9,11,51,2,161,15,39,18,0,2,14)
year2 =c(50,0,10,24,7,17,38,14,49,14,37,39,0,3,44)
year3 =c(18,1,5,13,6,6,18,11,34,11,18,36,2,5,18)


dat=data.frame(year1,year2,year3)
library(ggplot2)

```

```r
plot1= ggplot(dat,aes(x=year1)) + geom_dotplot(dotsize = 0.75,binwidth = 3.5) +xlab("1990")
plot2= ggplot(dat,aes(x=year2)) + geom_dotplot(dotsize = 0.75,binwidth = 3.5) + xlim(0,150) + xlab("1998")
plot3= ggplot(dat,aes(x=year3)) + geom_dotplot(dotsize = 0.75,binwidth = 4.0) + xlim(0,150) + xlab("2006")

library(grid)
grid.newpage()
grid.draw(rbind(ggplotGrob(plot1), ggplotGrob(plot2),ggplotGrob(plot3), size = "last"))

#Number of AQI exceedences by city


boston =c(0,0,0,0,0,0,0,0,0,4,0,3,9,8,1,4,1)
houston =c(51,36,32,27,38,65,26,46,38,51,42,28,21,31,22,28,18)
Newyork =c(15,30,4,11,13,17,11,22,14,22,19,19,27,11,6,15,11)

dat1=data.frame(boston,houston,Newyork)


plot4= ggplot(dat1,aes(x=boston)) + geom_dotplot(dotsize = 0.75,binwidth = 1.0)  +xlab("Boston") + xlim(0,70)
plot5=  ggplot(dat1,aes(x=houston)) + geom_dotplot(dotsize = 0.5,binwidth = 1.5) +xlab("Houston") + xlim(0,70)
plot6= ggplot(dat1,aes(x=Newyork)) + geom_dotplot(dotsize = 0.5,binwidth = 1.5) +xlab("New York") + xlim(0,70)

```

```
grid.newpage()
grid.draw(rbind(ggplotGrob(plot4), ggplotGrob(plot5),ggplotGrob(plot6), size = "last"))
```

R code Exa 1.8 AQI mean and median

```
data =c(12, 8, 10, 5, 17, 19, 31, 11, 88, 11, 19, 37, 1, 2, 12 )
cat("Mean =",mean(data))
cat("Median =",median(data))
```

R code Exa 1.9 accidents mean and median

```
n=c(3133,2065)
x=c(13.67,8.97)
cat(" the mean transportation time in Alabama is ", weighted.mean(x,n),"min")
```

R code Exa 1.10 AQI quantile

```
#The summary of AQI data for year 2003 are as follows

data=c(1, 2, 5, 8, 10, 11, 11, 12, 12, 17, 19, 19, 31, 37, 88)
quantile(data)
```

R code Exa 1.12 AQI sd and variance

```
data=c(1, 2, 5, 8, 10, 11, 11, 12, 12, 17, 19, 19,
   31, 37, 88)
cat("Mean :", mean(data))
cat("Variance :", var(data))
cat("Standard deviation :", sd(data))
```

R code Exa 1.13 histogram of cars

```
domestic =c(1995 , 2001 , 2004,  2000 , 2002 , 2002
   , 2000 , 2001,  1999 , 2002 , 2004,  2001,  2002
   ,1996,
             1990,  1995 , 1992  ,1995 , 1999 , 1996
                , 1999 , 1999  ,1998 , 2001,  2002 ,
                2004 ,2004 , 2004,
             2001,  1997 , 2002 , 2001 , 2002  ,2001,
                  2000 , 2002 , 1999 , 2001,  2002
                ,2000 , 2003,  2001  ,
             2001 , 1999 , 2002 , 2001 , 2002 , 2001
                , 2000 , 2002  ,2001 , 2002, 2000 ,
                2000 , 2002 , 2001 ,
             2002 , 2002 , 2001 , 2001  ,2002  ,2002
                , 2003 , 2003 , 2002, 2001 , 2002 ,
                2001,  2002 , 2003 ,
           2002
)

summary(domestic)

foriegn =c(1997 ,2000,  2002  ,2002  ,2001 , 2003 ,
   1995 , 1990 , 1992  ,1991 , 1997 , 2000 , 2000,
   1998 ,
            2000 , 1998,  2001 , 2004 , 2001 , 2000,
                2001 , 2000 , 2002 , 2003  ,2003 ,
```

```
                 2002 ,2002 , 2002 ,
              2003 , 1999 , 2000 , 2001 , 2003,  2003 , 2000 , 2001)
summary(foriegn)

hist(domestic,xlab = "Year of car advertised",main = "Domestic",xlim = c(1988,2006),breaks = 16)
hist(foriegn,xlab = "Year of car advertised",main = "Foreign",xlim = c(1988,2006),breaks = 16)
```

R code Exa 1.14 lawnmower mean and sd

```
prev_mean = 500
prev_Sd = 125

#a
# Increase the price of each lawnmower by $50.00.
cat("Changed mean= ",prev_mean+50)
cat("SD remains unchanged")

#b
#  Increase the price of each lawnmower by 10%.
cat("Changed mean= ",(1.1)*prev_mean)
cat("Changed sd= ",prev_Sd*(1.1))
```

R code Exa 1.16 summary of CO and SO2

```
# Mean,sd and no. of observations for CO and SO2 are as follows:
CO =c(  37298.013 , 84369.21 ,3143)
SO2 =c( 5616.0483 , 18869.243 ,3143)

z_CO = (189966.99-CO[1]) /CO[2]
```

```
z_SO =  (91310.67-SO2[1]) /SO2[2]

cat("The z-score when  carbon monoxide emission
   estimates at 189,966.99 :",z_CO)
cat("The z-score when  sulfur dioxide emission
   estimates at 91,310.67 :",z_SO)
```

R code Exa 1.17 Boxplots

```
#Boxplots

boston =c(0,0,0,0,0,0,0,0,0,4,0,3,9,8,1,4,1)
houston =c
   (51,36,32,27,38,65,26,46,38,51,42,28,21,31,22,28,18)

Newyork =c
   (15,30,4,11,13,17,11,22,14,22,19,19,27,11,6,15,11)


boxplot(boston,houston,Newyork,names = c("Boston","
   Houston","New YOrk"))
```

Chapter 2

Exploring Bivariate Distributions and Estimating Relations

R code Exa 2.1 Projections on net new workers

```
White =c(23,24)
Black =c(9,6)
Asian =c(7,6)
Hispanic =c(13,12)
gender =c(2,1)
df =data.frame(White,Black,Asian,Hispanic,gender)
means<-aggregate(df,by=list(df$gender),mean)
means<-means[,2:length(means)]
library(reshape2)
means.long<-melt(means,id.vars="gender")
library(ggplot2)
ggplot(means.long,aes(x=variable,y=value,fill=factor
    (gender)))+
  geom_bar(stat="identity",position="dodge")+
  scale_fill_discrete(name="Gender",
                      breaks=c(1,2),
                      labels=c("Men", "Women"))+
```

```
xlab("Category")+ylab("Mean Percentage")
```

R code Exa 2.3 Timeplot of CO and VOC emission

```
year =c
    (1989,1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000)
CO =c
    (106.439,99.119,101.797,99.007,99.791,103.713,94.058,104.600,105.
VOC =c
    (22.513,21.053,21.249,20.862,21.099,21.683,20.918,19.924,20.325,1
plot(year,CO)
lines(year,CO)
plot(year,VOC)
lines(year,VOC)
```

R code Exa 2.4 Average number of days with AQI greater than 100

```
x =c
    (24.60,28.00,19.93,22.40,20.93,25.73,14.46,18.13,23.06,26.53,20.5
year =c
    (1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002
plot(year,x)
lines(year,x)
```

R code Exa 2.5 Fuel consumption and efficiency of cars and vans

```
year=c(1970,1975,1980:1999)
Car =c
   (67.8,74.3,70.2,69.3,69.3,70.5,70.8,71.7,73.4,73.5,73.5,74.1,69.8

Van =c
   (12.3,19.1,23.8,23.7,22.7,23.9,25.6,27.4,29.1,30.6,32.7,33.3,35.6


#Graph for fuel consumption
plot(year,Car,ylim = c(10,80),ylab = "Fuel
   Consumption",xlab = "Year" )
lines(year,Car)
par(new =TRUE)
plot(year,Van,ylim = c(10,80),ylab = "Fuel
   Consumption",xlab = "Year")
lines(year,Van)

Car1 =c
   (13.5,14.0,16.0,16.5,16.9,17.1,17.4,17.5,17.4,18.0,18.8,18.0,20.3

Van1 =c
   (10.0,10.5,12.2,12.5,13.5,13.7,14.0,14.3,14.6,14.9,15.4,16.1,16.1


#Graph for fuel efficiency
plot(year,Car1,ylim = c(10,30),ylab = "Miles per
   gallon",xlab = "Year" )
lines(year,Car1)
par(new =TRUE)
plot(year,Van1,ylim = c(10,30),ylab = "Miles per
   gallon",xlab = "Year")
lines(year,Van1)
```

R code Exa 2.6 Annual temperatures at Newnan

```
temp =c(59.64 , 61.98 , 60.78 , 61.61 , 61.57 ,
    63.02 , 63.20 , 63.49 , 62.81 , 62.22 , 65.00 ,
    61.57 , 63.33 , 62.78 , 63.43 , 64.10  ,62.39 ,
    63.55 , 63.87 ,
          61.91 , 65.24 , 64.27 , 62.07 , 60.89 ,
             64.03 , 62.79 , 65.14 , 62.39 , 63.03 ,
             62.76 , 65.06 , 64.20 , 65.10 , 62.51 ,
             62.89  ,62.74 , 61.40 , 62.43 ,
          62.38 , 59.32 , 61.92 , 60.52 , 61.63 ,
             63.09 , 62.39 , 63.32 , 61.19 , 62.85 ,
             62.90 , 61.65 , 62.04 , 62.71 , 62.53 ,
             62.92  ,62.32 , 62.47 , 62.64 ,
          60.39 , 61.92 , 60.72 , 61.12 , 62.23 ,
             60.58 , 61.10 , 61.09 , 59.78 , 60.53 ,
             60.00 , 60.25 , 61.69 , 61.29 , 61.47 ,
             61.48  ,61.13 , 61.41 , 59.57 ,
          60.70 , 60.49 , 60.15 , 61.08 , 60.38 ,
             60.38 , 58.65 , 60.22 , 60.71 , 61.92 ,
             60.60 , 60.05 , 60.17 , 62.93 , 61.90 ,
             60.13  ,60.41 , 60.88 , 61.16 ,
          60.59 , 59.98 , 61.48 , 61.34 , 59.06
)
year=c(1901:2000)
plot(year,temp)
lines(year,temp)
```

R code Exa 2.7 scatterplot of heat exchangers

```
OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50

```

```
plot(OT,WR)
```

R code Exa 2.8 correlation coeff for power plant

```
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

r= cor(x,y)
cat("The fact that the value of r i.e,",r,"is
    positive and near 1 indicates that the peak power
     load is very strongly associated with the daily
    maximum temperature")
```

R code Exa 2.9 correlation coeff for O3 and SO2

```
# Mean,sd and no. of observations for CO and SO2 are
    as follows:
O3 =c
   (0.023,0.029,0.032,0.014,0.016,0.024,0.021,0.017,0.044,0.038,0.028

SO2 =c
   (0.019,0.030,0.075,0.047,0.009,0.043,0.031,0.039,0.010,0.046,0.027


r=cor(O3,SO2)
cat("The value of r is very small and doesn't shw
    any association between the O3 and SO2 levels
    measured.")
```

R code Exa 2.12 equation of the least squares regression line

```
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- lm(y~x)
coef = coefficients(pol)

#a
cat("beta1 : ",coef[2])
cat("y intercept i.e beta0: ",coef[1])

#b
cat("Peak power load increased by :", 5*coef[2],"
    megawatts when the maximum temperature increases
    by 5 F. ")

#c
cat("y = ",coef[1]," + ",coef[2],"x")

# Straight-line fit to power load and temperature
    data
abline(pol)
```

R code Exa 2.13 linear fit OT and WR

```
OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
pol <- lm(WR~OT)
summary(pol)
coef=coefficients(pol)
cat("WR = ",coef[1]," + ",coef[2],"OT")
```

R code Exa 2.14 Predict peak power load

```
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- data.frame(x,y)

line <- lm(y~x,data = pol)

#a
# here x0 <- 95 F
data_a <- data.frame(x=95)
res_a <- predict(line,data_a)
cat(" the fitted relation tells us that the likely
    peak load will be around ",res_a," megawatts. ")

#b
# here x0 <- 98 F
data_b <- data.frame(x=98)
res_b <- predict(line,data_b)
cat(" It predicts the peak power load of ",res_b, "
    megawatts for the day with maximum temperature 98
     F. ")

#c
# here x= <- 102 F
data_c <- data.frame(x=102)
res_c <-predict(line,data_c)
cat(" It predicts the peak power load of ",res_c, "
    megawatts for the day with maximum temperature
    102 F.")
```

R code Exa 2.15 Predict wall reduction

```r
OT =c(-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c(1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
pol <- lm(WR~OT)
data <- data.frame(OT=0.010)
res <- predict(pol,data)
cat("A wall reduction of",res,"is predicted for OT = 0.010. ")
```

R code Exa 2.16 coeff of determination for the peak power load

```r
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- data.frame(x,y)

line <- lm(y~x,data = pol)
r_sq <- summary(line)$r.squared
cat(" the sample variability of the peak load about their mean is reduced by ",r_sq*100," when the mean peak loads
   is modeled as a linear function of daily high temperature")
```

R code Exa 2.18 Least squares fit and residual plot for WR OT data

```
OT =c
   (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c
   (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50

#Least square fit

fit =lm(WR~OT)
summary(fit)
plot(WR~OT)
plot(OT,WR)
abline(reg = fit)

#Residual Plot
plot(OT,resid(fit))
abline(h=0)
```

R code Exa 2.19 linear regression model for population density

```
year=c
   (1790,1800,1810,1820,1830,1840,1850,1860,1870,1880,1890,1900,1910
pop =c
   (4.5,6.1,4.3,5.5,7.4,9.8,7.9,10.6,10.9,14.2,17.8,21.5,26.0,29.9,3

pol <- data.frame(x,y)

line <- lm(pop~year,data = pol)
summary(line)
coef<- coefficients(line)

#Regression Analysis: PopDens versus Year
```

```
#The regression equation is :

cat("Pop_Dens = ",coef[1]," + ",coef[2],"year")


line2 <- lm(log(pop)~year,data = pol)
summary(line2)
coef2<- coefficients(line2)

#Regression Analysis: ln(PopDens) versus Year
#The regression equation is :

cat("log(y) = ",coef2[1]," + ",coef2[2],"year")
```

R code Exa 2.20 linear regression modal for length and wt

```
wt =c
   (130,51,640,28,80,110,33,90,36,38,366,84,80,83,70,61,54,44,106,84
len =c
   (94,74,147,58,86,94,63,86,69,72,128,85,82,86,88,72,74,61,90,89,68

pol <- data.frame(len,wt)

#Regression Analysis: Weight versus Length
#The regression equation is :

line1 <- lm(wt~len,data = pol)
summary(line1)
coef1<- coefficients(line1)

cat("Weight = ",coef1[1] ," + ",coef1[2],"length")

# Output for the linear growth model
```

```
plot(len,wt)

#Residual plot
plot(len,resid(line1))


#Regression Analysis: ln(Weight) versus Length
#The regression equation is :

line2 <- lm(log(wt)~len,data = pol)
summary(line2)
coef2<- coefficients(line2)

cat("log(Weight) = ",coef2[1] ," + ",coef2[2]," length")

#Fit of ln(weight) versus length of alligators
plot(len,log(wt))

#Residual plot
plot(len,resid(line2))


#Regression Analysis: ln(Weight) versus ln(Length )
#The regression equation is :

line3 <- lm(log(wt)~log(len),data = pol)
summary(line3)
coef3<- coefficients(line3)

cat("log(Weight) = ",coef3[1] ," + ",coef3[2],"log(length)")

#Fit of ln(weight) versus ln(length) for alligators
plot(log(len),log(wt))

#Residual plot
plot(log(len),resid(line3))
```

Chapter 4

Probability

R code Exa 4.1 prob of eastern and western cities

```
prob_east <- 4
prob_west <- 2
prob_east_and_west <- prob_east * prob_west
prob_east_and_west
```

R code Exa 4.2 die example

```
dice_outcomes = c(1,2,3,4,5,6)
A= c(2,4,6)
B= c(1,3,5)
C= c(5,6)
prob_even = length(A)/length(dice_outcomes)
prob_even
prob_odd = length(B)/length(dice_outcomes)
prob_odd
prob_greater_than_4 = length(C)/length(dice_outcomes)
prob_greater_than_4
```

R code Exa 4.5 Venn diag EE

```
total_students = 100
calculus = 30
signal_processing = 25
calculus_and_signal_processing = 10
not_calculus = total_students-calculus
not_calculus
calulus_or_signal_processing = calculus + signal_
    processing - calculus_and_signal_processing
calulus_or_signal_processing
not_calculus_and_not_sigal_processing = total_
    students - calulus_or_signal_processing
not_calculus_and_not_sigal_processing
```

R code Exa 4.6 Venn diag labor statistics

```
all_woman = 52
all_white = 37
white_woman = 23
woman_or_white = all_woman + all_white - white_woman
woman_or_white
```

R code Exa 4.7 Venn diag electric motors

```
total_elecric_motors = 20
defect_free = 11
finish_defect = 8
assembly_defect = 3
```

```
defected_motors = total_elecric_motors - defect_free
finish_defect_and_assembly_defect = finish_defect +
   assembly_defect - defected_motors
finish_defect_and_assembly_defect
finish_defect_or_assembly_defect = defected_motors
finish_defect_or_assembly_defect
only_one_defect = finish_defect_or_assembly_defect -
    finish_defect_and_assembly_defect
only_one_defect
```

R code Exa 4.12 product rule example

```
ways_to_select_I_item <- 2
ways_to_select_II_item <- 2
ways_to_select_both_items <- ways_to_select_I_item *
    ways_to_select_II_item
ways_to_select_both_items
```

R code Exa 4.13 Product rule for cities

```
prob_E_selected_in_west <- 0.5
prob_E_gets_selected <- prob_E_selected_in_west
prob_E_gets_selected
```

R code Exa 4.16 Permutation of employees

```
cat("No. of possible assignments ", factorial(10)/
   factorial(7))
```

R code Exa 4.17 Permutation of operations

```
cat("No. of orderings", factorial(4))
```

R code Exa 4.18 Permutation of divisions

```
cat("prob D2 hiighest if all have equal preferences
   is ", factorial(3)/factorial(4))
cat("prob D2 is I and D3 II is", factorial(2)/
   factorial(4))
```

R code Exa 4.19.a selection of employees

```
cat("No. of possible selections ", choose(10,3))
```

R code Exa 4.19.b Probability of female candidate

```
cat("No. of ways to select 1 female ", choose(2,1))
a <- choose(2,1)
cat("No. of ways to select other 2 employees ",
   choose(8,2))
b <- choose(8,2)
cat("total ways of selecting 3 employees",choose
   (10,3))
c <-  choose(10,3)
d <- (a*b)/c
cat("prob is", d)
```

R code Exa 4.20 Combination of applicants

```
cat("No. of ways to select 2 applicants", choose
   (5,2))
a <- choose(5,2)
cat("No. of ways to select one of the best ", choose
   (2,1))
b <- choose(2,1)
cat("No. of ways to select other employee", choose
   (3,2))
c <-choose(3,2)
cat("prob is", (b*c)/a)
```

R code Exa 4.21.a Partitioning of employees

```
cat("No. of ways of assigning job", factorial(10)/(
   factorial(3)*factorial(4)*factorial(3)))
```

R code Exa 4.21.b Partitioning of employees specifically

```
cat("No. of ways three employees of a certain ethnic
    group get assigned to job I", factorial(7)/(
   factorial(4)*factorial(3)))
a <-  factorial(7)/(factorial(4)*factorial(3))
cat("No. of ways of assigning job", factorial(10)/(
   factorial(3)*factorial(4)*factorial(3)))
b <- factorial(10)/(factorial(3)*factorial(4)*
   factorial(3))
cat("prob is", a/b)
```

R code Exa 4.25 conditional probability of motors

```
cat("probability to select 2 non-defective motors ",
    4*3/20)
a <- 12/20
cat("probability to select 2 motors in which I is
    non defective ", 4*4/20)
b <- 16/20
cat("probability II motor is non defective if I one
    is non-defective ", a/b)
```

R code Exa 4.28 independent event of job

```
cat("Prob worker 1 or 2 is selected ", (1/4)+(1/4))
A <- (1/4)+(1/4)
cat("Prob worker 1 or 3 is selected" , (1/4)+(1/4))
B <- (1/4)+(1/4)
cat("Prob worker 1 is selected ",(1/4))
C <- 1/4
cat("Prob(AB) is worker 1 is selected ", 1/4)
AB <- 1/4
 A*B
cat("Since A*B is equal to AB, A and B are
    independent")
cat("Prob(AC) is worker 1 is selected", 1/4)
 A*C
cat("Since A*C is not equal to AC,  A and C are not
    independent")
```

R code Exa 4.29 Complementary Events

```
cat("ways in which no line chosen more than once ",
    factorial(10)/factorial(5))
a <- factorial(10)/factorial(5)
cat("total no. of ways to select a line ", 10 ^ 5)
b <- 10^5
cat("prob no line chosen more than once " ,a/b)
c <- a/b
cat("prob atleast one line chosen more than once " ,
     1-c)
```

R code Exa 4.31 additive rule

```
cat("Prob students enrolled in calculus class", 50/
    100)
a<- 50/100
cat("Prob students enrolled in class" , 45/100)
b <- 45/100
cat("prob students enrolled in both courses" ,10/
    100)
c <-10/100
cat("prob students enrolled in atleat one course", a
    +b-c)
cat("prob of students enrolled in none of the
    courses", 1-(a+b-c))
```

R code Exa 4.32 multiplicative rule for defects

```
defect_free_prob <-0.75
cat("prob of defected item ", 1- defect_free_prob)
defected <- 1- defect_free_prob
shaft <- 0.20
```

```
bushing <- 0.10
both_defect_prob <- shaft + bushing -defected
cat("prob of only shaft defect is ", shaft - both_defect_prob)
```

R code Exa 4.33 multiplicative rule for relays

```
cat("prob both relays r1 nad r2  open ", 0.2 * 0.2)
E1 <- 0.2 * 0.2
cat("Prob r1 open r2 closed" , 0.2*0.8)
E2 <- 0.2*0.8
cat("prob r1 closed r2 open ",0.8*0.2)
E3<-0.8*0.2
cat("Prob r1 r2 both closed ", 0.8*0.8)
E4 <-0.8*0.8
cat("Prob current will flow", E2+E3+E4)
```

R code Exa 4.35 example of ballpoint pen

```
A1<- 80
A2<- 120
DTA1<-8
DTA2<-2
DFA1<-13
DFA2<-27
NDA1<-59
NDA2<-91
TOTAL<-200
cat("overall defective trash rate",( DTA1+DTA2)/TOTAL)
cat("prob that pen is defective and produced by assembly line1 ", DTA1/TOTAL)
```

```
cat("Prob pen is defective if produced by line1 ", (DTA1/TOTAL)/(A1/TOTAL))
```

R code Exa 4.37 Bayes Rule

```
supplier1<-0.40
supplier2<-0.60
defective_supplier1<-0.10
defective_supplier2<-0.05
cat("prob tire comes from supplier1 if it is defective", (supplier1*defective_supplier1)/((supplier1*defective_supplier1)+

```

R code Exa 4.39 flight accident

```
x= c( 36 ,11 ,51 ,2 ,4  ,13 ,12 ,66 ,9 ,18  ,22 ,14 ,54 ,10 ,78 ,21 ,14 ,56 ,9 ,100)
m <- matrix(x,byrow = TRUE,nrow =4 )
rownames(m) <- c("Business","Instructional ", "Personal","All")
colnames(m) <- c("Fatal","Minor","Serious","None","All")

#a) An accident on a business flight
m ["Business","All"]  /100
```

```
#b)  An accident that resulted in a fatal injury
m["All","Fatal"]/100

#c) An accident that resulted in a minor injury given that it was on a business flight
m["Business","Minor"]/100

#d)  An accident on a business flight that resulted in a minor injury
m["Business","Minor"]/100 * m["Business","All"]/100

#e)  An accident on a business flight given that it was fatal
(m["Business","All"] * m["Business","Fatal"] / m["All","Fatal"])/100
```

R code Exa 4.40 Bays rule for tower malfunction

```
cat("prob of exactly 1 plugging malfunction", (0.096*0.904*0.904)+(0.904*0.096*0.904)+(0.904*0.904*0.096))
cat("prob of atleat 1 plugging malfunction", 1-(0.904*0.904*0.904))
cat("prob of exactly two plugging malfunctions when atleat 1 is due to plugging",
  ((0.096*0.096*0.904)+(0.096*0.904*0.096)+(0.904*0.096*0.096))/(1-(0.904*0.904*0.904)))
```

R code Exa 4.41 Odd ratio

```
x =c(139,10898,11037,239,10795,11034,378,21693,22071)
```

```
m <- matrix(x,byrow = TRUE,nrow = 3)
rownames(m) <- c("Aspirin","Placebo","Total")
colnames(m) <- c("MI","NoMI","Total")

cat("For the aspirin group, the odds in favor of M.I
    . are ",m["Aspirin","MI"]/m["Aspirin","NoMI"])

cat("For the placebo group, the odds in favor of M.I
    . are",m["Placebo","MI"]/m["Placebo","NoMI"])

cat("Odds ratio= ",(m["Aspirin","MI"]/m["Aspirin","
    NoMI"]) /(m["Placebo","MI"]/m["Placebo","NoMI"])
    )
```

R code Exa 4.42 Odd ratio

```
#Employment Status by Gender

x =c(64046,3141,55433,2556)
m <- matrix(x,byrow = TRUE,nrow = 2)
rownames(m) <- c("Emp","Unemp")
colnames(m) <- c("M","F")

cat("Odds ratio:",( m[1,1]*m[2,2])/(m[1,2]*m[2,1]))

#Employment Status by Education
y =c( 36249,  1962, 39250, 1165)
m <- matrix(y,byrow = TRUE,nrow = 2)
rownames(m) <- c("Emp","Unemp")
colnames(m) <- c("HS","College")

cat("Odds ratio:",( m[1,1]*m[2,2])/(m[1,2]*m[2,1]))
z= ( m[1,1]*m[2,2])/(m[1,2]*m[2,1])

```

```
cat(" risk of unemployment for those with a high
   school education is",1/z ," higher than the risk
   of unemployment for those with college education"
   )
```

Chapter 5

Discrete Probability Distributions

R code Exa 5.1 Relay example

```
a <- function(x){

(0.8^x)*(0.2^(2-x))
}

# The Distribution..
cat("The probability distribution for x=0 is",a(0))

cat("The probability distribution value for x=1 is"
   ,2*a(1))

cat("The probability distribution for x=2 is",a(2))
```

R code Exa 5.2 multiplicative rule of probability

```
a <- function(x){
```

```

  ((factorial(6)/factorial(6-x))*(factorial(4)/
     factorial(4-(2-x))))/(factorial(10)/factorial
     (8))

}

# The Distribution..
cat("The probability distribution for x=0 is",a(0))

cat("The probability distribution value for x=1 is"
   ,2*a(1))

cat("The probability distribution for x=2 is",a(2))
```

R code Exa 5.3 distribution function

```
a <- function(x){

  (0.8^x)*(0.2^(2-x))
}

# The Distribution..
cat("The  distribution function for b<0 is",0)

cat("The distribution function for 0<b<1 is",a(0))

cat("The distribution function for 1<b<2 is",2*a(1)
   + a(0))

cat("The distribution function for b>=2 is",a(2) + 2
   *a(1) + a(0) )

b= c(0,1,2,3)
fb= c(0,0.04,0.36,1)
```

```
plot(b,fb)
```

R code Exa 5.4 mean and sd

```
mp =c(3,9,16,21,30,40,55,75,92)
year1 =c(7.6,12.8,5.3,10.8,17.3,15.1,18.6,11.3,1.2)
year2 =c(6.4,11.6,5.2,9.0,12.5,12.2,22.5,16.0,4.6)

mean1 =weighted.mean(mp,year1/100)
mean2 =weighted.mean(mp,year2/100)

sd1=sqrt(sum(((mp- mean1)^2)*year1/100))
sd2=sqrt(sum(((mp- mean2)^2)*year2/100))

cat("mean for year 1990s: ",mean1)
cat("mean for year 2050s: ",mean2)
cat("SD for year 1990s: ",sd1)
cat("SD for year 2050s: ",sd2)
```

R code Exa 5.5 expected daily demand and variance

```
px <- c(0.1,0.5,0.4)
x <- 0:2
E <-  weighted.mean(x,px)
cat("expected daily demand for the tool is ",E)

V <- sum(((x - E)^2)*px)
cat("variance is:", V)
```

R code Exa 5.6 mean and variance of the daily costs

```
px <- c(0.1,0.5,0.4)
x <- 0:2
E <-  weighted.mean(x,px)

# E(100X) = 100 E(X)
cat("Daily cost of using tool" , 100*E)


V <- sum(((x - E)^2)*px)
cat("variance is:", V)

# V(100X) =(100^2)V(X)
cat("Variance of daily cost is ", 100*100*V)
```

R code Exa 5.7 mean score and the standard deviation

```
prob_marks <- c(0.1,0.2,0.4,0.2,0.1)
marks <- 0:4
E <-  weighted.mean(marks,prob_marks)

cat("Mean score is:" , E)

V <- sum(((marks-E)^2)*prob_marks)
cat("Standard deviation is ", sqrt(V))
```

R code Exa 5.8.a Tchebyshefs Theorem

```
mean <- 120
sd <- 10
lower_limit <- 100
k <- (mean - lower_limit)/sd

```

```
cat(1-1/k*k, "fraction of days will have prod. between 100 and 140" )
```

R code Exa 5.8.b shortest interval

```
mean <- 120
sd <-10
fraction <- 0.90
k <-( 1/(1-fraction))^0.5
cat("shortest interval is from ", mean-(k*sd), "–", mean+(k*sd))
```

R code Exa 5.11 binomial distribution of fuses

```

# given probability of defective fuse =0.10

# a)
cat("Probability exactly one fuse in the sample of four is defective", dbinom(1,4,prob = 0.10))

# b)
# Probability atleast one is defective P(X>=1) = 1-P(X=0)
none_defective <- dbinom(0,4,prob = 0.10)
cat("Proability that atleast one bulb is defective : " , 1- none_defective)

# c)
n=4
p=0.1
E =n*p
V =n*p*(1-p)
```

```
E_Y_sq = V+ E^2
cat("E(C) =", 3*E_Y_sq)
cat(" we could expect to pay an average of $",3*E_Y_
    sq*10, "in repair costs for each shipment of four
     fuses. ")
```

R code Exa 5.12 binomial distribution of battery

```
# Probability battery exceeding lifetime of 4 hours
    is 0.135

# a
 cat("probability that only one battery lasts 4
     hours or more is ",  dbinom(1,3,prob = 0.135))

#b
# probability that at least one battery lasts 4
    hours or more = P(Y>=1)= 1- P(Y=0)
 no_battery <- dbinom(0,3,prob = 0.135)
 cat(" probability that at least one battery lasts 4
      hours or more is " , 1- no_battery)
```

R code Exa 5.13 binomial distribution of chemicals

```
#a
cat("The probability of exactly 3 out of 10 plants
    calling in orders is ",
     dbinom(3,10,prob = 0.2))

#b
# The probability of at most 3 out of 10 plants
    calling in orders = P(Y<=3) = p(0)+p(1)+p(2)+p(3)
```

```
cat("The probability of at most 3 out of 10 plants
    calling in orders is " ,  pbinom(3,10,prob = 0.2)
    )

#c
#  The probability of at least 3 out of 10 plants
    calling in orders = 1− P(Y<3)
cat(" The probability of at least 3 out of 10 plants
     calling in orders is" , 1-  pbinom(2,10,prob =
    0.2))
```

R code Exa 5.15 binomial distribution of contracts

```
# a
# The probability that the firm will get none of
    those contracts = P(X=0)

cat("the probability that the firm will get none of
    those contracts" ,dbinom(0,8,prob = 0.40))

#b
# The probability that the firm will get five out of
     eight contracts = P(X=5)

cat(" the probability that the firm will get five
    out of eight contracts", dbinom(5,8,prob = 0.40))

#c
#  The probability that the firm will get all eight
    contracts

cat(" the probability that the firm will get all
    eight contracts" , dbinom(8,8,prob = 0.40))
```

R code Exa 5.17 geomatric distribution of interviews

```
#  the probability that the first applicant having
   advanced training is found on the fifth interview
    = P(Y=5)
# using geomatric distribution
library(stats)
p=0.30
cat(" the probability that the first applicant
   having advanced training is found on the fifth
   interview is", dgeom(4,0.30) )
cat("Total cost of interviewing is ", 300/p)

# V(C)=(300^2)V(Y)

cat("V(C) is" , ((300^2)*(1-p))/(p^2))
```

R code Exa 5.18 expected time

```
#Using negative binomial distribution

r=3
p=0.2
E = 10*(r/p) + r*20
V= 10*10*(r*(1-p)/(p^2))

cat(", the total time to use up the kits has an
   expected value of ",E," minutes and a standard
   deviation of",sqrt(V),"minutes")
```

R code Exa 5.19 geomatric distribution of interviews

```
#  Prob of the applicants for a certain position have
#advanced training in computer programming of the applicants for a certain position
#have advanced training in computer programming of the applicants for a certain position have
#advanced training in computer programming = 0.30

cat(" The probability that the third qualified applicant is found on the fifth interview" ,
   dnbinom(2,3,0.30))
```

R code Exa 5.20 industrial accidents

```
# average no. of accidents in a week is 3

# using poisson distriution

#a
cat("Prob of no accidents in a week p(0)= ", dpois(0,3))

#b
cat("Prob of two accidents in a given week p(2)= ", dpois(2,3))

#c Prob atmost 4 accidents occur in a given week is
# p(0)+p(1)+p(2)+p(3)

cat("Prob atmost 4 accidents occur in a given week is ", ppois(4,3))

#d Average no. of accidents on a given day = 3/7
```

```r
d <- 3/7
cat("Prob of two accidents on any given day is ",
    dpois(2,d))
```

R code Exa 5.22 The Hypergeometric Distribution

```r
# using hypergeometric distribution

# the probability that the female is selected for
    one of the jobs =P(Y=1) = p(1)

cat("the probability that the female is selected for
     one of the jobs is", dhyper(1,1,5,2))
```

R code Exa 5.23.a The Hypergeometric Distribution of boxes

```r
# using hypergeometric distribution

#a
# the probability that all five boxes will fit
    properly = P(Y=0) = p(0)

cat("The probability that all five boxes will fit
    properly is", dhyper(0,2,18,5))
```

Chapter 6

Continuous Probability Distributions

R code Exa 6.1 continuous random variable

```
a <- function(x){
  0.5*exp(-x*0.5)
}


cat("the probability that battery will last longer
   than 400 hours",integrate(a,4,'infinite')$value)
cat("probability that the lifetime exceeds 9 is",
   integrate(a,9,'infinity')$value)
```

R code Exa 6.2 probability density function

```
a <- function(x){
  0.5*exp(-x*0.5)
}

```

```
#a
cat(" the probability that the lifetime of a
    particular battery of this type is less than 200
    hours",
    integrate(a,0,2)$value)
# ans given in book is wrong ,P(X<2)= 0.6321

#b
cat(" the probability that a battery of this type
    lasts more than 30 hours given that it has
    already been in use for more than 200 hours. ",
    (integrate(a,3,'infinity')$value)/(integrate(a
        ,2,'infinity')$value))
```

R code Exa 6.4 distribution function

```
a = function(x){
  if(x<0){
    0
  } else if(x>=0 && x<=1){
    x
  } else if(x>=1 && x<=2){
    0.5
  } else {
    0
  }
}

cat("the probability that demand will exceed 150
    gallons on a given week",integrate(Vectorize(a)
    ,1.5,2)$value)
```

R code Exa 6.5 Expected value and variance

```
a = function(x){
  3*x*x*x
}

E= integrate(a,0,1)$value

cat(" on the average, the lathe is in use ", E*100,"
   percent of the time")

b =function(x){
  3*x*x*x*x
}

cat("Variance is ",integrate(b,0,1)$value - E^2 )
```

R code Exa 6.6 expected weekly demand

```
a = function(x){
  if(x<0){
    0
  } else if(x>=0 && x<=1){
    x*x
  } else if(x>=1 && x<=2){
    x/2
  } else {
    0
  }
}

b =integrate(Vectorize(a),0,1)$value
c=  integrate(Vectorize(a),1,2)$value
cat("The expected weekly demand for kerosene is ",b+
   c," gallons")
```

R code Exa 6.7 Expected interval

```
E =445
V =236
sd =sqrt(V)
p= 0.75
k =sqrt((1/(1-p)))

cat(" This interval isgiven by ",E- k*sd," - ",E+ k*
  sd)
```

R code Exa 6.8.a probability of uniform distribution

```

# Unniform distriution problem

#a
#  the probability that the delivery time is two or
  more days = P(X>=2)

cat(" the probability that the delivery time is two
  or more days is " , 1-punif(2,1,5))
```

R code Exa 6.9.a Probability of exponential distribution

```

# Using exponential distriution

#a
```

```
# The probability that any given plant processes
    more than 5 tons of raw sugar on a given day = P(
    Y>=5)

mean <- 4
Theeta <- 1/mean
a <- 1 - pexp(5, rate = Theeta)
cat("The probability that any given plant processes
    more than 5 tons of raw sugar on a given day is",
    a)
```

R code Exa 6.9.b cumilative probability

```

# Using binomial distribution
#b

mean <- 4
Theeta <- 1/mean
a <- 1 - pexp(5, rate = Theeta)

cat(" the probability that exactly two of the three
    plants process more than 5 tons of raw sugar on a
    given day"
    , dbinom(2,3,prob = a) )
```

R code Exa 6.10 The Gamma Distribution of components

```

# using exponential distribution , alpha=1
# As Y= X1 + X2 , alpha =2  beta =400

#b
```

```
cat("Expected value E(Y) =", 2 * 400)

#c
# using Gamma distriution
 cat(" The probability that the system will survive
    for more than 1000 hours is equal to P(Y>1000) =
    ",
      1 - pgamma(1000,shape = 2 ,scale = 400))
```

R code Exa 6.11 The mean and variance for the length of maintenance times

```
alpha <- 3
beta <- 2
sd <- ((alpha * beta * beta)^2)
cat("P(|Y-6| >=14 ) = " , (3.46/14)^2)


cat(" Because P(Y>20 min) is so small, we must
   conclude that our new maintenance man is somewhat
    slower than his predecessor. ")
```

R code Exa 6.12 z value

```
cat("P(Z< =1.53) =",pnorm(1.53,lower.tail = TRUE))
```

R code Exa 6.13 Normal distribution

```
# using normal distribution

```

```
#a
cat("P(Z <= 1) =", pnorm(1,lower.tail = TRUE))

#b
cat("P(Z < -1.5) =" , pnorm(-1.5,lower.tail = TRUE))

#c
cat("P(Z > 1) =", pnorm(1,lower.tail = FALSE))

#d
cat("P(-1.5 <= Z < =0.5) =", pnorm(0.5)-pnorm(-1.5))

#e
cat("The value of z such that P(Z<=z)= 0.99 is",
    qnorm(0.99,lower.tail = T))
```

R code Exa 6.14 standard normal distribution of bottles

```

# area using normal distribution


# given population mean = 16 and sd = 1

cat(" the probability that the machine will dispense
     more than 17 ounces of liquid into any one
    bottle. ",pnorm(17,16,1,lower.tail = F))
```

R code Exa 6.15 Normal distribution

```

# using normal distribution

```

```
z <- qnorm(0.05,lower.tail = F)
cat("The value of z such that P(Z>=z)= 0.05 is" ,z)
sd <- 1.2
cat("Therefore, the proper setting for the dial so
   that 17−ounce bottles will overflow only 5% of
   the time is when mean =",
    17 - sd*z)
```

R code Exa 6.16 Normal distribution of maths score

```

# using normal distribution

#a
# SAT mathematics scores mean = 480 and sd = 100
a <- pnorm(550,480,100,lower.tail = T)
cat(" percent of students would score less than 550
   in a typical year is P(X<550)", a * 100)

#b
# ACT mathematics scores mean = 18 and sd = 6
b <- (550-480)/100
cat(" The engineering school set as a comparable
   standard on the ACT math test would be" , 18 + 6*
   b)

#c
cat(" the probability that a randomly selected
   student will score over 700 on the SAT math test
   = P(X>700)",
    pnorm(700,480,100,lower.tail = F))
```

R code Exa 6.17.a percentile score

```r

# mean and sd of the batting league is 0.358 and 0.027 respectively

#a
# z sore for Ted Williams is (0.406-0.358)/0.027
z1 <- (0.406-0.358)/0.027
cat("z-score for Ted Williams is ", z1)
cat("Percentile score for Ted Williams is", pnorm(z1,lower.tail = T))

# z score for George Brett is (0.390-0.358)/0.027
z2 <- (0.390-0.358)/0.027
cat("z score for George Brett is ", z2)
cat("Percentile score for George Brett is",pnorm(z2,lower.tail = T))


cat("The percentile score for Ted Williams is 0.96 while that for George Brett is 0.88. Both the performances are outstanding; however, Ted Williams did slightly better than George Brett. ")
```

R code Exa 6.17.b probability

```r

#b

cat(" The chance of the League leader hitting over 0.400 in a given year can be ",
  pnorm(0.400,0.358,0.027,lower.tail = F))
```

R code Exa 6.18 octane rating

```
octane =c(88.5 ,95.6 ,88.3 ,94.2 ,89.2, 93.3 ,89.8
   ,91.8 ,90.4, 92.2 ,87.7, 93.3 ,87.6 ,92.7, 88.3
   ,91.8 ,89.6, 91.6, 89.3, 92.2, 83.4, 94.7 ,84.3
   ,93.2 ,85.3, 92.3, 87.4, 90.4, 89.7, 91.2, 86.7,
   91.1 ,86.7, 91.0, 87.9 ,90.4, 88.9 ,91.1, 90.3
   ,91.0, 87.5, 91.0, 88.2, 90.3 ,88.6, 90.1 ,91.2,
   92.6, 91.6 ,92.2 ,91.5, 94.2, 90.8, 93.4 ,90.9
   ,93.0, 89.3 ,89.8, 90.5, 90.0, 88.6, 87.8, 88.3
   ,88.5 ,89.0 ,88.7 ,94.4, 90.6 ,90.7 ,100.3 ,93.7,
    89.9, 98.8 ,90.1, 96.1 ,89.9, 92.7 ,91.1 ,92.7)
summary(octane)
p_value =pnorm(96,mean=mean(octane),sd = sd(octane),
   lower.tail = FALSE)
cat("P(ocatane rating) > 96 =",p_value)

hist(octane)
```

R code Exa 6.20 QQ plot

```
percentile =c(5,10,20,30,40,50,60,70,80,90,95,99.9)
z_score =(qnorm(percentile /100))

florida =c
   (180.448,198.383,220.102,235.762,249.11,261.651,278.539,303.200,3

qqnorm(florida,ylim = c(150,500),xlim =c(-3,3))
qqline(florida)

nebraska =c
   (165.275,178.736,195.036,206.79,216.832,226.219,235.606,245.649,2

```

```
qqnorm(nebraska,ylim = c(100,350),xlim=c(-3,3))
qqline(nebraska)
```

R code Exa 6.22 CO z score

```
CO =c( 1.7, 1.8, 2.1, 2.4, 2.4, 3.4, 3.5, 4.1, 4.2,
   4.4, 4.9, 5.1, 8.3, 9.3, 9.5)
qqnorm(CO)

i =1:15
z =i/(15+1)
z_score =qnorm(z)
```

R code Exa 6.23 lognormal distribution

```

# Given X = beryllium contamination follows
   lognormal distribution
# Parameters of X ,mu = -2.291 , sd = 1.276

# Using lognormal distribution
#a
cat(" the probability that wipe sample  will have
   beryllium contamination exceeding 0.50 gm/100 cm2
    = P(X>0.50) "
     , 1 - plnorm(0.50,meanlog = -2.291,sdlog =
        1.276))

#b
cat("The 95th percentile for contamination
   distribution is", qlnorm(0.95,meanlog = -2.291,
   sdlog = 1.276))
```

```
cat("The result exceeds the DOE safety limit of 0.20 and thus, we can conclude that the beryllium contamination at this smelter is at an unhealthy level for workers")


#c
cat("E(X)= ", exp(-2.291 + (1.276^2)/2))
cat("V(X)= ", (exp((2*(-2.291)) + (1.276^2)))*((exp(1.276^2)) - 1 ))
```

R code Exa 6.24 beta distribution

```

# Let X denote the proportion of the total supply sold in a given week
# Given alpha = 4, beta = 2

# Using beta distribution

#a
cat(" The expected proportion of supply sold in a given week is ", 4/(4+2))

#b
cat("P(X>0.90) = 1 - P(X<=0.90) =", 1 - pbeta(0.90,4,2))
cat("It is not very likely that 90% of the stock will be sold in a given week. ")
```

R code Exa 6.25 Relative frequency histograms and densities

```
lifetime_sqroot =c(0.637, 1.531, 0.733, 2.256, 2.364 ,1.601, 0.152 ,1.826 ,1.868, 1.126, 0.828 ,1.184, 0.484 ,1.207, 0.719, 0.715 ,0.474 ,1.525, 1.709, 1.305, 2.186, 1.228, 2.006 ,1.032, 1.802 ,1.668 ,1.230, 0.577, 1.274, 1.623 ,1.313 ,0.542, 1.823, 0.880 ,1.526, 2.535, 1.793 ,2.741, 0.578, 1.360 ,2.868, 1.493 ,1.709, 0.872, 1.032, 0.914 ,1.952 ,0.984 ,2.119, 0.431)

#Exponential density plot

hist(lifetime_sqroot^2,breaks = 9,probability = TRUE ,main = "Histogram of Lifetime",xlab = "Lifetime" ,ylab = "Probability")

# Weibull distribution plot differs from the one given in textbook
#Weibull distribution

hist(lifetime_sqroot,breaks = 12,probability = TRUE, main = "Histogram of Sq root of Lifetime",xlab = "sq root",ylab = "Probability")
```

R code Exa 6.26 Weibull Distribution

```

# USing Weibull Distribution

Theeta =50
gama = 2

#a
cat("P(X>10) = ", 1 - pweibull(10,shape = gama,scale = sqrt(Theeta)))

```

```
#b
cat(" Expected lifetime of thermisters is E(X) =", (Theeta^(1/gama))*(gamma(1+(1/gama))))
```

R code Exa 6.27 Weibull Distribution

```
cables =c(5300,5410,5500,5700,5710,5750,5800,5810,5870,5900,5900,5990,6050

L =c(-3.481,-2.772,-2.351,-2.046,-1.806,-1.606,-1.434,-1.281,-1.144,-
fit  =lm(L~log(cables))
plot(log(cables),L,main = "Plot of LF( x ) versus ln( x ) for  the cable strength data")
abline(reg = fit)
```

Chapter 7

Multivariate Probability Distributions

R code Exa 7.2 joint probability distribution of X1 and X2

```

# Joint Probability Distribution
# a
library(MASS)
# formula for Joint Probability Distribution

a <- function(x,y)
{
  if(x+y <= 2)
  {
    if(x==1 || y==1)
    {
     ans = ( 2 * (1/3) * (1/3))

    }
    else{
      ans = ((1/3) * (1/3))
    }
  }
```

```

  else{
    ans = (0)
  }
}

for(i in 0:2){
  for(j in 0:2)
  {
    cat("p(", i,",",j,") =")
    print(fractions(a(i,j)))
  }

}

# b
#  the probability that one of the customers visits counter B given that one of the customers is known to have
# visited counter A.
cat("P(X2 =1 | X1 =1) =", a(1,1)/(a(1,0) + a(1,1) + a(1,2)))
```

R code Exa 7.3 Joint probability distribution of imprities

```

# Joint Density Function of Continuous Random Variable..

a <- function(x1,x2){
  2 * (1-x1)

}


```

```
#  to find the probability of a region A:{(x1,x2)|
   0<x1<0.5, 0.4<x2<0.7}
cat("The probability is", " ")

integrate(function(x2){

  sapply(x2,function(x2){

    integrate(function(x1)a(x1,x2),0,0.5)$value
  })
},0.4,0.7)
```

R code Exa 7.4 marginal probability density functions

```

a <- function(x1,x2){
  2 * (1-x1)
}
cat("f1(x1) = 2(1-x1)")
cat("Since , f2(x2) =", integrate(a,0,1)$value,",
   our conjecture is verified")
```

R code Exa 7.5 The marginal probability density functions

```

a <- function(x1){
  3 * x1
}

b <- function(x2){
  1.5 * (1-( x2*x2))
}

```

```
cat("the probability that X2 will be between 0.2 and 0.4 for a given week = ", integrate(b,0.2,0.4)$value)
```

R code Exa 7.6 conditional probability

```

# using conditional probability distribution

f <- function(x,y){
  3*x
}

f1 <-  function(x){
   3*x*x
 }


d <- function(y){

  f(0.5,y)/f1(0.5)

   }


cat("The value of conditional Probability P(0<X2<0.2|X1= 0.5) is ", integrate(Vectorize(d), 0, 0.2)$value)
```

R code Exa 7.7 conditional probability density function

```

# using conditional probability distribution
```

```r

a <- function(x,y){
    0.5
}

b <- function(y){
  0.5 * y
}

c <- function(x){
  a(x,1)/b(1)
}

cat(" The probability of interest is ", integrate(
    Vectorize(c),0,0.5)$value)

#  if the machine had contained 2 gallons at the
    start of the day
d <- function(x){
  a(x,2)/b(2)
}

cat(" The probability of interest is ", integrate(
    Vectorize(d),0,0.5)$value)
```

R code Exa 7.10 covariance between two random variables

```r

# to find the covariance between two variables X and
     Y

rm("c")

a <- matrix(c(0.04,0,0,0.16,0.10,0,0.20,0.30,0.20),
    nrow = 3, ncol = 3)
```

```
# individual column sum
ax <- apply(a,2,sum)


# indivdual row sum
ay <- apply(a,1,sum)

# E[X]..
ex <- sum(0:2*ax)

# E[Y]..
ey <- sum(0:2*ay)


# E(XY)..
exy <- 0

for(i in 0:2){

  for(j in 0:2){

    exy <- exy + i*j*a[i+1,j+1]

  }
}
Cov <- exy - ex*ey
cat("CoV(X1,X2 =)", Cov)

df<- function(l,m)
{
  (l-m) * (l-m)
}
cat("V(X1) = " , sum(df((0:2),ey)*ay))
cat("V(X2) =" , sum(df((0:2),ex)*ax))

V1 <- sum(df((0:2),ey)*ay)
```

```
V2 <- sum(df((0:2),ex)*ax)

Ro <- Cov/(sqrt(V1 * V2))

cat("Correlation = ", Ro)
```

R code Exa 7.12 mean and variance

```
a <- function(x)
{
  x*(3*x*x)
}
EX1 =integrate(a,0,1)$value

b <- function(y)
{
  y*1.5*(1- y^2)
}
EX2 =integrate(b,0,1)$value

c <- function(x)
{
  x*x*(3*x*x)
}
EX1sq =integrate(c,0,1)$value

d <-function(y)
{
  y*y*1.5*(1- y^2)
}
EX2sq =integrate(d,0,1)$value

VX1 = EX1sq - EX1^2
VX2 = EX2sq - EX2^2

```

```
e <- function(x)
{
  1.5*(x^4)
}
EX1X2 =integrate(e,0,1)$value

Cov = EX1X2 - EX1*EX2

cat("E(Y) = ",EX1 - EX2)
cat("V(Y) = ",VX1 + VX2 + 2*1*(-1)*Cov)
```

R code Exa 7.13 mean and variance of the total weekly amount

```
# The dollar amount spent per week is given by Y =3
    X1 +5X2
# given , E(X1) = 40 , V(X1) = 4
# E(X2) = 65, V(X2) = 8

cat("E(Y) = 3E(X1) + 5(E(X2) :" , 3*40 + 5*65)
cat("V(Y) = (3^2)*V(X1) + (5^2)*V(X2))" , (3^2)*4 +
    (5^2)*8)
```

R code Exa 7.15 multinomial probability distribution

```

# to find the probability that three bulbs have no
    defects , one has a type A defect , and two have
    type B defects out of 6
# bulbs chosen from a lot

# Using multinomial distribution

p_bulb <- c(0.70,0.20,0.10)
```

```

# No. of bulbs chosen without defect , with defect A, with defect B are:
x <- c(3,1,2)

cat(" the probability that three bulbs have no defects , one has a type A defect , and two have type B defects is",
    dmultinom(x,6,p_bulb))
```

R code Exa 7.19 conditional expectation

```

a <- function(x1,x2){
  0.5
}

b <- function(x2){
  0.5 * x2
}

c <- function(x1){
 x1 * a(x1,1)/b(1)
}

cat("  the conditional expectation of amount of sales X1 given that X2 = 1 is ", integrate(Vectorize(c),0,1)$value)
```

R code Exa 7.20 conditional distribution of X1 given X2

```

# E(Y) = E(E(Y|p))
```

```
# using binomial distribution
# E(Y|p)= E(n*p) = n*E(p)
# given , n=10

a <- function(p)
{
  10 * (4*p)
}

cat(" the expected value of Y for any given day",
   integrate(a,0,0.25)$value)
```

Chapter 8

Statistics Sampling Distributions and Control Charts

R code Exa 8.2 central limit theorem

```

# given , mean =10, sd=10

# area for the region P(a<=X<=b) =0.95
# area for the region P(X<=b)
x <- (1-0.95)/2 +0.95

ll <- qnorm(x,lower.tail = FALSE)
ul <- qnorm(x,lower.tail = T)
cat("Therefore , Z lies between", ll," to ",ul)

y =function(n){
  a <- ( 10 + (ll*(10/sqrt(n))))
  b <- ( 10 + (ul*(10/sqrt(n))))
  output <- list(a,b)
  return(output)
}
```

```

output1 <- y(25)
cat("interval when n=25 is ")
cat("lower limit")
output1[1]
cat("Upper limit")
output1[2]


output2 <- y(50)
cat("interval when n=50 is ")
cat("lower limit")
output2[1]
cat("Upper limit")
output2[2]


output3 <- y(100)
cat("interval when n=100 is ")
cat("lower limit")
output3[1]
cat("Upper limit")
output3[2]
```

R code Exa 8.3 central limit theorem for average fracture strength

```
# given , sigma =2 , n=100

#a
# according to central limit theorem
sigma =2
n = 100
sd = sigma/sqrt(n)
cat("the probability that the average fracture
    strength of glass exceeds 14.5 is",pnorm(14.5,14,
```

```
    sd,lower.tail = F),
      "which is very small")

#b
x <- (1-0.95)/2 +0.95

ll <- qnorm(x,lower.tail = F)
ul <- qnorm(x,lower.tail = T)
cat("The limit is a = ", 14+ ll*sd,"to b = ",14 + ul
    *sd)
```

R code Exa 8.4 Probability of sample mean

```
#given sigma=1, n=25
sigma =1

#a
n = 25
sd <= sigma/sqrt(n)

cat(" the probability that the sample mean will be
    within 0.3 ounces of the true population mean is"
    ,
      pnorm(.3/sd) - pnorm(-0.3/sd))

#b

# given , P(-0.3<|X-mu|<0.3) = 0.95
x <- (1-0.95)/2 +0.95

ll <- qnorm(x,lower.tail = F)
ul <- qnorm(x,lower.tail = T)

n<- (ul/0.3)^2
cat("Value of n so that the sample mean will be
```

```
within 0.3 ounces of the population mean with
probability 0.95 is ",n)
```

R code Exa 8.5 Normal distribution of failure strengths

```

# Application of t distribution
#a
xbar1 <- 2000

mu <- 3000

sd <- 989

n <- 12

t_value = (xbar1-mu)/(sd/sqrt(n))
cat("The required probability is ", pt(t_value, df =
    n-1))

#b
xbar2 <- 2500

t_value = (xbar2-mu)/(sd/sqrt(n))
cat("The required probability is ", pt(t_value, df =
    n-1))
```

R code Exa 8.6 The Sampling Distribution of large samples

```
Y <- 12
p <- 0.2
n <- 100
mu <- p
```

```
sd <- sqrt(p*(1-p)/n)


z_value = (Y/n - mu)/sd
cat("The required probability is ", pnorm(z_value))
cat("There is only a small probability of" ,pnorm(z_
   value) , " of accepting any lot that has 20%
   nonconforming wafers. ")
```

R code Exa 8.7 The Sampling Distribution of S2

```
# USing chi-square distribution
# give , vraiance =0.8,n =10
var=0.80
n =10
ll = 0.05
ul = 0.95
a= var*qchisq(0.05,df=9)/(n-1)
b= var*qchisq(0.95,df=9)/(n-1)
cat("value of a and b such that the sample variance
   of the amounts dispensed will be between a and b
   with probability 0.90 is",
    a,"-",b )
```

R code Exa 8.8 The Sampling Distribution of S 2

```
# using chi-sq distribution

# given , variance =100 , n=25

var =100
n=25

```

```
#a
#P(S^2 > 50)
a <- (n-1)*50/var
cat("approximate probability = ", pchisq(a,df=n-1, lower.tail = F))

#b
#P(S^2 > 150)
b <- (n-1)*150/var
cat("approximate probability = ", pchisq(b,df=n-1, lower.tail = F))

#c
cat("E(S^2) = ", var)
cat("V(S^2) = ", 2*var*var/(n-1))
```

R code Exa 8.9 General distribution large samples

```
rm("c")
#Sampling Distribution of the difference of 2 means ..

# Mean, variance and no. of samples for both machines are as follows

A <- c(1,200,25)
B <- c(1,200,25)

diff_mean <- A[1]-B[1]

diff_sd <- sqrt((A[2]/A[3])+(B[2]/B[3]))

cat(" the probability that the difference in sample means for two machines will be at most 10 ml is "
,
```

```
    pnorm(10,diff_mean,diff_sd) - pnorm(-10,diff_mean,diff_sd) )
```

R code Exa 8.10 Small samples case equal variances

```

#Sampling Distribution of the difference of 2 means..

# Mean, variance and no. of samples for both groups are as follows

A <- c(450,17.795,6)
B <- c(250,9.129,4)

diff_mean <- A[1] - B[1]

diff_sd <- sqrt(((A[3]-1)*A[2]*A[2] + (B[3]-1)*B[2]*B[2])/ (A[3]-1 + B[3]-1))

# P(X1-X2 >= 150)
x <- (150 - diff_mean)/sqrt(diff_sd*diff_sd*(1/A[3] + 1/B[3]))

#degree of freedom = 6+4-2
cat(" probability that the sample mean tensile proof stress for group 1 is at least 150 MPa larger than that for group 2 is ",pt(x,df=8,lower.tail = F))
```

R code Exa 8.11 The sampling distribution of XDbar

```
A <- c(10.18,12.19,12)
```

```
B <- c(-14.71,4.40,12)
mu <- 0
#a
#P(d>=10.18)
a <- (A[1]- mu)/(A[2]/sqrt(A[3]))

cat("probability of observing a mean difference of a
    At least 10.18 kN for EOF measurements is",pt(a,
   df=A[3]-1,lower.tail = F))

#b
#P(d<= -14.71)
b <- (B[1]- mu)/(B[2]/sqrt(B[3]))
cat("probability of observing a mean difference of a
    At most -14.71 kN for EOF measurements is", pt(b
   ,df=B[3]-1,lower.tail = T))
```

R code Exa 8.12 sampling distribution of p1 minus p2

```
p1= 0.04
p2= 0.025
n1 = 200
n2 = 200
diff_mean = p1-p2
diff_sd = sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)

cat("P(|p1 - p2|) =",  pnorm(0.02,diff_mean,diff_sd,
   lower.tail = FALSE) )

# There is a fairly high chance (38.88%) of
   observing a difference of at most 2 percentage
   points between the sample proportion defectives

#Answer given in book is wrong. Answer will be twice
    of what given in book.
```

R code Exa 8.13.a sampling distribution of S1sq divided S2sq

```
n1 = 10
n2 = 10

# 2 cases: s1*s1 > 2*s2*s2 and s2*s2 > 2*s1*s1
# Let s1^2/s2^2 =X

# The probability of observing one sample variance
    at least 2 times larger than the other is
#P(X<0.5) + P(X>2)

# Using F distribution

cat("P(F(9,9)<0.5) + P(F(9,9)>2) = ", pf(0.5,9,9) +
    pf(2,9,9,lower.tail = F))

cat(" There is approximately ", (pf(0.5,9,9) + pf
    (2,9,9,lower.tail = F))*100, " chance that one
    sample variance will be at least 2 times larger
    than the other
, even if the population variances are equal.")
```

R code Exa 8.13.b probability of sampling distribution

```
n1 = 10
n2 = 10

# 2 cases: s1*s1 > 4*s2*s2 and s2*s2 > 4*s1*s1
# Let s1^2/s2^2 =X

```

```
# The probability of observing one sample variance at least 2 times larger than the other is
#P(X<0.25) + P(X>4)

# Using F distribution

cat("P(F(9,9)<0.25) + P(F(9,9)>4) = ", pf(0.25,9,9) + pf(4,9,9,lower.tail = F))

cat(" There is approximately ", (pf(0.25,9,9) + pf(4,9,9,lower.tail = F))*100, " chance that one sample variance will be at least 2 times larger than the other
   , even if the population variances are equal.")
```

R code Exa 8.14 Xbar and R charts

```
m1 =c(16.1,16.2,16.0,16.1,16.5,16.8,16.1,15.9,15.7,16.2,16.4,16.5,16.7
m2 =c(16.2,16.4,16.1,16.2,16.1,15.9,16.9,16.2,16.7,16.9,16.9,16.9,16.2
m3 =c(15.9,15.8,15.7,15.9,16.4,16.1,16.2,16.8,16.1,16.1,17.1,17.2,16.4
m4 =c(16.0,16.1,16.3,16.4,16.4,16.3,16.5,16.1,16.4,17.0,16.2,16.1,15.8
m5 =c(16.1,16.2,16.1,16.6,16.2,16.4,16.5,16.4,16.8,16.4,16.1,16.4,16.6

obseravation <- c(m1,m2,m3,m4,m5)

```

```
samples <- c(rep(1:20,5))


dat <- data.frame(obseravation,samples)


print("The xbar and S chart for the above data is:")

#install the package qicharts for xbar chart

library(qicharts)
# Run the below two code individually..
#xbar chart
qic(obseravation,
    x= samples,
    data = dat,
    chart = 'xbar',
    xlab = 'Sample Number')

#install the package qcc for R chart

# R chart
library(qcc)
dat1=data.frame(m1,m2,m3,m4,m5)
qcc(dat1,type = "R")
```

R code Exa 8.15 X bar and S charts

```
m1 =c
   (16.1,16.2,16.0,16.1,16.5,16.8,16.1,15.9,15.7,16.2,16.4,16.5,16.7

m2 =c
   (16.2,16.4,16.1,16.2,16.1,15.9,16.9,16.2,16.7,16.9,16.9,16.9,16.2

m3 =c
```

```
    (15.9,15.8,15.7,15.9,16.4,16.1,16.2,16.8,16.1,16.1,17.1,17.2,16.4

m4 =c
    (16.0,16.1,16.3,16.4,16.4,16.3,16.5,16.1,16.4,17.0,16.2,16.1,15.8

m5 =c
    (16.1,16.2,16.1,16.6,16.2,16.4,16.5,16.4,16.8,16.4,16.1,16.4,16.6


obseravation <- c(m1,m2,m3,m4,m5)

samples <- c(rep(1:20,5))


dat <- data.frame(obseravation,samples)


print("The xbar and S chart for the above data is:")


#Use the package qicharts

library(qicharts)
# Run the below two code individually..
#xbar chart
qic(obseravation,
     x= samples,
     data = dat,
     chart = 'xbar',
     xlab = 'Sample Number')

# S chart
qic(obseravation,
     x = samples,
     chart = 's',
     xlab = 'Sample Number',
     data = dat)

```

```
#UCL and LCL values have been rounded off.
```

R code Exa 8.16 p chart

```
rm("c")
x =c
   (3,1,4,2,0,2,3,3,5,4,1,1,1,2,0,3,2,2,4,1,3,0,2,3)

sample =1:24
dat <- data.frame(sample,x)

p <- mean(dat$x/50)


u <- p + 3*sqrt(p*(1-p)/50)

l <- p- 3*sqrt(p*(1-p)/50)

cat("The LCL and UCL are",0,"and", u,"respectively")

#Since l is neg. , we take lower limit to be 0.

#install the package qcc

library(qcc)
qcc(dat$x, sizes =50,type="p")
```

R code Exa 8.17 c chart

```
defective =c
   (6,3,4,0,2,7,3,1,0,0,4,3,2,2,6,5,0,7,2,1)
sample =1:20
dat <- data.frame(sample,defective)
```

```
n=20
c =sum(defective)/n
u =c + 3*sqrt(c)
l =c - 3*sqrt(c)

cat("The LCL and UCL are",0,"and", u,"respectively")
#Since l is neg. , we take lower limit to be 0.

# install the package qcc.
library(qcc)
qcc(dat$defective, type = "c")
```

R code Exa 8.18 u chart

```
defect =c(1,4,1,2,1,4,3,5,3,1,2,1)
hours =c
    (58.33,80.22,209.24,164.70,253.70,426.90,380.20,527.70,319.30,340

part =1:12
dat = data.frame(part,defect,hours)
u_bar= sum(dat$defect)/sum(dat$hours)

ucl1=u_bar + 3*sqrt(u_bar/dat$hours[1])
ucl2=u_bar + 3*sqrt(u_bar/dat$hours[2])

u1 =dat$defect[1]/dat$hours[1]
u2 =dat$defect[2]/dat$hours[2]


#install the package qicharts for u chart

library(qicharts)
# Run the below  code ..
# u chart
qic(defect,hours,
```

```
    x= part   ,
    data = dat,
    chart = 'u',
    xlab = 'Sample Number')
```

R code Exa 8.19 total proportion out of specification

```
m1 =c
   (16.1,16.2,16.0,16.1,16.5,16.8,16.1,15.9,15.7,16.2,16.4,16.5,16.7
m2 =c
   (16.2,16.4,16.1,16.2,16.1,15.9,16.9,16.2,16.7,16.9,16.9,16.9,16.2
m3 =c
   (15.9,15.8,15.7,15.9,16.4,16.1,16.2,16.8,16.1,16.1,17.1,17.2,16.4
m4 =c
   (16.0,16.1,16.3,16.4,16.4,16.3,16.5,16.1,16.4,17.0,16.2,16.1,15.8
m5 =c
   (16.1,16.2,16.1,16.6,16.2,16.4,16.5,16.4,16.8,16.4,16.1,16.4,16.6

obseravation <- c(m1,m2,m3,m4,m5)
m =mean(obseravation)
sigma1 =0.361
sigma2 =0.367
USL =17
LSL =16
zUSL = (USL - m)/sigma1
zLSL = (m - LSL)/sigma2
zmin = min(zUSL,zLSL)
Cpk =zmin/3
cat("The area below zLSL is",1- pnorm(zLSL))
cat("The area above zUSL is",1- pnorm(zUSL) )
```

```
prop= 1- pnorm(zLSL) + 1- pnorm(zUSL)
cat("proportion out of specification =",prop)
```

Chapter 9

Estimate

R code Exa 9.4 Large Sample Confidence Interval for a Mean

```
#Confidence interval evalualtion
#For confidence coefficient to be 95%, , it leaves
   (1 - 0.025 = 0.975) area to the left

n <- 50

mu <- 2.268

sig <- 1.932

a <- qnorm(0.975)*sig/sqrt(n)


cat("The 95% confidence interval for the above
   parameters is",mu-a,mu+a," hours")
```

R code Exa 9.6 Determining Sample Size to Estimate Mean

```r
# Determining Sample Size to Estimate Mean...

sigma =2
B = 0.1

cat("Thus, at least ", round((qnorm(0.975)*sigma/B)
    ^2) ,"employees should be sampled to achieve the
    desired results. ")
```

R code Exa 9.7 lower limit for the mean lifetime of batteries

```r
# givn confidence level =95%, we'll calculate z(1-
    0.95)

n <- 50

mu <- 2.266

sig <- 1.935

a <- qnorm(0.05, lower.tail = F)

# To calculate lower limit
cat("Lower limit is ", mu - a*sig/sqrt(n) )
```

R code Exa 9.8 Confidence Interval for a Mean Based on t distribution

```r

# Using T distribution

uts <- c(253,261,258,255,256)

mu = mean(uts)
```

```
sig = sd(uts)
n = 5
alpha = 1 -0.95

a <- qt(alpha/2, df= n-1)

cat("Therefore, the interval is ", mu + a*sig/sqrt(n
), " - ", mu - a*sig/sqrt(n))
```

R code Exa 9.9 Large Sample Confidence Interval for a Proportion

```
# confidence interval =90%

alpha = 1-0.90
p =0.20
n =100
a =qnorm(1 - alpha/2,lower.tail = F)

cat(" the true probability p of finding this
    microorganism in a sample is somewhere between ",
     p - a*sqrt(p*(1-p)/n),
     " - ", p + a*sqrt(p*(1-p)/n))
```

R code Exa 9.10 number of workers

```
#Determining sample size
#Given , confidence limit =90%
#Because no prior knowledge of p is available, use p
    = 0.5

alpha=1-0.90
B =0.05
p=0.5
```

```
a <- qnorm(alpha/2,lower.tail = F)

cat("A random sample of at least ",round(((a/B)^2)*p*(1-p)," workers is required in order to estimate the true proportion favoring
  the revised policy to within 0.05. ")
```

R code Exa 9.11 Confidence Interval for a Variance

```
readings <- c(9.54,9.61,9.32,9.48,9.70,9.26)

var= (sd(readings))^2
n=6

# given Confidence interval =90%
alpha= 1-0.90

# Using chi-sq distribution

a <- qchisq(1 - (alpha/2),df= n-1,lower.tail = F)
b <- qchisq((alpha/2),df= n-1,lower.tail = F)

cat("Thus, confidence interval is ", (n-1)*var/b," - ", (n-1)*var/a)
```

R code Exa 9.12 Confidence Interval for a Difference in Means

```

# Mean , variance and no. of observations for both machines are as follows:
A <- c(12,6,100)
B <- c(9,4,100)

```

```
# given confidence nterval =90%
alpha = 1-0.90
diff_mean <- A[1] - B[1]

diff_sd <- sqrt(A[2]/A[3] + B[2]/ B[3])


a = qnorm(alpha/2, lower.tail = F)
cat("We are about 90% confident that the difference
    in mean daily downtimes is between, ", diff_mean
    - a*diff_sd," − ", diff_mean + a*diff_sd," min")
```

R code Exa 9.13 Confidence Interval for a Linear Function of Means

```
# Mean , variance and no. of observations for 3
    machines are as follows:

A <- c(12,6,100)
B <- c(9,4,100)
C <- c(14,5,100)

#Expected daily cost for downtime on 3 machines is 3
    *mu1 + 5*mu2 + 2*mu3

mu = 3*A[1] + 5*B[1] + 2*C[1]
cat("The estimated daily cost is ",mu )

var =  9*A[2]/A[3] + 25*B[2]/B[3] + 4*C[2]/C[3]
cat("Estimated variance is ", var)

# Confidence interval =95%
alpha=1-0.95
z = qnorm(1 - alpha/2)

cat(" We are 95% confident that the mean daily cost
```

```
    of downtimes on these machines is between $",
    mu - z*sqrt(var)," and $",mu + z*sqrt(var))
```

R code Exa 9.14 Normal Distributions with Common Variance

```
# Mean , variance and no. of observations for 2
    batches are as follows:

A <- c(0.22,0.0010,4)
B <- c(0.17,0.0020,5)

diff_mean= A[1]-B[1]

common_var = ((A[3]-1)*A[2] + (B[3]-1)*B[2])/ (A[3]+
    B[3]-2)
sigma = sqrt(common_var)

# Confidence interval =95%
alpha=1-0.95
t = qt(alpha/2, df=A[3]+B[3]-2,lower.tail = F)

c = t*sigma*sqrt(1/A[3]+ 1/B[3])

cat("Thus, we are 95% confident that the difference
    in the mean porosity measurements for two batches
     is between ",
    diff_mean - c," and ", diff_mean +c)
```

R code Exa 9.15 confidence interval for difference in mean denier

```

#Normal Distributions with Common Variance

```

```
M1 <- c
   (9.17,12.85,5.16,6.37,6.64,8.42,7.33,8.91,9.45,11.39,10.90,6.34,1

A <- c(mean(M1),sd(M1),length(M1))

M2 <- c
   (18.86,8.86,17.11,17.38,9.38,11.64,11.25,15.00,12.77,18.89,16.88,

B  <- c(mean(M2),sd(M2),length(M2))

diff_mean= A[1]-B[1]


common_var = ((A[3]-1)*A[2]*A[2] + (B[3]-1)*B[2]*B
   [2])/ (A[3]+B[3]-2)
sigma = sqrt(common_var)


# Confidence interval =95%
alpha=1-0.95
t = qt( alpha/2, df=A[3]+B[3]-2,lower.tail = F)


c = t*sigma*sqrt(1/A[3]+ 1/B[3])

cat("Thus, we are 95% confident that the difference
   in the mean denier is between ",
    diff_mean - c," and ", diff_mean +c)
```

R code Exa 9.16 95 percent confidence level

```

#Normal Distributions with Common Variance
rm("c")
```

```
previous = c(13.18, 9.42, 10.55, 10.11, 7.28, 8.53, 7.52, 8.04, 8.34, 6.91, 10.70, 9.21, 7.84, 9.46, 6.49)


after = c(5.31, 5.77, 3.36, 5.26, 2.43, 6.08, 3.77, 3.20, 3.49, 3.39, 2.99, 4.79, 6.99, 4.81, 3.99, 4.41, 7.12, 3.83, 3.57, 5.41)

t.test(previous,after)
```

R code Exa 9.17 95 percent confidence interval for normal distribution

```
# Mean , variance and no. of observations for 3 batches are as follows:

A <- c(0.22,0.0010,4)
B <- c(0.17,0.0020,5)
C <- c(0.12,0.0018,10)

mu =( A[1]*A[3] + B[1]*B[3])/(A[3]+B[3])

# Difference between average f 2 batches and the 3rd batch
diff_mean = mu - C[1]

diff_sd = sqrt(((A[3]-1)*A[2] + (B[3]-1)*B[2] + (C[3]-1)*C[2])/(A[3]+B[3]+C[3]- 3))

a1 = A[3]/(A[3]+B[3])
a2 = B[3]/(A[3]+B[3])
a3 = -1

#Given , confidence interval =95%
alpha= 1-0.95
```

```
b = qt(alpha/2,df= A[3]+B[3]+C[3]-3,lower.tail = F)*
    diff_sd*sqrt(a1*a1/A[3] + a2*a2/B[3] + a3*a3/C
    [3])

cat(" we are 95% confident that the difference
    between the mean of the third batch and the
    average
    of means of the first two batches is between ",
        diff_mean - b ," — ", diff_mean + b)
```

R code Exa 9.18 Normal Distributions with Unequal Variances

```
#Normal Distributions with Unequal Variances

M1 <- c
    (9.17,12.85,5.16,6.37,6.64,8.42,7.33,8.91,9.45,11.39,10.90,6.34,1


M3 <- c
    (12.17,11.22,11.42,11.73,12.33,12.21,12.21,10.93,12.16,11.61,10.4

t.test(M1,M3)
```

R code Exa 9.19 Two sample T test for Chemical vs Atmospheric

```
#Two sample T test for Chemical vs Atmospheric

chemical <- c
    (2.30143,2.29890,2.29816,2.30182,2.29869,2.29940,2.29849,2.29889,

```

```r

atmp <- c
   (2.31017,2.30986,2.31010,2.31001,2.31010,2.31024,2.31028,2.31163,

t.test(chemical,atmp)
```

R code Exa 9.20 Large Sample Confidence Interval for a Difference in Proportions

```r

#Large Sample Confidence Interval for a Difference
   in Proportions
#Data for motors

n1 = 250
y1 = 25
p1 = y1/n1
n2 = 200
y2 = 30
p2 = y2/n2

diff_prop = p1 - p2

# givem confidence interval =95%
aplha = 1-0.95
a = qnorm(aplha/2, lower.tail = F)*sqrt(p1*(1-p1)/n1
    + p2*(1-p2)/n2)


cat("We are 95% confident that the true difference
   in proportion of defective motors produced by two
    shifts is between " ,diff_prop -a," - ", diff_
   prop+a)
```

R code Exa 9.21 95 percent confidence interval

```
p11=0.7
p12=0.9
p21=0.8
p22=0.9

#Estimated mean
p= (p12 - p11) - (p22 - p21)

#Estimated variance
var= sum(p11*(1-p11),p12*(1-p12),p21*(1-p21),p22*(1-
    p22))/100

# givem confidence interval =95%
aplha = 1-0.95
a = qnorm(aplha/2, lower.tail = F)*sqrt(var)

cat("we are 95% confident that the difference in the
     change in probability for males and females is
    between ",p-a," - ",p+a)
```

R code Exa 9.22 Confidence Interval for a Ratio of Population Variances

```
#Confidence Interval for a Ratio of Population
    Variances

n1 =10
var1 =2.31
n2 =16
var2= 3.68
x = var2/var1
```

```

#given , confidence interval=90%
alpha=1-0.90
cat(" we are 90% confident that the ratio is between ",x/qf(alpha/2,n2 -1,n1 -1,lower.tail = F)," - ", x*qf(alpha/2,n1 -1,n2-1,lower.tail = F))
```

R code Exa 9.23 A Prediction Interval

```
#Prediction interval
mu =16.1
s=0.01
n=16

#Given, confidence interval =95%
alpha=1-0.95
x= qt(alpha/2,df = n-1,lower.tail = F)*s*sqrt(1 + 1/n)
cat("We are about 95% confident that the next observation will lie between ", mu-x ," - ", mu+x)
```

R code Exa 9.24 Tolerance Intervals

```
n=45
mu=498
s=4
delta=0.90
alpha=1-0.95
cat("For these data, k=2.021")
k=2.021
```

```
cat(" We are 95% confident that 90% of the population resistances in the population lie between "
  ,mu- k*s,"-",mu+ k*s)
```

R code Exa 9.25 the confidence coefficient

```

# find the confidence coefficient

delta=0.90
n=50
a= 1 - n*(delta^(n-1)) + (n-1)*(delta^n)
cat("We are ",a*100," confident that the interval (2,150, 2,610) contains at least 90% of the lifelength measurements for the population under study. ")
```

R code Exa 9.29 a 95 percent confidence interval for theeta

```

#Method of maximum liklihood

X <- c( 0.406, 2.343, 0.538, 5.088, 5.587, 2.563, 0.023, 3.334, 3.491, 1.267)
mean = mean(X)*10

#Given confidence interval=95%

alpha =1-0.95
a <- qchisq(alpha/2,df=2*(length(X)),lower.tail = F)
b <- qchisq(alpha/2,df=2*(length(X)))

```

```
cat(" We are about 95% confident that the true mean lifelength is between",2*mean/a, "and", 2*mean/b)
```

R code Exa 9.30 confidence interval

```
data=c(0.406 ,0.685 ,4.778 ,1.725 ,8.223, 2.343 ,1.401 ,1.507 ,0.294, 2.230, 0.538, 0.234 ,4.025 ,3.323, 2.920, 5.088 ,1.458, 1.064, 0.774 ,0.761 ,5.587 ,0.517, 3.246, 2.330 ,1.064 ,2.563 ,0.511 ,2.782 ,6.426 ,0.836 ,0.023 ,0.225, 1.514 ,3.214 ,3.810 ,3.334 ,2.325 ,0.333 ,7.514 ,0.968 ,3.491, 2.921 ,
      1.624, 0.334, 4.490, 1.267, 1.702, 2.634 ,1.849 ,0.186)
x =mean(data)
t=5
n=50
alpha=0.05
z =qnorm(1- alpha/2)
u = exp(-t/x) + (z/sqrt(n))*(t/x)*exp(-t/x)
l = exp(-t/x) - (z/sqrt(n))*(t/x)*exp(-t/x)

cat(" we are about 95% confident that the probability is between ",l,"and",u)
```

Chapter 10

Hypothesis Testing

R code Exa 10.8 Testing for mean

```
#H0 : mu =2
#H1 : mu Not = 2

n= 100
mu =2
sample_mean=2.005
sd = 0.03
alpha =0.05

stat = (sample_mean-mu)/(sd/sqrt(n))

compare = qnorm(alpha/2,lower.tail = F)

if(stat<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is not accepted")
}
error <- qnorm(0.975)*sd/sqrt(n)
cat(" The 95% confidence interval for mu is ",2 -
   error, "- ",2+ error)
```

R code Exa 10.9 Hypothesis testing at 5 percent significance level

```
#H0 : mu<=15
#H1 : mu>15

n= 36
mu =15
sample_mean=17
sd = 3
alpha =0.05

stat = (sample_mean-mu)/(sd/sqrt(n))

compare = qnorm(alpha,lower.tail = F)

if(stat<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.10 Observed Significance Level or p value

```
n= 36
mu =15
sample_mean=17
sd = 3
alpha =0.05

stat = (sample_mean-mu)/(sd/sqrt(n))
p_value= pnorm(stat,lower.tail = F)
```

```
cat("Thus the p−value for this test is ",p_value,"
   and we would reject H0 for any significance level
    greater than or equal to this p−value")
```

R code Exa 10.11 p value for the situation

```
n= 100
mu =2
sample_mean=2.005
sd = 0.03
alpha =0.05

stat = (sample_mean-mu)/(sd/sqrt(n))
p_value= pnorm(stat,lower.tail = F)
cat("The probability of getting a sample of holes
   created with this drill with a mean diameter
    that is at least 1.67 standard deviations away
       from the depth setting of 2 inches is ",2*p_
       value)
```

R code Exa 10.12 hypothesis about the population mean

```
#H0 :  mu<2
#H1 :  mu>2

data=c(0.406 ,0.685 ,4.778 ,1.725 ,8.223, 2.343
   ,1.401 ,1.507 ,0.294, 2.230, 0.538, 0.234 ,4.025
   ,3.323, 2.920, 5.088 ,1.458, 1.064, 0.774 ,0.761
   ,5.587 ,0.517, 3.246, 2.330 ,1.064 ,2.563 ,0.511
   ,2.782 ,6.426 ,0.836 ,0.023 ,0.225, 1.514 ,3.214
   ,3.810 ,3.334 ,2.325 ,0.333 ,7.514 ,0.968 ,3.491,
    2.921 ,
```

```
        1.624, 0.334, 4.490, 1.267, 1.702, 2.634
           ,1.849 ,0.186)

# install the package DescTools

library(DescTools)
p_value =ZTest(data,alternative = "greater",mu=2,sd_
   pop = sd(data),conf.level = 0.95)$p.value
alpha =0.05

if(p_value>alpha){
  cat("Null hypothesis is accepted")
}else{
  cat("Null hypothesis is rejected")
}
```

R code Exa 10.13 the probability of a type II error

```
#H0 : mu<=100
#H1 : mu=103

n= 30
mu =100
sd = 4
alpha =0.01

z = qnorm(alpha,lower.tail = F)
sample_mean = mu + z*sd/sqrt(n)

true_mean =103
s = sd/sqrt(n)

#P(X<= sample_mean)
p_value = pnorm(sample_mean,true_mean,s)
cat("Therefore, the probability of type II error
```

```
   when true avg pressure is 103 =",p_value )
```

R code Exa 10.14 Determining Sample Size

```
mu0 =100
mu1= 103
sd=4
alpha = beta =0.01
n = ((qnorm(alpha)+qnorm(beta))^2)*sd*sd/((mu1 - mu0
   )^2)
cat("By taking ",ceiling(n),"measurements, we can
   reduce to 0.01 while also holding at 0.01. ")
```

R code Exa 10.15 Testing a Mean Normal Distribution Case

```
#H0 : mu=1200
#H1 : mu=not=1200

n= 10
mu =1200
sample_mean=1290
sd = 110
alpha =0.05

t = (sample_mean -mu)/(sd/sqrt(n))

#using rejection region approach
compare = qt(1 - alpha/2,df=n-1,lower.tail = T)

if(t<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
```

```
}
```

R code Exa 10.16 Hypothesis about contradicting the manufacturers claim

```
#H0 : mu=3000
#H1 : mu<3000

n= 8
mu =3000
sample_mean=2959
sd = 39.4
alpha =0.05

t = (sample_mean -mu)/(sd/sqrt(n))

compare = qt(alpha,df=n-1)

if(t>compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.18 Testing for proportion Large sample case

```
#H0 :  p<=0.10
#H1 :  p>0.10

n=100
p_bar=0.15
p0=0.10
alpha=0.01

```

```
z=(p_bar-p0)/sqrt(p0*(1-p0)/n)
compare= qnorm(alpha,lower.tail = F)

if(z<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.19 Testing for variance Normal distribution case

```
#Testing for variance: Normal distribution case

n=10
var=0.0002
sample_var=0.0003
alpha=0.05

K= (n-1)*sample_var/var
compare= qchisq(alpha,df=n-1,lower.tail = F)


if(K<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.20 Testing the difference between two means

```
# H0: mu1-mu2=0
# H1: mu1-m2 not=0

```

```
M =c(42,18,50)
W =c(38,14,50)

diff_mean = M[1]-W[1]
D0=0
alpha=0.05

z=(diff_mean-D0)/sqrt(M[2]/M[3] + W[2]/W[3])

#Using rejection region approach
compare= qnorm(alpha/2,lower.tail = F)

if(z<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.21 Checking the condition of equal variances

```
#H0 : mu1 =mu2
#H1 : mu1>mu2

sm <- c(35.22,24.44,9)
nm <- c(31.56,20.03,9)

diff_mean = sm[1]-nm[1]
D0=0
alpha=0.05

s = sqrt(((sm[3]-1)*sm[2] + (nm[3]-1)*nm[2])/(sm
    [3]-1 + nm[3]-1))
t = (diff_mean - D0)/(s*sqrt(1/sm[3] + 1/nm[3]))

compare= qt(alpha,df=sm[3]-1 + nm[3]-1,lower.tail =
```

```
    F)

if(t<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.22 Checking the condition of equal variances

```
#H0 : mu2=mu3
#H1 : mu2 not= mu3
Class2 = c(253, 261, 258, 255, 256 )
Class3 = c(274, 275, 271, 277, 256 )

p_value= t.test(Class2,Class3)$p.value
if(p_value>alpha){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}

# Mean calculated for class III is incorrect and
    therefore gives the wrong answer.
```

R code Exa 10.23 Testing the Difference between 2 Means Unequal Variances Case

```
#H0 : mu1 - mu2 =D0
#H1 : mu1 - mu2 not=D0

C1 = c( 5.83, 5.66, 4.75, 3.00, 3.37, 3.63, 4.00,
    4.63, 4.25, 4.13 )
```

```
C2 = c( 3.38, 2.81, 7.00, 1.50, 5.88, 5.25, 4.08,
    7.63, 4.50, 4.88 )

p_value= t.test(C1,C2)$p.value
alpha= 0.05
if(p_value>alpha)
{
  cat("Hypothesis accepted")
}else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.24 Testing the Difference between Means for Paired Samples

```
#H0 :muD >=0
#H1 :muD < 0

A = c
    (38.25,31.68,26.24,41.29,44.81,46.37,35.42,38.41,42.68,46.71,29.2

B = c
    (38.25,31.71,26.25,41.33,44.80,46.39,35.46,38.42,42.70,46.76,29.1


p_value= t.test(A,B,paired = TRUE,alternative = "
    less")$p.value

alpha= 0.05
if(p_value>alpha)
{
  cat("Null Hypothesis accepted")
}else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.25 Testing the Difference between Means for Paired Samples

```
#H0 : muD=0
#H1 : muD not=0

E = c( 2727.6, 2902.6, 2463.1 ,3744.5 ,3855.3,
   3807.3, 3610.1 ,3596.3 ,3457.0 ,3507.1, 3184.2,
   3104.7 )
A = c( 2741.0 ,2885.0 ,2476.0 ,3745.0 ,3862.0
   ,3812.0 ,3609.0 ,3568.0 ,3465.0 ,3541.0 ,3213.0,
   3092.0)

alpha= 0.05
p_value=t.test(E,A,paired = TRUE)$p.value
if(p_value>alpha)
{
  cat("Null Hypothesis accepted")
}else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.26 Testing the ratio of variances Normal distributions case

```
#H0 : sigma1^2 = sigma2^2
#H1 : sigma1^2 < sigma2^2


n1= 10
n2= 20
var1=0.003
```

```r
var2=0.001
alpha=0.05

F = var1/var2

#Left-tailed test
compare= qf(1-alpha,n1 -1,n2 -1,lower.tail = T)

if(F<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}


#Alternative solution

p_value = 1-pf(F,n1 -1,n2 -1,lower.tail = T)
if(p_value>alpha)
{
  cat("Null Hypothesis accepted")
}else{
  cat("Hypothesis is rejected")
}

#Note:  t.test function cnnot be used as  numeric vector of data values is not given.
```

R code Exa 10.27 Testing Parameters of the Multinomial Distribution ChiSquare Test

```r
#H0 : p1=4/7,p2=2/7,p3=1/7
#H1 : The proportions differ from those indicated in the null hypothesis.

```

```
X=c(20,16,14)
p=c(4/7,2/7,1/7)
n=50
EX=n*p
alpha=0.05

stat= sum(((X-EX)^2)/EX)
compare= qchisq(1-alpha,df=2,lower.tail = T)


if(stat<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.28 Testing Equality among Binomial Parameters ChiSquare Test

```

#H0 : equal kill rates for the four chemicals
#H1 : at least two mixtures have different kill
   rates.

dead = c(124,147,141,142)
not_dead = c(76,53,59,48)

observed =as.data.frame(rbind(dead,not_dead))
names(observed) <- c('Mix1','Mix2','Mix3','Mix4')

stat= chisq.test(observed)$statistic

alpha=0.05
compare=qchisq(1-alpha,df=3,lower.tail = F)

```

```
if(stat<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}

# *chi-sq value =10.72 given in book is wrong
```

R code Exa 10.29 Test of Independence ChiSq test

```
#H0: The defective/nondefective classification is
    independent of machinist classification
#H1: The defective/nondefective classification
    depends on machinist classification

def= c(10,8,14)
not_def =c(52,60,56)

observed =as.data.frame(rbind(def,not_def))
names(observed) <- c('Machinst A','Machinist B','
    Machinist C')

stat= chisq.test(observed)$statistic

alpha=0.01
dof=(3-1)*(2-1)

#Using rejection region approach
p_value=1 -pchisq(stat,df=2,lower.tail = F)

if(p_value>alpha){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}
```

R code Exa 10.30 ChiSq test

```r
#H0 : Y follows a Poisson distribution
#H1 :  Y does not follow a Poisson distribution

x= rep(0:2, times=c(32,12,6))
table(x)
mean(x)
probs = dpois(0:1, lambda=mean(x))
comp= 1- sum(probs)

stat = chisq.test(x=c(32,12,6), p= c(probs,comp),
   simulate.p.value = TRUE)$statistic

alpha= 0.05
#degree of freedom = (3-1)- 1, as 1 parameter is
   estimated
compare = qchisq(1- alpha,df=1)

if(stat<compare){
  cat("Hypothesis is accepted");
} else{
  cat("Hypothesis is rejected")
}


#Alternative soln..

p_value = 1-pchisq(stat,df=1,lower.tail = T)

if(p_value>0.05)
{
  cat("Hypothesis is accepted")
} else{
```

```
    cat("Hypothesis is rejected")
}

#Both solutions generate same results
```

R code Exa 10.31 Kolmogorov Smirnov test

```
#H0 : F(y) is exponential with theeta=2
#H1 : F(y) is not exponential with theeta=2

y = c
    (0.023,0.406,0.538,1.267,2.343,2.563,3.334,3.491,5.088,5.587)

Fy= 1- exp(-y/2)
n=10
i = 1:10

D_plus = i/n - Fy
D_minus = Fy - (i-1)/n
D = max(max(D_plus),max(D_minus))

# the critical value for a two-sided test with n =
    10 and alpha= 0.05 is 0.409.
D0=0.409
if(D0>D){
    cat("Hypothesis is accepeted")
} else{
    cat("Hypothesis is rejected")
}
```

R code Exa 10.32 Kolmogrov Smirnov Normality Test

```
data =c(0.3780, 0.5090, 0.6230 ,0.6860 ,0.7350 ,
   0.7520, 0.7580, 0.8690, 0.8890, 0.8890, 0.8990,
   0.9370, 0.9820 ,1.0220 ,1.0370 , 1.0880, 1.1230
   ,1.2060, 1.3340 ,1.4230
)

ks.test(data,"pnorm",mean(data),sd(data))

# the answers are different from those given in the
   book.
```

R code Exa 10.33 Kolmogrov Smirnov Normality Test

```
data =c
   (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,34

#Exponential Distribution
ks.test(data,"pnorm",mean(data),sd(data))

#Lognormal Distribution
ks.test(log(data),"pnorm",mean(data),sd(data))

# the answers are different from those given in the
   book.
```

Chapter 11

Inference for Regression Parameters

R code Exa 11.2 SSE for the least squares line

```
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)
n=length(x)

pol <- data.frame(x,y)

line_eq <- lm(y~x,data=pol)

cat("s=",summary(line_eq)$sigma)
```

R code Exa 11.3 95 percent confidence interval for the slope beta1

```
# to find the confidence interval of the given data
    ..

x = c(95,82,90,81,99,100,93,95,93,87)
```

```r
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- data.frame(x,y)

line_eq <- lm(y~x,data=pol)


cat("The 95% confidence interval for beta0(i.e intercept) in the regression line is",confint(line_eq,'x',level=0.95))
```

R code Exa 11.4 Testing the Slope of a Straight Line Model T test

```r
#H0 : beta1 =0
#H1 : beta1 not= 0

#Given, alpha=0.05
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- data.frame(x,y)

line_eq <- summary(lm(y~x,data=pol))
t_value =(coefficients(line_eq)[2,1] - 0)/coefficients(line_eq)[2,2]

cat("Since the T value is",t_value," greater than ",qt(1- 0.025,df=10-2,lower.tail = F)," suggesting strong evidence that beta1 < 1.0")
```

R code Exa 11.5 fitting a line

```
#H0 :(OT contributes no information for the prediction of WR using a simple linear model
#H1 :(OT contributes information for the prediction of WR using a simple linear model

OT =c(-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c(1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
pol <- lm(WR~OT)
summary(pol)

# p_value <0.0001
cat(", we would reject the null hypothesis and conclude that the overtolerance provides information for the prediction of wall reduction")
```

R code Exa 11.6 association between the test strength

```
#H0 : beta1 =1
#H1 : beta1 not=1

NAS <- c(25.1,20.9,25.5,21.3,25.4,20.9,38.5,31.6,63.8,52.6,61.8,51.2)
AS<-  c(35.7,33.6,35.0,33.2,35.7,33.1,52.5,48.7,80.8,75.3,78.3,73.2)
pol <- data.frame(NAS,AS)

line_eq <- summary(lm(AS~NAS,data=pol))
t_value <-(coefficients(line_eq)[2,1] -1)/
```

```
    coefficients(line_eq)[2,2]


cat("Since the T value is",t_value," greater than ",
    qt(1- 0.025,df=12-2,lower.tail = F)," suggesting
    strong evidence that beta1 < 1.0")
```

R code Exa 11.7 tool life and the cutting speeds

```
speed =c
    (340,570,340,570,340,570,340,570,440,440,440,440,305,635,440,440,

life =c
    (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3


pol <- data.frame(speed,life)

l1 <- lm(life~speed,data = pol)
c =coefficients(l1)
cat("Tool Life =",c[1]," + ",c[2],"Cutting Speed")

p_value =summary(l1)$coefficients[,4][2]

cat("Because the p-value <0.001, we reject H0 :beta1
    =0  and conclude that cutting speed contributes
    significant amount of information toward the
    prediction of tool life. ")
```

R code Exa 11.8.a confidence interval for the mean peak power load

```
# To find the confidence interval of the mean
    response..
```

```

x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- data.frame(x,y)

line <- lm(y~x,data = pol)

# here x0 <- 90 F
data <- data.frame(x=90)

c <- predict(line,data, interval = "confidence")

cat(" we can be 95% confident that the mean peak
    power load is between",c[2],c[3],"   megawatts
    for days with a maximum temperature of 90F  ")
```

R code Exa 11.8.b Predict the peak power load for a day

```
# To find the confidence interval of the mean
    response..

x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- data.frame(x,y)

line <- lm(y~x,data = pol)

# here x0 <- 90 F
data <- data.frame(x=90)
res <- predict(line,data, interval = "prediction")

cat(" we are 95% confident that the peak power load
    will be between ",res[2],"-" ,res[3]," megawatts
```

```
on a particular day when the maximum temperature
is 90 F ")
```

R code Exa 11.9.a Estimate the mean wall reduction

```
OT =c
  (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c
  (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50

pol <- data.frame(OT,WR)

line <- lm(WR~OT,data = pol)
data <- data.frame(OT=0.0020)

c <- predict(line,data, interval = "confidence")
cat(" The 95% confidence interval for estimating the
   mean wall reduction at OT = 0.0020 is ",c[2],c
  [3])
```

R code Exa 11.9.b Predict the amount of wall reduction

```
OT =c
  (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
WR =c
  (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50

pol <- data.frame(OT,WR)

```

```
line <- lm(WR~OT,data = pol)
data <- data.frame(OT=0.0020)
res <- predict(line,data, interval = "prediction")

cat(" we are 95% confident that the peak power load
    will be between ",res[2],"-" ,res[3]," megawatts
    on a particular day when the maximum temperature
    is 90 F ")
```

R code Exa 11.10 Polynomial Regression of degree 2

```
#Multiple Regression Analysis

usage = function(size){
  -1216.14 + 2.39893*size - 0.00045*size*size
}

cat("For a house with 2,000 square feet area, the
    predicted electricity usage is ",usage(2000))
```

R code Exa 11.14 Fitting the model The least squares approach

```
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- lm(y~x+I(x^2))
coef = coefficients(pol)

cat("load =", coef[1]," ",coef[2],"temperature +",
    coef[3],"temp^2")

cat("The SSE for this best line of fit is equal to "
    , anova(pol)["Residuals","Sum Sq"])
```

R code Exa 11.15 Estimation of error variance s2

```
#Estimation of error variance sigma^2

x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol <- lm(y~x+I(x^2))

s= summary(pol)$sigma

cat(" the mean square for error , or MSE= ", s*s)
```

R code Exa 11.16 Testing the Utility of a Multiple Regression Model The Global F test

```
#H0 : beta1 = beta2 =0
#H1 : at least one of the coefficients is nonzero

x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)
n=10
k=2
dof=n- (k+1)
pol <- lm(y~x+I(x^2))
F= summary(pol)$fstatistic["value"]
F
compare = qf(1-0.05,k,dof)

if(F < compare)
{
```

```
  cat("Null Hypothesis is accepted")
} else{
  cat("Null Hypothesis is rejected")
}
```

R code Exa 11.17 least square fit of the modal

```
#H0 :beta1 =beta2 =0
#H1 :: At least one is nonzero, i.e., the model is
    useful for predicting Y. The rejection region for
     this test at is .

y =c
    (121,169,172,116,53,177,31,94,72,171,23,177,178,65,146,129,40,167,

x1=c
    (6490,7244,7943,6478,3138,8747,2020,4090,3230,8786,1986,9653,9429

x2=c
    (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,


fit=lm(y~x1+x2)
x=summary(fit)
coef =coefficients(x)

#a
F_value = 112.9
compare =qf(1-0.05,2,37,lower.tail = F)
if(F_value>compare){
  cat("Null hypothesis is accepted")
} else{
  cat("Null hypothesis is rejected")
}

```

```
#b
cat("y =",coef[1]," + ",coef[2],"x1 ",coef[3],"x2")
```

R code Exa 11.18 Estimating and testing hypotheses about beta2

```
#H0 : beta2 =0 , (No quadratic relationship exists
#H1 : beta2 >0 , (The peak power load increases at
    an increasing rate as the daily maximum
    temperature increases.)

x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)
n=10
k=2
dof=n- (k+1)

pol<- data.frame(x,y)
line <- lm(y~x+I(x^2),data = pol)
summary(line)
T =coef(summary(line))[3,3]
compare= qt(1-0.05,df=dof)
if(T < compare)
{
  cat("Null Hypothesis is accepted")
} else{
  cat("Null Hypothesis is rejected")
}
s=coef(summary(line))[3,2]
t =qt(1- 0.05,df=dof)
beta2 =coef(summary(line))[3,1]

cat(" confidence interval for the parameter beta2 as
     follows:",beta2- t*s," - ",beta2+ t*s)
```

R code Exa 11.19 model for mean lost work hours

```
#H0 : beta3 =0
#H1 : beta3 <0

y =c
   (121,169,172,116,53,177,31,94,72,171,23,177,178,65,146,129,40,167

x1=c
   (6490,7244,7943,6478,3138,8747,2020,4090,3230,8786,1986,9653,9429

x2=c
   (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,

fit=lm(y~x1+x2+I(x1*x2))
x=summary(fit)
coef =coefficients(x)
cat("The regression equation is y= ",coef[1]," + ",
   coef[2],"x1 + ",coef[3],"x2 ",coef[4],"x1*x2")

anova(fit)
dof=36
T =coef(summary(fit))[4,3]
compare= qt(0.05,df=dof)
if(T > compare)
{
  cat("Null Hypothesis is accepted")
} else{
  cat("Null Hypothesis is rejected")
}
```

R code Exa 11.20 multiple regression model for estimation and prediction

```r
x = c(95,82,90,81,99,100,93,95,93,87)
y = c(214,152,156,129,254,266,210,204,213,150)

pol<- data.frame(x,y)
line <- lm(y~x+I(x^2),data = pol)
coef = coefficients(line)
y_cap = function(xp){
  coef[1] + coef[2]*xp + coef[3]*xp*xp
}

cat(" the electrical usage for a particular day on
    which the high temperature is 90F, y_cap=", y_cap
    (90))
data <- data.frame(x=90)

d <- predict(line,data, interval = "prediction")

cat("The 95% prediction interval for y0 when x0=90
    is",d[2],"-" ,d[3])

f <- predict(line,data,interval = "confidence")
cat("The 95% prediction interval for y0 when x0=90
    is",f[2],"-" ,f[3])
```

R code Exa 11.21 least squares equation to predict tool life

```r
speed =c
    (340,570,340,570,340,570,340,570,440,440,440,440,305,635,440,440,4

life =c
    (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3

feed =c
```

```
   (0.00630,0.00630,0.01410,0.01416,0.00630,0.00630,0.01416,0.01416,0
        0.00905,0.00905,0.00905,0.00905,0.00472,0.01732,
            0.00905, 0.00905)
depth =c
   (0.02100,0.02100,0.02100,0.02100,0.02100,0.04000,0.04000,0.04000,
    0.02900, 0.02900, 0.02900, 0.02900, 0.02900,
   0.02900, 0.02900, 0.02900,
         0.01350,
            0.04550,0.02900,0.02900,0.02900,0.02900,
             0.01350,0.04550 )

dat= data.frame(life,speed,feed,depth)
fit =lm(life~speed+feed+depth)
#a
cor(dat)

#b
x=summary(fit)
coef= coefficients(x)
cat("the leastsquares equation : Tool life = ",coef
   [1],coef[2],"Speed  ",coef[3],"Feed rate  ",coef
   [4],"Depth of cut")



#c
y =data.frame(depth=0.03,speed=450,feed =0.01)

val= predict(fit,y,interval = "confidence")
cat(" A tool that is used to cut depths of 0.03 inch
    at a speed of 450 fpm with a feed rate of 0.01
   ipr is expected to last on the average ",val[1])
cat(" we are 95% confident that the mean life of
   such a tool used to cut depths of 0.03 inch at a
   speed of 450 fpm with a feed rate of 0.01 ipr
   will be between",val[2],"-",val[3])
```

R code Exa 11.22 A Test for a Portion of a Model

```
# A Test for a Portion of a Model


y=c
   (48.5,55.0,68.0,137.0,309.4,17.5,19.6,24.5,34.8,32.0,28.0,49.9,59

area =c
   (1.1,1.01,1.45,2.4,3.3,0.4,1.28,0.74,0.78,0.97,0.84,1.08,0.99,1.0

bedroom =c
   (3,3,3,3,4,1,3,3,2,3,3,2,2,3,3,3,3,3,3,3,3,3,3,3,3,4,3,2,2,3,3,3,

bathroom =c
   (1,2,2,3,3,1,1,1,1,1,1,2,1,2,2,1,2,2,2,2,2,1,2,2,
   rep(2,59),3,2,2,2,2,2,3,2,3,3)
style =c(0,0,0,1,rep(0,31)
   ,1,1,1,1,0,0,0,0,0,0,0,0,1,1,1,0,0,1,1,0,1,1,1,0,1,0,1,1,0,1,0,1,


# Multiple regression model for selling price of
   houses
fit =lm(y~area+bathroom+style)
summary(fit)

# Full model for selling prices
# Multiple regression model for selling price of
   houses
fit =lm(y~area+bedroom+bathroom+style)
summary(fit)
```

R code Exa 11.23.b Testing a Portion of a Model F test

```
speed =c(340,570,340,570,340,570,340,570,440,440,440,440,305,635,440,440,
life =c(70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3
feed =c(0.00630,0.00630,0.01410,0.01416,0.00630,0.00630,0.01416,0.01416,
        0.00905,0.00905,0.00905,0.00905,0.00472,0.01732,0.00905, 0.00905)
depth =c(0.02100,0.02100,0.02100,0.02100,0.02100,0.04000,0.04000,0.04000, 0.02900, 0.02900, 0.02900, 0.02900, 0.02900, 0.02900, 0.02900, 0.02900,
         0.01350,0.04550,0.02900,0.02900,0.02900,0.02900, 0.01350,0.04550 )

dat= data.frame(life,speed,feed,depth)
fit =lm(life~speed+feed+depth+I(speed*feed)+I(feed*depth)+I(speed*depth)+I(speed*depth*feed))
x=summary(fit)
coef =coefficients(x)
cat("the leastsquares equation : Tool life = ",coef[1],coef[2],"Speed  ",coef[3],"Feed  ",coef[4],"Depth  + ",coef[5],"speed*feed + ",coef[6],
"feed*depth + ",coef[7],"speed*depth ",coef[8],"speed*feed*depth")
```

R code Exa 11.25 Representation of Mean Profit in the Additive Model

```
# Correspondence between Means and Model Parameters
```

```
y =c(0.065,0.073,0.068,0.036,0.078,0.082,0.050,0.043,0.048,0.046,0.06
x1 =c(0,0,0,1,0,0,1,1,0,0,1,1)
x2 =c(0,0,0,0,1,1,1,1,0,0,0,0)
x3 =c(0,0,0,0,0,0,0,0,1,1,1,1)

# a
# Main effects model
fit =glm(y~x1+x2+x3)
coef=coefficients(summary.glm(fit))
cat(" The least-squares prediction equation is yv= ",coef[1],coef[2],"x1 + ",coef[3],"x2 ",coef[4]," x3")

# b
# Complete model including interactions
fit1 =glm(y~x1+x2+x3+(x1*x2)+(x1*x3))
coef1=coefficients(summary.glm(fit1))
cat(". The least-squares prediction equation is ",coef1[1],coef1[2],"x1 +",coef1[3],"x2 ",coef1[4],"x3 ",coef1[5],"x1*x2 + ",coef1[6],"x1*x3")

# c
# H0:The interaction terms do not contribute to the model.(beta4 =beta5 =0)
# H1: At least one of interaction parameters is nonzero.

F_value= 64.04
compare =qf(1- 0.05,2,6)
if(F_value <compare){
  cat("Null hypothesis is accepted")
} else{
  cat("Null hypothesis is rejected")
}
```

R code Exa 11.26 Response surface method

```r
# Response surface method


x1 =c(80,80,80,80,80,80,80,80,80,90,90,90,90,90,90,90,90,90,100,100,10
x2 =c(50,50,50,55,55,55,60,60,60,50,50,50,55,55,55,60,60,60,50,50,50,5
y = c(50.8,50.7,49.4,93.7,90.9,90.9,74.5,73.0,71.2,63.4,61.6,63.4,93.8

dat <- data.frame(x1,x2,y)
model <- lm(y~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data = dat)
c =coefficients(summary(model))
cat(" The least-squares model is as follows:")
cat(c[1][1],"+",c[2][1],"x1 +",c[3][1],"x2 + ",c[4][1],"x1^2 + ",c[5][1],"x2^2 + ",c[6][1],"x1x2")

data <- data.frame(x1=86.25  ,x2= 55.58 )

d <- predict(model,data, interval = "confidence")

cat("The 95% confidence interval for y when x1=86.25 and x2=55.58 is",d[2],"-" ,d[3])
cat("The 95% confidence interval for y when x1=86.25 and x2=55.58 is",d[1])
```

R code Exa 11.27 Modeling a time trend

```
year = c(1970,1975,1980,1981,1982,1983,1984,1985,1986,1987,1988,1989,1990
suv = c(12.3,19.1,23.8,23.7,22.7,23.9,25.6,27.4,29.1,30.6,32.7,33.3,35.6

pol <- data.frame(year,suv)

l1 <- lm(suv~year,data = pol)
r1 = summary(l1)$r.squared
fit <- data.frame(year,log(suv))
l2 = lm(log(suv)~year,data=fit)
r2=summary(l2)$r.squared

cat("Fitting log(FC) as a function of year produces
   a slightly better fit R^2= ",r2*100,"%, than the
   I modal")
```

R code Exa 11.28 Logistic regression

```
#Logistic Fit of Test by Pressure

pressure =c(3943,4163,3812,3888,3926,3900,3942,3732,4480,3940,4143,4146,3962
res =c(1,0,1,1,0,1,1,1,0,0,0,0,0,0,1,1,1,1,0,1,1,0,0,1,0)

```

```
dat=data.frame(pressure,res)
modal =glm(res~pressure,family = "binomial",data =
   dat)

summary(modal)
anova(modal,test = "Chisq")
```

Chapter 12

Analysis of Variance

R code Exa 12.2 Test to Compare k Treatment Means for a Completely Randomized Design

```
#H0 : the mean stopping times at the three types of
   signals are the same i.e mu1 =mu2 =mu3

#H1 : The mean stopping times for at least two types
    of signals are different.


a =c
   (36.6,39.2,30.4,37.1,34.1,17.5,20.6,18.7,25.7,22.0,15.0,10.4,18.9

b =c(rep(1,5),rep(2,5),rep(3,5))
dat =data.frame(a,b)

c <- aov(a~factor(b),data = dat)
summary(c)

#From summary table we obtain data as:
cat("SST = ",1202.6)
cat("SSE = ",137.83)
cat("TSS = ",1202.6+137.83)
```

```
#F value=52.35
compare = qf(1-0.05,2,12)

cat("Since F0.05(2,12) < Fvalue, we  we reject the
   null hypothesis of equal means and conclude that
   at least two types of signals have different mean
    stop times")
```

R code Exa 12.3 mean score for the three groups of managers

```
#H0 : mua =mub =muc
#H1 : At least two group means are different

a =c( 82, 114, 90, 80, 88, 93, 80, 105, 128, 90,
   130, 110, 133, 130, 104, 156, 128, 151, 140 )
b =c(rep(1,8),rep(2,7),rep(3,4))
dat =data.frame(a,b)

x =lm(a~factor(b),data = dat)
summary(x)

c <- aov(a~factor(b),data = dat)
summary(c)

cat("The p-value for the F-test is < 0.0001, which
   means we would reject the null hypothesis of
   equal population mean scores for three groups of
   managers and conclude that at least two groups
   differ in their mean scores. ")
```

R code Exa 12.4 test for mean counts show significant differences

```
#H0 : muAN = muLC = muEC
#H1 : the mean counts differ for the tree metals


a =c
   (10,9,9,9,10,11,14,11,8,11,7,6,9,7,8,10,12,14,9,8,8,9,8,7,10,
      14,12,15,14,10,17,16,11,13,14,15,11,16,12,6,13,20,17,10,16,10,1
      42,46,44,39,50,34,42,40,36,37,46,42,43,50,32,41,37,49,28,34,34,

b =c(rep(1,25),rep(2,25),rep(3,25))

dat =data.frame(a,b)

# the analysis must be done on the transformed data
   because of the lack of homogeneity of variances.

x =lm(sqrt(a)~factor(b),data = dat)
summary(x)
c <- aov(sqrt(a)~factor(b),data = dat)
summary(c)

cat(" the p-value <0.001 so, we reject the null
   hypothesis of equality of mean square root count
   for three metals, and conclude that at least two
   mean square root counts are different. ")
```

R code Exa 12.5 Equivalence between a t test and an F test

```
#H0 : mu1 =mu2
#H1 : mu1 not= mu2

a =c(253,261,258,255,256,264,265,261,257,256)
```

```
b =c(rep(1,5),rep(2,5))
dat =data.frame(a,b)
c <- aov(a~factor(b),data = dat)
summary(c)

t.test(a[1:5],a[6:10],alternative = "two.sided",var.equal = TRUE,mu=0,conf.level = 0.05)

# t=-1.768 and F value = 3.125 which is equal to t^2.

cat("Because p-value = 0.1151 is larger than , we fail to reject the null (using either test) and conclude that there is no evidence of significant difference between the mean UTI for wires provided by the two suppliers")
```

R code Exa 12.6 common variance using a pooled sample variance

```
a =c(36.6,39.2,30.4,37.1,34.1,17.5,20.6,18.7,25.7,22.0,15.0,10.4,18.9
b =c(rep(1,5),rep(2,5),rep(3,5))
dat =data.frame(a,b)

c <- aov(a~factor(b),data = dat)
cat("MSE =" ,anova(c)[["Mean Sq"]][2])
```

R code Exa 12.7 modal for the test score of one manager

```
#  ANOVA and regression analysis

```

```
a =c( 82, 114, 90, 80, 88, 93, 80, 105, 128, 90,
    130, 110, 133, 130, 104, 156, 128, 151, 140 )
b =c(rep(1,8),rep(2,7),rep(3,4))
dat =data.frame(a,b)

x =lm(a~factor(b),data = dat)
summary(x)
anova(x)
```

R code Exa 12.8 Confidence Intervals for Means in the Completely Randomized Design Bonferroni Method

```
a =c( 82, 114, 90, 80, 88, 93, 80, 105, 128, 90,
    130, 110, 133, 130, 104, 156, 128, 151, 140 )
b =c(rep(1,8),rep(2,7),rep(3,4))

dat =data.frame(a,b)

x =lm(a~factor(b),data = dat)
s =summary(x)$sigma

# Bonferroni Method ,c=3
alpha =0.05
c=3
k=(alpha/2)/c
t = qt(1-k,df=16)

#Three intervals are constructed as follows:

a_b = mean(a[1:8]) - mean(a[9:15])
x1= t*s*sqrt(1/8 + 1/7)
cat("Interval muA - muB =", a_b-x1,a_b +x1)

a_c = mean(a[1:8]) - mean(a[16:19])
x2 =t*s*sqrt(1/8 + 1/4)
```

```
cat("Interval muA - muC =", a_c-x2,a_c +x2)

b_c = mean(a[9:15]) - mean(a[16:19])
x3 =t*s*sqrt(1/7 + 1/4)
cat("Interval muB - muC =", b_c-x3,b_c +x3)
```

R code Exa 12.9 confidence intervals for the pairwise difference in mean stop times

```
a =c
   (36.6,39.2,30.4,37.1,34.1,17.5,20.6,18.7,25.7,22.0,15.0,10.4,18.9
b =c(rep(1,5),rep(2,5),rep(3,5))
dat =data.frame(a,b)

x =lm(a~factor(b),data = dat)
s =summary(x)$sigma

# Bonferroni Method ,c=3
alpha =0.05
c=3
k=(alpha/2)/c
t = qt(1-k,df=12)

#a
#Three intervals are constructed as follows:

a_b = mean(a[1:5]) - mean(a[6:10])
x1= t*s*sqrt(1/5 + 1/5)
cat("Interval muPre - muSA =", a_b-x1,a_b +x1)

a_c = mean(a[1:5]) - mean(a[11:15])
x2 =t*s*sqrt(1/5 + 1/5)
cat("Interval muPre - muFA =", a_c-x2,a_c +x2)

```

```r
b_c = mean(a[6:10]) - mean(a[11:15])
x3 =t*s*sqrt(1/5 + 1/5)
cat("Interval muSA - muFA =", b_c-x3,b_c +x3)

#b
y = mean(a[11:15])
d =s/sqrt(5)
cat("a 95% confidence interval for the mean stop time of the best signal i.e FA =",
    y - d," - ",y+d)
```

R code Exa 12.10 Test to Compare k Treatment Means for a Randomized Block Design

```r
c1 =c(5,9)
c2 =c(3,8)
c3 =c(8,13)
c4 =c(4,6)

dat <- rbind(c1,c2,c3,c4)   # combining rows to make matrix..

a <- c(t(as.matrix(dat)))   # concatenate different rows into a vector..

b <- c("b1","b2") # treatment levels

n_tr <- 2  # no. of treatment levels

n_cont <- 4 # no. of control blocks..
block <- gl(n_tr,1,n_cont*n_tr,factor(b))  # vector of treatment factors corresponding to each element of vector a..

chemical <- gl(n_cont,n_tr,n_tr*n_cont) # vector of
```

```
blocking factors corresponding to each element in vector a..

print("The Analysis of Variance table is:")

summary(aov(a~block+chemical)) # anova table display ..

#A
#H0 : mu1 =mu2 =mu3 =mu4
#H1 : the mean resistance differs for at least two treatments.
#F ratio for block is:

F_value_chemical = 12.33
compare = qf(1-0.05,3,3)

if(F_value_chemical < compare){
  cat("Null hypothesis is accepted")
} else{
  cat("Null hypothesis is rejected")
}


#B
#H0 : m1 =mu2
#H1 : there is evidence of significant difference between the block (fabric) means
#F ratio for block is:

F_value_block = 32
compare = qf(1-0.05,1,3)

if(F_value_block < compare){
  cat("Null hypothesis is accepted")
} else{
  cat("Null hypothesis is rejected")
}
```

R code Exa 12.11 difference in mean gains

```
#H0 : muD=0 that is , that there is no difference in
    the mean gain by modified and conventional
    systems
#H1 : muD not=0 , that is , that the mean gain by the
     modified system differs from the mean gain by
    the conventional system

M= c( 1.776, 1.637 ,1.554, 1.460 ,1.405)
C= c(1.901, 1.730 ,1.629 ,1.517 ,1.451)
p_value= t.test(M,C,paired = TRUE,alternative = "two
    .sided")$p.value

if(p_value <0.05){
  cat("Null hypothesis is rejected")
} else{
  cat("Null hypothesis is accepted")
}
```

R code Exa 12.12 ANOVA for RBD and regression analysis

```
c1 =c(5,9)
c2 =c(3,8)
c3 =c(8,13)
c4 =c(4,6)

dat <- rbind(c1,c2,c3,c4)   # combining rows to make
     matrix..

a <- c(t(as.matrix(dat)))   # concatenate different
    rows into a vector..
```

```
b <- c("b1","b2") # treatment levels

n_tr <- 2  # no. of treatment levels

n_cont <- 4 # no. of control blocks..
block <- gl(n_tr,1,n_cont*n_tr,factor(b))  # vector of treatment factors corresponding to each element of vector a..

chemical <- gl(n_cont,n_tr,n_tr*n_cont) # vector of blocking factors corresponding to each element in vector a..

print("The Analysis of Variance table is:")

summary(aov(a~block+chemical)) # anova table display ..

#F ratio for block is:

F_value_chemical = 12.33
```

R code Exa 12.13 Confidence Intervals for Means in the Randomized Block Design

```
c1 =c(5,9)
c2 =c(3,8)
c3 =c(8,13)
c4 =c(4,6)

dat <- rbind(c1,c2,c3,c4)   # combining rows to make matrix..

a <- c(t(as.matrix(dat)))   # concatenate different
```

```
rows into a vector..

b <- c("b1","b2") # treatment levels

n_tr <- 2  # no. of treatment levels

n_cont <- 4 # no. of control blocks..
block <- gl(n_tr,1,n_cont*n_tr,factor(b))  # vector of treatment factors corresponding to each element of vector a..

chemical <- gl(n_cont,n_tr,n_tr*n_cont) # vector of blocking factors corresponding to each element in vector a..

print("The Analysis of Variance table is:")

summary(aov(a~block+chemical)) # anova table display ..

#MSE = 1
s=sqrt(1)


# Bonferroni Method ,c=3
alpha =0.10
c=6
k=(alpha/2)/c
t = qt(1-k,df=3)

x= t*s*sqrt(1/2 + 1/2)
cat("mu1 -mu2 =", mean(c1) -mean(c2) -x,mean(c1) -mean(c2) +x)
cat("mu1 -mu3 =", mean(c1) -mean(c3) -x,mean(c1) -mean(c3) +x)
cat("mu1 -mu4 =", mean(c1) -mean(c4) -x,mean(c1) -mean(c4) +x)
cat("mu2 -mu3 =", mean(c2) -mean(c3) -x,mean(c2) -
```

```
mean(c3) +x)
cat("mu2 -mu4 =", mean(c2) -mean(c4) -x,mean(c2) -mean(c4) +x)
cat("mu3 -mu4 =", mean(c3) -mean(c4) -x,mean(c3) -mean(c4) +x)
#Simialrly, other differences can be computed
```

R code Exa 12.14 Analysis of variance for the factorial experiment

```
y1 =c(9,8,8,7)
y2 =c(5,6,3,4)

observations <- c(y1,y2)
A <- c(rep(1,4),rep(2,4))

a <- c(rep(1,2),rep(2,2))

B <- c(rep(a,2))
dat <- data.frame(observations,A,B)
d <- aov(observations~factor(B)*factor(A),data = dat)

summary(d)
```

R code Exa 12.15 confidence intervals for the six possible differences between treatment means

```
y1 =c(9,8,3,4)
y2 =c(5,6,8,7)

observations <- c(y1,y2)
A <- c(rep(1,4),rep(2,4))

```

```
a <- c(rep(1,2),rep(2,2))

B <- c(rep(a,2))
dat <- data.frame(observations,A,B)
d <- aov(observations~factor(A)*factor(B),data = dat)
summary(d)

# NOTE: Sum Sq value for factor(A)*factor(B)= 24.5 and Mean Sq value for A is 0.5 . (misprinted in textbook)

s=sqrt(anova(d)[["Mean Sq"]][4])
# Bonferroni Method ,c=3
alpha =0.05
c=6
k=(alpha/2)/c
t = qt(1-k,df=4)

x= t*s*sqrt(1/2 + 1/2)
cat(" Thus, all six intervals will be of the form : yi_bar - yj_bar +- ",x)
```

R code Exa 12.16.a Fitting higher order models

```
x1 =c(14.05,14.93,16.56,15.85)
x2 =c(10.55,9.48,13.63,11.75)
x3 =c(7.55,6.59,9.23,8.78)

observations <- c(x1,x2,x3)
A <- c(rep(1,4),rep(2,4),rep(3,4))
a <- c(rep(1,2),rep(2,2))

B <- c(rep(a,3))

```

```r
dat <- data.frame(observations,A,B)

d <- aov(observations~factor(A)*factor(B),data = dat)

summary(d) # analysis of variance table..
```

R code Exa 12.17 Factorial Design

```r
E1 =c(14,10)
E2 =c(14,11)
E3 =c(12,11)
E4 =c(11,12)
W1 =c(4,5)
W2 =c(4,5)
W3 =c(3,6)
W4 =c(5,6)

r=c(E1,E2,E3,E4,W1,W2,W3,W4)
r
f1 = c("X", "Y")          # 1st factor levels
f2 = c("Dry", "Damp")     # 2nd factor levels
k1 = length(f1)           # number of 1st factors
k2 = length(f2)           # number of 2nd factors
n = 4                     # observations per treatment

A = gl(k1, 1, n*k1*k2, factor(f1))
A

B = gl(k2, n*k1, n*k1*k2, factor(f2))
B

av = aov(r ~ A * B)  # include interaction

```

```
summary(av)
```

R code Exa 12.18 effect on defrosted fish by freezing method defrosting method and duration

```
#H0: there is no 3 factor interaction Duration*
    Defrost*Freeze effect on the quality.
#H1: 3 factor interaction Duration*Defrost*Freeze
    affects quality.


A= c(73,70,65,65,68,75,74,69,67,67,74,73)
B= c(75,71,69,70,69,68,76,72,69,72,81,61)
C= c(74,70,70,72,70,70,74,65,65,69,80,74)

obs= c(A,B,C)
freeze= c(rep(1,12),rep(2,12),rep(3,12))
defrost= c(rep(1:3,12))
a= c(rep("1 day",3),rep("8 days",3))

duration= c(rep(a,6))
dat=  data.frame(obs,freeze,defrost,duration)
d=  aov(obs~(factor(freeze)*factor(defrost)*factor(
    duration)),data = dat)

print("The Analysis of Variance Table is shown as
    follows:")

summary(d)

cat("pvalue =0.20507  is higher than any reasonable
    level of significance. Therefore, fail to reject
    the null hypothesis and conclude that there is no
     evidence of the presence of a 3-factor
    interaction. ")
```

```
x = lm(obs~(factor(freeze)*factor(defrost)*factor(duration)),data = dat)
print(”Summary of fit is shown below:”)
summary(x)
```

R code Exa 12.19 yield differ significantly by temperature pressure and reaction time

```
p1 =c(68.5,72.8,72.5,74.5,72.0,75.5,70.5,69.5,65.0)
p2 =c(73.0,80.1,72.5,75.0,81.5,70.0,72.5,84.5,66.5)
p3 =c(68.7,72.0,73.1,74.6,76.0,76.0,74.7,76.0,70.5)

obs =c(p1,p2,p3)
pressure =c(rep(30,9),rep(70,9),rep(100,9))
temp  =c(rep(1:3,9))
a =c(rep(”1 hour”,3),rep(”2 hours”,3),rep(”3 hours”,3))
time =c(rep(a,3))
dat=  data.frame(obs,pressure,temp,time)
d=  aov(obs~(factor(pressure)*factor(temp)*factor(time)),data = dat)

print(”The Analysis of Variance Table is shown as follows:”)

summary(d)
```