R Textbook Companion for Introduction to Probability and Statistics by William Mendenhall, Robert J Beaver, and Barbara M Beaver[1]

Created by
Shivam Sharma
B.Tech.
Information Technology
Inderprastha Engineering College, Ghaziabad
Cross-Checked by
R TBC Team

June 11, 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: Introduction to Probability and Statistics

Author: William Mendenhall, Robert J Beaver, and Barbara M Beaver

Publisher: Brooks Cole, USA

Edition: 13

Year: 2008

ISBN: 9780495389538

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

DESCRIBING DATA WITH GRAPHS

R code Exa 1.3 Pie and Bar Chart

```
rating<-c("A","B","C","D")
frequency<-c(35,260,93,12)
total<-sum(frequency)
relative_frequency_percent<-c(frequency/total)*100
label<-paste(rating,relative_frequency_percent)
label<-paste(label,"%",sep=" ")
par(mfrow=c(1,2))
pie(relative_frequency_percent,labels = label, clockwise = TRUE)
DF<-data.frame(rating,frequency)
barplot(frequency, ylab="frequency",xlab="rating", names.arg = c("A","B","C","D"))
#The answer may slightly vary due to rounding off values
```

R code Exa 1.4 Pareto Chart of Candies Problem

```
color<-c("Brown","Green","Brown","Blue","Red","Red","Green","Brown","Yellow","Orange","Green","Blue","Brown","Blue","Blue","Brown","Orange","Blue","Brown","Orange","Yellow")
color.frequency<-table(color)
barplot(color.frequency[order(color.frequency,decreasing = T)],ylab = "frequency",xlab = "color" )
#please zoom the plot to view the plot clearly
```

R code Exa 1.5 Bar and Pie Chart

```
category<-c("Millitary personnel","Operation and maintenance","Procurement","Research and development","Military construction","Other")
amount<-c(127.5,188.1,82.3,65.7,5.3,5.5)
label<-paste(category,amount)
label<-paste(label,"%",sep="")
par(mfrow=c(1,2))
pie(amount,labels = label,clockwise = TRUE)
barplot(amount, ylab="amount(in Billions)",xlab="category",ylim=c(0,200),names.arg = c("Millitary personnel","Operation and maintenance","Procurement","Research and development","Military construction","Other"))
```

R code Exa 1.6 Line Chart

```
years <- c(2010,2020,2030,2040,2050)
equal_85_and_over <- c(6.1,7.3,9.6,15.4,20.9)
par(mfrow=c(1,2))
plot(years,equal_85_and_over,type = "b")
```

```
plot(years,equal_85_and_over,ylim=c(0,100),type = "b
    ")
```

R code Exa 1.7 Stem and Leaf Plot

```
prices<-c
    (90,65,75,70,70,68,70,70,60,68,70,74,65,75,70,40,70,95,65)
stem(prices, scale = 2)
```

R code Exa 1.8 Stem and Leaf Plot for Weights

```
weights<-c
    (7.2,7.8,6.8,6.2,8.2,8.0,8.2,5.6,8.6,7.1,8.2,7.7,7.5,7.2,7.7,5.8,
stem(weights)
```

R code Exa 1.10 Dotplot

```
GPAs<-c(2.8,3.0,3.0,3.3,2.4,3.4,3.0,0.21)
stripchart(GPAs,method = "stack")
summary(GPAs)
#summary function is used to find the outlier
```

R code Exa 1.11 Relative Frequency Histogram

```r
visits <- c
   (6,4,6,5,3,7,6,5,5,5,1,4,6,5,7,5,6,3,7,5,6,8,4,6,5)
h<-hist(visits, plot=F)
h$counts <- h$counts / sum(h$counts)
plot(h, freq=TRUE, ylab="Relative Frequency")
```

Chapter 2

DESCRIBING DATA WITH NUMERICAL MEASURES

R code Exa 2.1 Dotplot and Sample Mean

```
measurement<-c(2,9,11,5,6)
stripchart(measurement, xlab = "Measurements")
sample_mean<-mean(measurement)
cat("sample mean is",sample_mean)
```

R code Exa 2.2 Median

```
set<-c(2,9,11,5,6)
median<-median(set)
cat("median is",median)
```

R code Exa 2.3 Median for Set

```
set<-c(2,9,11,5,6,27)
```

```
median <- median(set)
cat("Median for the set of measurements is",median)
```

R code Exa 2.5 Variance and Standard Deviation

```
measurement<-c(5,7,1,2,4)
variance <- var(measurement)
standard_deviation <- sd(measurement)
cat("the variance is",variance)
cat("the standard deviation is",standard_deviation)
```

R code Exa 2.6 Tchebysheff theorem

```
x_bar <- 75
variance <- 100
standard_deviation <- sqrt(variance)
lower1 <- x_bar - 2 * standard_deviation
upper1 <- x_bar + 2 * standard_deviation
lower2 <- x_bar - 3 * standard_deviation
upper2 <- x_bar + 3 * standard_deviation
cat("atleast 3/4 of the 25 measurements lie in the
   interval",lower1,"to",upper1,".")
cat("atleast 8/9 of the 25 measurements lie in the
   interval",lower2,"to",upper2,".")
```

R code Exa 2.7 Empirical Rule

```
x_bar <- 12.8
standard_deviation <- 1.7
lower1 <- x_bar - 1 * standard_deviation
```

```
upper1 <- x_bar + 1 * standard_deviation
lower2 <- x_bar - 2 * standard_deviation
upper2 <- x_bar + 2 * standard_deviation
lower3 <- x_bar - 3 * standard_deviation
upper3 <- x_bar + 3 * standard_deviation
cat("approximately 68% ofmeasurements lie in the interval",lower1,"to",upper1,".")
cat("approximately 95% ofmeasurements lie in the interval",lower2,"to",upper2,".")
cat("approximately 99.7% ofmeasurements lie in the interval",lower3,"to",upper3,".")
```

R code Exa 2.9 Range Approximation

```
measurements<-c(5,7,1,2,4)
x<-range(measurements)
range <- x[2]-x[1]
sd <- range /4
cat("standard deviation from range approximation is",sd)
#the value of standard deviation from range approximation is not accurate
```

R code Exa 2.10 Range Approximation

```
data <- c(26.1,26.0,14.5,29.3,19.7,22.1,21.2,26.6,31.9,25.0,15.9,20.8,20.2
x <- range(data)
range <- x[2]-x[1]
sd <- range /4
cat("standard deviation from range approximation is",sd)
```

R code Exa 2.11 Z Score

```
measurement<-c(1,1,0,15,2,3,4,0,1,3)
mean <- mean(measurement)
s <- sd(measurement)
x <- 15
cat("z_score is",(x-mean)/s)
```

R code Exa 2.13 Lower and Upper Quartiles

```
measurement<-c(16,25,4,18,11,13,20,8,11,9)
cat("Lower quantile is",quantile(measurement, 0.25))
cat("Upper quantile is",quantile(measurement, 0.75))
cat("IQR =",IQR(measurement))
summary(measurement)
#the answers provided in the textbook is wrong
```

R code Exa 2.14 Box Plot and Outlier

```
data<-c(340,300,520,340,320,290,260,330)
boxplot(data,horizontal = TRUE)
summary(data)
cat("q1 =",quantile(data,0.25))
cat("m =",quantile(data,0.50))
cat("q3 =",quantile(data,0.75))
IQR(data)
boxplot.stats(data)
#outlier is 520
#the answers provided in the textbook is wrong
```

Chapter 3

DESCRIBING BIVARIATE DATA

R code Exa 3.1 Side by Side Bar Chart

```
data<- matrix(c(94.8,65.9,56.4,118.1,76.0,65.1),ncol
   =3,byrow=TRUE)
colnames(data) <- c("Full Professor","Associate
   Professor","Assistant Professor")
rownames(data) <- c("public","private")
data <- as.table(data)
data
barplot(data, ylab="Average Salary ($ Thousands)",
   ylim=c(0,120), col=c("blue","grey"),
        legend = c("public","private"), beside=TRUE)
```

R code Exa 3.2 Comparative Charts

```
public<-c(24,57,69)
private<-c(60,78,112)
private_percent<-(private*100)/250
```

```
public_percent<-(public*100)/150
private_percent
label<-paste(public_percent,"%",sep=" ")
label1<-paste(private_percent,"%",sep=" ")
par(mfrow = c(1,2))
color<-c("white","aliceblue","cadetblue1")
pie(private_percent,labels = label1,clockwise = TRUE
    ,main="Private",col = color)
legend("bottomleft",c("Full Professor","Assosiate
    Professor","Assistant Professor"),cex=0.35,fill =
     color)
pie(public_percent,labels = label,clockwise = TRUE,
   main="Public",col = color)
legend("bottomleft",c("Full Professor","Assosiate
    Professor","Assistant Professor"),cex=0.35,fill =
     color)
cat("proportion of assistant professor is roughly
   same for both private and public colleges")
cat("public colleges have smaller proportion of full
     professors and a large proportion of associate
    professors")
```

R code Exa 3.3 Scatterplot

```
x<-c(2,2,3,4,1,5)
y<-c(95.75,110.19,118.33,150.92,85.86,180.62)
plot(x,y)
```

R code Exa 3.4 Scatterplot for Data

```
#install package("ggplot2")
#install library("ggplot2")
cases<-c(23,21,19,18,15,17,19,20,25,24)
```

```
price<-c(10,10,11,11,12,12,13,13,14,14)
DF<-data.frame(price,cases)
library(ggplot2)
#read library("ggplot2)
ggplot(DF,aes(price,cases)) + geom_point()
cat("there exists Linear relationship ")
```

R code Exa 3.5 Scatter Plot for Size of Living Area and Selling Price

```
x<-c
   (1360,1940,1750,1550,1790,1750,2230,1600,1450,1870,2210,1480)
y<-c
   (278.5,375.7,339.5,329.8,295.6,310.3,460.5,305.2,288.6,365.7,425.
plot(x,y)
cat("Plot represents the linear pattern in data")
```

R code Exa 3.6 Correlation Coefficient

```
x<-c
   (1360,1940,1750,1550,1790,1750,2230,1600,1450,1870,2210,1480)
y<-c
   (278.5,375.7,339.5,329.8,295.6,310.3,460.5,305.2,288.6,365.7,425.
correlation_Coefficient <- round(cor(x, y),4)
cat("correlation coefficient of x and y is",
   correlation_Coefficient)
```

R code Exa 3.7 Line and Dataplot

```
#install package("ggplot2")
#install library("ggplot2")
x<-c(2,3,4,5,6,7)
y<-c(6.00,7.50,8.00,12.00,13.00,15.50)
install.packages("ggplot2")
DF<-data.frame(x,y)
library(ggplot2)
#read library("ggplot2")
ggplot(DF,aes(x,y))+geom_point()+geom_smooth(method=
    "lm" , se= F)
```

Chapter 4

PROBABILITY AND PROBABILITY DISTRIBUTIONS

R code Exa 4.5 Probabilities Using Simple Events

```
E2 <- 1/4;
E3 <- 1/4;
cat("Probability of observing exactly one head in
   two tosses is",E2+E3)
```

R code Exa 4.6 Probabilities of Blood Phenotypes

```
A <- 0.41;
B <- 0.10;
AB <- 0.04;
o <- 0.45;
cat("Probability that person is either type A or
   type AB is",A+AB)
```

R code Exa 4.7 Probability of Candies Problem

```
R1R2 <- 1/6;
R2R1 <- 1/6;
cat("Probability that both candies are red is",R1R2 + R2R1)
```

R code Exa 4.8 Probability Using Counting Rules

```
m <- 6;
n <- 6;
cat("Total number of simple events in the sample space S are",m*n)
```

R code Exa 4.9 Candy Dish

```
first_candy <- 3;
second_candy <- 2;
cat("Simple events in the sample space S are",first_candy * second_candy)
```

R code Exa 4.10 Simple Events of Coins

```
coin_ways <- 2
cat("Simple events in the sample space when three coins are tossed are",coin_ways * coin_ways * coin_ways)
```

R code Exa 4.11 Routes

```
Routes_A_B <- 3;
Routes_B_C <- 4;
Routes_C_D <- 3;
cat("Possible A to D routes are",Routes_A_B * Routes
    _B_C * Routes_C_D)
```

R code Exa 4.12 Simple Events Using Permutation

```
simple_events <- factorial(50)/factorial(50-3);
cat("Total Simple events are",simple_events)
```

R code Exa 4.13 Permutaion of Tests

```
Total_tests <- factorial(5)/factorial(5-5);
cat("Total number of tests are",Total_tests)
```

R code Exa 4.14 Counting Rule for Combinations

```
suppliers <- 5
choose <- 3
cat("Total number of ways in which three suppliers
    are to be choosen from five are",choose(suppliers
    , choose))
```

R code Exa 4.15 Manufacturers Problem

```
Total_ways <- 10;
Two_out_of_best <- choose(3, 2);
one_out_of_not_best <- choose(2, 1);
cat("Probability of selecting exactly two of best three are",(Two_out_of_best * one_out_of_not_best) / Total_ways)
```

R code Exa 4.17 Oil Prospecting Firm Problem

```
Prob_A <- 0.80;
Prob_B <- 0.18;
Prob_c <- 0.02;
cat("probability of A or B",Prob_A + Prob_B)
cat("probability of B or C",Prob_B + Prob_c)
```

R code Exa 4.18 College Education Expenses Problem

```
Child_Too_High <- 0.35;
Child_Right_Amount <- 0.08;
Child_Too_Little <- 0.01;
No_Child_Too_High <- 0.25;
No_Child_Right_Amount <- 0.20;
No_Child_Too_Little <- 0.11;
too_high <- 0.60;
right_ammount <- 0.28;
too_little <- 0.12;
child_college <- Child_Too_High + Child_Too_Little + Child_Right_Amount;
```

```
cat("probability that respondent has a child in
   college is",child_college)
cat("probability that respondent does not have a
   child in college is",1-child_college)
cat("probability that respondent has child in
   college and with too high load is",too_high +
   child_college - Child_Too_High)
```

R code Exa 4.19 Multiplication Rule for Probability

```
prob_r <- 2/8;
prob_g <- 6/8;
prob_rr <- 1/7;
prob_rg <- 6/7;
prob_gr <- 2/7;
prob_gg <- 5/7;
cat("probability that child choose the two red toys
   is",prob_r * prob_rr)
```

R code Exa 4.20 Independent Events

```
s <- c("hh","ht","th","tt")
prob_a <- 1/2
prob_b <- 1/2
prob_a_and_b <- 1/4
if((prob_a * prob_b) == prob_a_and_b){
  print("events must be independent")
}else{
    print("events are not independent")
  }
```

R code Exa 4.21 Multiplication Rule for Independent Events

```r
Child_Too <- 0.35;
Child_Right_Amount <- 0.08;
Child_Too_Little <- 0.01;
No_Child_Too_High <- 0.25;
No_Child_Right_Amount <- 0.20;
No_Child_Too_Little <- 0.11;
too_high <- 0.60;
right_amount <- 0.28;
too_little <- 0.12;
child_college <- 0.44;
No_child_college <- 0.56;
x = too_high * child_college
cat("x and child_college values are not same so both events are dependent ")
```

R code Exa 4.22 Cards Problem

```r
ace_on_first <- 4/52;
ten_on_second_when_ace_on_first <- 4/51;
ten_on_first <- 4/52;
ace_on_Second <- 4/51;
x <- ace_on_first * ten_on_second_when_ace_on_first
y <- ten_on_first * ace_on_Second
cat("probability of ace on first and ten on second draw",x)
cat("probability of ten on first and ace on second draw",y)
cat("probability that the draw includes an ace and a ten is",x+y)
```

R code Exa 4.23 Law of Total Probability

```r
pg1 <- 0.09
pg2 <- 0.20
pg3 <- 0.31
pg4 <- 0.23
pg5 <- 0.17
p_a_g1 <- 0.26
p_a_g2 <- 0.20
p_a_g3 <- 0.13
p_a_g4 <- 0.18
p_a_g5 <- 0.14
pA <- (pg1 * p_a_g1) + (pg2 * p_a_g2) +(pg3 * p_a_g3
    ) +(pg4 * p_a_g4) +(pg5 * p_a_g5)
cat("the required probability is",pA)
```

R code Exa 4.24 Sneakers Problem

```r
pg1 <- 0.09
pg2 <- 0.20
pg3 <- 0.31
pg4 <- 0.23
pg5 <- 0.17
p_a_g1 <- 0.26
p_a_g2 <- 0.20
p_a_g3 <- 0.13
p_a_g4 <- 0.18
p_a_g5 <- 0.14
required_probability <- (pg5 * p_a_g5) / ((pg1 * p_a
    _g1) + (pg2 * p_a_g2) +(pg3 * p_a_g3) +(pg4 * p_a
    _g4) +(pg5 * p_a_g5))
cat("the required probability is",required_
    probability)
```

R code Exa 4.26 Mean Variance and Standard Deviation

```
x <- c(0,1,2,3,4,5)
prob_x <- c(0.10,0.40,0.20,0.15,0.10,0.05)
k<- c(x*prob_x)
mean <- weighted.mean(x, prob_x)
l <- c((x-mean)*(x-mean))
m <- c(l*prob_x)
variance <- sum(m)
standard_deviation=round(sqrt(variance),2)
cat("mean is",mean)
cat("variance is",variance)
cat("standard deviation is",standard_deviation)
```

R code Exa 4.27 Expected Gain

```
gain<-c(-20,23980)
prob_gain<-c((7998/8000),(2/8000))
prob_gain
expected_gain <- weighted.mean(gain, prob_gain)
#expected_gain is in dollar
cat("expected gain per lottery would be a loss of",
   expected_gain)
```

Chapter 5

SEVERAL USEFUL DISCRETE DISTRIBUTIONS

R code Exa 5.3 Binomial Probability Distribution

```
x <- 2;
n<- 10;
p <- 0.1;
prob <- round((dbinom(x,n,p)),4)
cat("the required probability is",prob)
```

R code Exa 5.4 Probabilities for free throws

```
total_throws <- 4;
prob <- 0.8;
x <- 2;
y <- 0;
case_one <- dbinom(x,total_throws,prob)
cat("probability that he will make exactly two free
    throws is",case_one)
case_two <- 1-dbinom(y,total_throws,prob)
```

```
cat("probability that he will make atleast one free
    throw is",case_two)
```

R code Exa 5.5 Probabilities for Successes

```
total <- 5;
prob <- 0.6;
x <- 3;
y <- 2
case_one <- pbinom(x,total,prob)-pbinom(y,total,prob
    )
cat("probability of exactly three successes is",case
    _one)
case_two <- 1-pbinom(y,total,prob)
cat("probability of three or ore successes is",case_
    two)
#the answer may slightly vary due to rounding off
    values
```

R code Exa 5.6 Probability of Survivors

```
total <- 10;
prob <- 0.5;
x <- 7;
eight_or_more <- 1 - pbinom(x,10,0.5)
cat("probability of exactly three success is",eight_
    or_more)
```

R code Exa 5.7 Multiple Choice Test

```
prob_correct <- 0.2
prob_incorrect <- 1 - prob_correct
n <- 100
mu0 <- n * prob_correct
sigma <- sqrt(n * prob_correct * prob_incorrect)
cat("a large proportion of score will lie within two
    standard deviations of the mean, or from",(mu0 -
    2 * sigma),"to",(mu0 + 2 * sigma),"." )
cat("allmost all the score will lie within three
   standard deviations of the mean, or from",(mu0 -
   3 * sigma),"to",(mu0 + 3 * sigma),"." )
cat("guessing will be better than zero score but the
    student will not pass the exam")
```

R code Exa 5.8 Traffic Accidents using Poisson Distribution

```
mean <- 2;
accidents_in_case_one <- 0;
case_one <- round((dpois(accidents_in_case_one,mean)
   ),6)
cat("probability of no accident on this section of
   highway during a 1-week period is",case_one)
case_two_mean <- 2*mean;
case_two_mean
case_two <- round((dpois(0,case_two_mean)+ dpois(1,
   case_two_mean) + dpois(2,case_two_mean) + dpois
   (3,case_two_mean)),6)
cat("probability of atmost three accidetns on this
   section of highway during a 2-week period is",
   case_two)
```

R code Exa 5.9 Life Insurance Company

```
total_men <- 5000;
prob <- 0.001;
claims <- 4;
mean <- total_men * prob
exact_prob <- round((dpois(claims,mean)),3)
cat("probability that the company will have to pay 4 claims during a given year is",exact_prob)
```

R code Exa 5.10 Lawn Mowers Problem

```
total <- 1000;
prob <- 0.001;
mean <- total * prob;
none_defective <- dpois(0,mean)
cat("probability that none is defective is",none_defective)
three_defective <- dpois(3,mean)
cat("probability that three is defective",three_defective)
four_defective <- dpois(4,mean)
cat("probability that four are defective",four_defective)
#the asnwer may slightly vary due to rounding off
```

R code Exa 5.11 Probability Distribution Mean Variance

```
total_bottles <- 12;
spoiled_wine <- 3;
sample <- 4;
prob_zero <- (choose(spoiled_wine,0) * choose(total_bottles - spoiled_wine,(sample - 0)))/choose(total_bottles,sample)
```

```
cat("probability distribution of no bottle of spoiled wine is",prob_zero)
prob_one <- (choose(spoiled_wine,1) * choose(total_bottles - spoiled_wine,(sample - 1)))/choose(total_bottles,sample)
cat("probability distribution of  one bottle of spoiled wine in sample is",prob_one)
prob_two <- (choose(spoiled_wine,2) * choose(total_bottles - spoiled_wine,(sample - 2)))/choose(total_bottles,sample)
cat("probability distribution of  two bottle of spoiled wine in sample is",prob_two)
prob_three <- (choose(spoiled_wine,3) * choose(total_bottles - spoiled_wine,(sample - 3)))/choose(total_bottles,sample)
cat("probability distribution of  three bottle of spoiled wine in sample is",prob_three)
mean = sample * (spoiled_wine/total_bottles)
cat("mean is",mean)
variance <-  sample * (spoiled_wine/total_bottles)*(9/total_bottles)*((total_bottles-sample)/11)
cat("variance is",variance)
#"The answer may slightly vary due to  rounding off values"
```

R code Exa 5.12 Hypergeometric Probability Distribution

```
total_items <- 20;
sample_items <- 5;
defective <- 4;
prob_accept_lot <- choose(defective,0) * choose(total_items - defective,(sample_items - 0))/choose(total_items,sample_items) + choose(defective,1) * choose(total_items - defective,(sample_items - 1))/choose(total_items,sample_
```

```
    items)
cat("probability of get accepted is",prob_accept_lot
    )
```

Chapter 6

THE NORMAL PROBABILITY DISTRIBUTION

R code Exa 6.1 Probability

```
given_value <- 0.2
required_probability <- punif(given_value, max =
   0.5, min = -0.5) - punif(- given_value, max =
   0.5, min = -0.5)
cat("required probability is",required_probability)
```

R code Exa 6.2 Probability of waiting time

```
average_waiting_time <- 5
waiting_time <- 10
required_probability <-  1 - pexp(waiting_time, rate
    = 1 / average_waiting_time)
cat("the required probability is",required_
   probability)
```

R code Exa 6.3 Normal Probability Distribution

```
point <- 1.63;
prob <- round((pnorm(point)),4);
cat("the required probability is",prob)
```

R code Exa 6.4 Normal Probability Distribution

```
point <- -0.5;
prob <- 1-pnorm(point);
cat("the required probability is",round(prob,4))
```

R code Exa 6.5 Normal Probability Distribution

```
point_one <- -0.5;
point_two <- 1.0;
prob_one <- pnorm(point_one);
prob_two <- pnorm(point_two);
prob <- round((prob_two - prob_one),4);
cat("The required probability is",prob)
```

R code Exa 6.6 Normal Probability Distribution

```
point_one <- 1;
point_two <- 2;
prob_one <- round(pnorm(point_one) - pnorm(-point_
    one),5);
```

```
prob_two <- round(pnorm(point_two) - pnorm(-point_two),4);
cat(" probability that a normally distributed random variable will fall within One standard deviation of its mean",prob_one)
cat(" probability that a normally distributed random variable will fall within Two standard deviations of its mean",prob_two)
#the answers may slightly vary due to rounding off values
```

R code Exa 6.7 z value

```
shaded_area <- 0.95
A1 <- (1 - shaded_area) / 2
cumulative_Area_left_of_z0 <- shaded_area + A1
z_value <- round(qnorm(cumulative_Area_left_of_z0),2)
cat("the required z value is",z_value)
```

R code Exa 6.8 Normal Distributed Random Variable

```
mean <- 10;
standard_deviation <- 2;
lower_X <- 11;
upper_x <- 13.6;
prob <- round((pnorm(upper_x, mean, standard_deviation) - pnorm(lower_X, mean, standard_deviation)),4);
cat("the required probability is",prob)
```

R code Exa 6.9 Gasoline use for Compact car

```
mean <- 25.5;
standard_deviation <- 4.5;
value_x <- 30;
prob <- 1 - pnorm(value_x, mean, standard_deviation)
    ;
percentage <- prob *100;
cat("the required percentage is",round(percentage,2)
    )
```

R code Exa 6.10 Gasoline Use Rate

```
mean <- 25.5;
standard_deviation <- 4.5;
percent <- 95 / 100;
cat("The gasoline use rate for the new car must be",
    qnorm(percent, mean, standard_deviation))
```

R code Exa 6.11 Normal Approximation to Binomial Distribution

```
total <- 25;
prob <- 0.5;
actual_prob <- round((pbinom(10,total,prob)-pbinom
    (7,total,prob)),4)
mean <- total * prob;
standard_deviation <- sqrt(total*prob*0.5)
x_lower <- 7.5;
x_upper <- 10.5;
approx_prob <- round(pnorm(x_upper, mean, standard_
    deviation) - pnorm(x_lower, mean, standard_
    deviation) ,4)
cat("actual probability is",actual_prob)
```

```
cat("approximate probability is",approx_prob)
cat("approximate and actual probability are quite
    close")
```

R code Exa 6.12 Electric Fuse

```
x_value <- 26.5;
total <- 1000;
defective <- 0.02;
reliability <- 0.98;
mean = total * defective;
standard_deviation <-round( sqrt((total * defective)
     * reliability),2)
approx_prob <- 1 - round(pnorm(x_value, mean,
    standard_deviation),4)
cat("Approximate probablity of observing 27 or more
    defective is",approx_prob)
#the answers may slightly vary due to rounding off
    values
```

R code Exa 6.13 Soft Drinks Brand

```
consumers <- 2500;
brand_share <- 10;
prob_correct <- brand_share/100;
prob_not_correct <- 1 - prob_correct;
mean = consumers * prob_correct;
standard_deviation <- sqrt(consumers * prob_correct
    * prob_not_correct)
x_value <- 211.5;
required_prob <- round((pnorm(x_value, mean,
    standard_deviation)),4)
```

```
cat("probability of observing 211 or fewer consumers
    who prefer her band of soft  drink is",required_
  prob)
```

Chapter 7

SAMPLING DISTRIBUTIONS

R code Exa 7.1 Simple Random Sample

```
N <- 1000
n <- 5
sample(N, n, replace = FALSE, prob = NULL)
```

R code Exa 7.4 Alzheimer disease

```
sample_size <- 30;
x_value <- 7;
mean <- 8;
sd <- 4;
prob_less_7 <- round(pnorm(x_value, mean, round((sd
   / sqrt(sample_size)),2)),6)
cat("The approximate probability duration for
   average duration is less than 7 years is",prob_
   less_7)
prob_exceed_7 <- round((1 - prob_less_7),5)
cat("The approximate probability for average duation
    exceeds 7 years is",prob_exceed_7)
```

```
x1_value <- 9;
prob_of_interest <- (pnorm(x1_value, mean, round((sd / sqrt(sample_size)),2))) - (pnorm(x_value, mean, round((sd / sqrt(sample_size)),2)))
cat("The approximate probability for average duration lies within 1 year of the population mean=8 is",prob_of_interest)
#"The answers may sligthly vary due to rounding off values"
```

R code Exa 7.5 Sample Distribution of Sample Mean

```
no_of_bottles <- 10;
mean <- 12.1;
x_value <- 12;
standard_Deviation <- 0.2;
standard_error <- standard_Deviation/sqrt(no_of_bottles)
required_prob <- round((pnorm(x_value, mean, standard_error)),4)
cat("The required probability is",required_prob)
```

R code Exa 7.6 Sample Distribution of Sample Proportion

```
sample_space <- 500;
mean <- 0.60;
standard_error <- sqrt((mean * (1-mean))/sample_space)
standard_error <- round(standard_error,3)
cat("standard error is",standard_error)
p <- 2 * standard_error
cat("required value is",p)
```

R code Exa 7.7 Probability for Sample Proportion

```
# textbook refered to example 7.6 for values
sample_space <- 500;
proportion <- 0.55;
observed_value <- 0.60;
standard_error <- round(sqrt((proportion * (1-
   proportion))/sample_space),4)
cat("standard error is",standard_error)
req_prob <- round(1 - pnorm(observed_value,
   proportion, standard_error),4)
cat("probobality of observing a sample proportion as
    large as or larger than the observed value is",
   req_prob)
```

R code Exa 7.8 Xbar Chart

```
#install package ("qicharts")
#install library ("qicharts")
install.packages("qicharts")
library(qicharts)
n <- 4
bearing1 <- c
   (0.992,1.015,0.988,0.996,1.015,1.000,0.989,0.994,1.018,0.997,1.02
bearing2 <- c
   (1.007,0.984,0.993,1.020,1.006,0.982,1.009,1.010,1.016,1.005,0.98
bearing3 <- c
   (1.016,0.976,1.011,1.004,1.002,1.005,1.019,1.009,0.990,0.989,1.00
```

```
bearing4 <- c
   (0.991,1.000,0.981,0.999,1.001,0.989,0.994,0.990,1.011,1.001,0.98

sample <- c(1:25)
measurement <- c(bearing1 + bearing2 + bearing3 +
   bearing4)
sample_mean <- c((bearing1 + bearing2 + bearing3 +
   bearing4) / n)
df <- data.frame(sample, bearing1, bearing2,
   bearing3, bearing4,sample_mean)
df
qic(sample_mean, x = sample, data = df, chart = 'i',
   main = 'Xbar chart of Diameter', xlab = 'sample',
    ylab = 'sample mean')
#The results may slightly vary due to rounding off
   values
```

R code Exa 7.9 Control Chart

```
#install package ("qcc")
#install library ("qcc")
install.packages("qcc")
library(qcc)
proportion <- c
   (0.0200,0.0125,0.0225,0.0100,0.0150,0.0200,0.0275,0.0175,0.0200,0

n <- 400
defects <- c(proportion * n)
qcc(defects, type = "p",sizes = n, xlab = "day",
   ylab = "proportion")
```

Chapter 8

LARGE SAMPLE ESTIMATION

R code Exa 8.4 Point Estimation of Population Parameter

```
sample <- 50;
standard_deviation <- 105;
margin_of_error <- 1.96 * (standard_deviation/sqrt(
    sample))
margin_of_error
cat("the average weight of all arctic polar bears
    are within more or less of 29 pounds of sample
    estimate of 980 pounds")
```

R code Exa 8.5 Estimation of True Population

```
sample <- 100;
p <- 0.73;
q <- 1 - p;
k <- sqrt((p * q)/sample);
margin_of_error <- round((1.96 * k),2)
```

```
cat("margin of error is",margin_of_error)
```

R code Exa 8.6 Construction of Confidence Interval

```
sample_mean <- 756;
sample_size <- 50;
standard_deviation <- 35;
standard_error <- qnorm(0.975) *( standard_deviation / sqrt(sample_size))
left <- round((sample_mean - standard_error),2);
right <- round((sample_mean + standard_error),2);
cat("95% confidence interval for sample mean is from ",left, "to" ,right,"grams")
```

R code Exa 8.7 99 percent Confidence Interval

```
sample_mean <- 756;
sample_size <- 50;
standard_deviation <- 35;
z <- round((qnorm(0.995,lower.tail = T)),2);     #99% confidence interval -> 1 - 0.005 = 0.995
standard_error <- z *( standard_deviation / sqrt(sample_size))
left <- sample_mean - standard_error;
right <- sample_mean + standard_error;
cat("99% confidence interval is from ",left, "to" , right, "grams per day")
```

R code Exa 8.8 Large Sample Confidence Interval

```r
sample_size <- 985;
vote_for_republican <- 592;
point_estimate <- vote_for_republican/sample_size;
standard_error <- sqrt((point_estimate * (1 - point_estimate))/sample_size)
standard_error
point_estimate
coin_interval <- 90/100;
z <- qt((1+coin_interval)/2,df=sample_size-1)
value <- z * standard_error
left <- round((point_estimate - value),3);
right <- round((point_estimate + value),3);
cat("90% confidence interval is from ",left, "to" , right)
```

R code Exa 8.9 Difference Between Two Sample Mean

```r
miles_cover_by_type1 <- 26400;
miles_cover_by_type2 <- 25100;
type1_sample <- 100;
type2_sample <- 100;
variance1 <- 1440000;
variance2 <- 1960000;
point_estimate <- miles_cover_by_type1 - miles_cover_by_type2;
standard_error <- sqrt(((variance1)/type1_sample) + (variance2)/type2_sample);
confidence_interval_percent <- 0.99;
z_value <- round((qnorm(0.995)),2)
value <- z_value * standard_error;
left <- point_estimate - value;
right <- point_estimate + value;
cat("The difference in the average miles to wearout for the two types of tires is estimated to lie between ",left,"and",right,"miles of wear")
```

R code Exa 8.10 Confidence Interval for Difference

```
sample_size_men <- 50;
sample_size_women <- 50;
sample_mean_men <- 756;
sample_mean_women <- 762;
sample_standard_Deviation_men <- 35;
sample_standard_Deviation_women <- 30;
z_value <- 1.96
#we know from z table that for 95% confidence interval the z value is 1.96;
point_estimate <- sample_mean_men -sample_mean_women;
standard_error <- sqrt(((sample_standard_Deviation_men * sample_standard_Deviation_men)/sample_size_men + (sample_standard_Deviation_women * sample_standard_Deviation_women)/sample_size_women))
value <- z_value * standard_error;
left <- round((point_estimate - value),2);
right <- round((point_estimate + value),2);
cat("The 95% confidence interval is from",left,"to",right)
```

R code Exa 8.11 Bond proposal problem

```
p1_cap <- 0.76
p2_cap <- 0.65
q1_cap <- 0.24
q2_cap <- 0.35
n1 <- 50
n2 <- 100
```

```
p1_cap - p2_cap
standard_error <- round(sqrt(((p1_cap * q1_cap) / n1
    ) + ((p2_cap * q2_cap) / n2 )),4)
standard_error
alpha <- 0.01        #at 99% confidence interval
    alpha = 0.01
z_0.005 <- qnorm( 1 - alpha/2)
lower_value <- (p1_cap - p2_cap) - z_0.005 *
    standard_error
upper_value <- (p1_cap - p2_cap) + z_0.005 *
    standard_error
cat("the interval is from",lower_value,"to",upper_
    value)
n <- 150
point_estimation <- 103/n
margin_of_error <- 1.96 * sqrt((point_estimation *
    (1 - point_estimation) / n))
margin_of_error1 <- - 1.96 * sqrt((point_estimation
    * (1 - point_estimation) / n))
cat("margin of error is",margin_of_error,"and",
    margin_of_error1,".")
cat("interval is from",point_estimation - margin_of_
    error,"to",point_estimation + margin_of_error,"."
    )
```

R code Exa 8.12 95 percent upper confidence bound

```
alpha <- 0.05  #at 95% confidence interval the alpha
     = 0.05
x_bar <- 10.3
n <- 40
s <- 0.31
se <- s / sqrt(n)
z_0.05 <- qnorm(1 - alpha)
z_0.05
```

```
ucb <- x_bar + z_0.05 * se
ucb
cat("the 95% upper confidence bound is",ucb)
```

R code Exa 8.13 Plastic Pipe

```
bound_b <- 0.04
alpha <- 0.10    #at 0.90 confidence coefficient
z_0.05 <- qnorm(1 - alpha/2)
z_0.05
p <- q <- 0.5
n <- ((z_0.05 * 0.5) / bound_b) ** 2
cat("the producer must include atleast",round(n,0),"
   wholesalers in  survey")
```

R code Exa 8.14 Workers in Training groups

```
range <- 8
sigma1 <- sigma2 <- sigma <- range / 4
alpha <- 1 - 0.95  # at 0.95 confidence coefficient
z_0.05 <- qnorm(1 - alpha/2)
z_0.05
n <- (z_0.05 * sqrt(8)) ** 2
cat("there should be atleast",round(n,0),"workers in
    each group.")
```

Chapter 9

LARGE SAMPLE TESTS OF HYPOTHESES

R code Exa 9.3 Test Statistic

```
mean <- 14
s <- 2
n <- 100
se <- s / sqrt(n)
x_bar <- 15
z <- (x_bar - mean) / se
z
p_value <- 2 * pnorm(-abs(z))
cat("p-value is approximately zero")
```

R code Exa 9.4 Appropriate Hypothesis

```
xbar <- 725;              # sample mean
mu0 <-  670;                # hypothesized value
sigma <- 102;             #  standard deviation
sample_size <- 40;                  # sample size
```

```
z <- (xbar - mu0)/(sigma/sqrt(sample_size))
z                          # test statistic
alpha <- .01
z.alpha <- qnorm(1 - alpha)
z.alpha                    # critical value
cat("The value of z is",z)
cat("critical value is",z.alpha)
cat("null hypothesis rejected ")
cat("average weekly earning for male are higher than
    average for female")
#the answer may slightly vary due to rounding off
   values
```

R code Exa 9.5 Appropriate Hypothesis

```
sample_mean <- 871;              # sample mean
hypothesized_value <-  880;              #
   hypothesized value
sigma <- 21;               #  standard deviation
sample_size <- 50;                   # sample size
z <- round(((sample_mean - hypothesized_value)/(
   sigma/sqrt(sample_size))),2)
z                          # test statistic
alpha <- .05
z.alpha <- round((qnorm(1 - alpha/2)),2)
z.alpha                    # critical value
cat("The value of z is",z)
cat("critical value is",z.alpha)
cat("null hypothesis can be rejected")
cat("she is reasonably confident that the decision
   is correct")
```

R code Exa 9.6 Calculation P Value

```r
sample_mean <- 871;                 # sample mean
hypothesized_value <-  880;              #
    hypothesized value
sigma <- 21;           #  standard deviation
sample_size <- 50;                   # sample size
z_value <- (sample_mean - hypothesized_value)/(sigma
    /sqrt(sample_size))
z_value                         #test statistics
p_value <- round((1 - pnorm(-z_value)) + pnorm(z_
    value),4)
cat("p value is",p_value)
cat("reject null hypothesis  at either 5% or 1%
    level of significance")
```

R code Exa 9.7 Daily Sodium Intake

```r
sample_mean <- 3400;                 # sample mean
hypothesized_value <-  3300;              #
    hypothesized value
sigma <- 1100;           #  standard deviation
sample_size <- 100;                   # sample size
z_value <- (sample_mean - hypothesized_value)/(sigma
    /sqrt(sample_size))
z_value                         # test statistic
alpha <- 0.05
z.alpha <- qnorm(1 - alpha)
z.alpha                  # critical value
cat("The value of z is",z_value)
cat("critical value is",z.alpha)
p_value <- 1 - pnorm(z_value)
cat("p value is",p_value)
cat("null hyppothesis is not rejected")
cat("not enough evidence ")
#the answer may slightly vary due to rounding off
    values
```

R code Exa 9.8 Beta and power of beta

```
sample_mean <- 871;                    # sample mean
hypothesized_value <-  880;                    #
    hypothesized value
sigma <- 21;              #  standard deviation
sample_size <- 50;                      # sample size
alpha <- .05
z.alpha <- round((qnorm(1 - alpha/2)),2)
z.alpha                   # critical value
lower <- round(hypothesized_value - z.alpha * ((
    sigma/sqrt(sample_size))),2)
lower
upper <- round(hypothesized_value + z.alpha * ((
    sigma/sqrt(sample_size))),2)
upper
mu <- 870
z1 <- round(((lower - mu)/(sigma/sqrt(sample_size)))
    ,2)
z1
z2 <- round(((upper - mu)/(sigma/sqrt(sample_size)))
    ,2)
z2
beta <- round( 1 - pnorm( z1),4)
power_of_test <- 1 - beta
cat("beta is",beta)
cat("power of test is",power_of_test)
```

R code Exa 9.9 Car Ownership Affect

```
non_owners <- 100;
```

```
owners <- 100;
average_non_owner <- 2.70;
average_owner <- 2.54;
variance_non_owner <- 0.36;
variance_owner <- 0.40;
d_nod <- 0;
point_estimate <-  average_non_owner - average_owner
    ;
standard_error <- sqrt(((variance_non_owner)/non_
    owners) + (variance_owner)/owners);
z = round(((point_estimate - d_nod)/standard_error)
    ,2);
cat("value of z is",z)
alpha <- .05
z.alpha <- round((qnorm(1 - alpha/2)),2)
z.alpha                     # critical value
cat("The value of z is",z)
cat("critical value is",z.alpha)
p_value <- round(((1-pnorm(z)) + pnorm(-z)),4)
cat("p value is",p_value)
cat("null hypothesis cannot  be rejected and there
    is insufficient evidence")
```

R code Exa 9.10 Hypothesis Testing and Confidence Interval

```
# In textbook values are refered to example 9.9
non_owners <- 100;
owners <- 100;
average_non_owner <- 2.70;
average_owner <- 2.54;
variance_non_owner <- 0.36;
variance_owner <- 0.40;
d_nod <- 0;
point_estimate <-  average_non_owner - average_owner
    ;
```

```
standard_error <- sqrt(((variance_non_owner)/non_
    owners) + (variance_owner)/owners);
confidence_interval_percent <- 0.95;
z_value <- qnorm(1-(1 - confidence_interval_percent)
    /2)
value = z_value * standard_error;
left = round((point_estimate - value),2);
right = round((point_estimate + value),2);
cat("The confidence interval is  approximated from",
    left,"to",right)
cat("there is not enough evidence so we cannot
    conclude")
```

R code Exa 9.11 Hypothesis Testing and P Value

```
hypothesized_value <- 0.2;
observed_value <- 15/100;
standard_error <- sqrt((hypothesized_value * (1-
    hypothesized_value))/100)
z_value <- (observed_value - hypothesized_value)/
    standard_error;
p_value <- pnorm(z_value)
cat("pvalue is",p_value)
cat("null hypothesis cannot be rejected so there is
    insufficient evidence to make conclusion")
```

R code Exa 9.12 Hypothesis Test for Difference Between Two Binomial Proportions

```
admitted_men <- 52;
admitted_women <- 23;
sample_men <- 1000;
sample_women <- 1000;
```

```
p1_value <- admitted_men/sample_men;
p2_value <- admitted_women/sample_women;
pooled_estimate <- (admitted_men + admitted_women)/(sample_men + sample_women);
standard_error <- sqrt(pooled_estimate * (1 - pooled_estimate) * ((1/1000) + (1/1000)))
test_statistics <- (p1_value - p2_value)/standard_error
test_statistics
alpha <- 0.05;
k <- abs(qnorm(alpha))
lower_bound <- (p1_value  - p2_value) - k * sqrt((p1_value * (1 - p1_value)/sample_men) + (p2_value * (1 - p2_value)/sample_women))
cat("Lowest likely value for the difference is", lower_bound)
cat("the data present sufficient evidence to indicate that the percentage of men entering the hospital because of heart disease is higher than women")
```

Chapter 10

INFERENCE FROM SMALL SAMPLES

R code Exa 10.2 Students T Distribution

```
sample_size <- 10;
t <- round((qt(p = 0.01, df = sample_size - 1 ,lower
    .tail = T)),3)
cat("value of t is",t)
```

R code Exa 10.3 Average Weight of Diamonds

```
weights <- c(0.46,0.61,0.52,0.48,0.57,0.54)
t.test(weights, mu = 0.5, alternative = "greater",
    conf.level = 0.95)
cat("the range of possible values of mean is both
    greater and smaller than 0.5 so there is no
    sufficient evidence and our test fails.")
```

R code Exa 10.4 P value of paint probem

```
data <- c(310,311,412,368,447,376,303,410,365,350)
mean <- mean(data)
s <- sd(data)
mu0 <- 400
t.test(data,mu = mu0)
p.value <- t.test(data,mu = mu0)$p.value
p.value
cat(p.value < 0.05)
cat("Since p-value is less than 0.05 so null
   hypothesis is rejected and there is sufficient
   evidence to indiacte the coverage differs from
   400")
```

R code Exa 10.5 Hypothesis Testing and T Value for Student grades

```
online <- c(32,37,35,28,41,44,35,31,34)
classroom <- c(35,31,29,25,34,40,27,32,31)
alpha <- 0.05
df <- length(online) + length(classroom) - 2
stem(online)
stem(classroom)
cat("stem and leaf plot of the data show at least a
   mounding pattern so the assumption of normality
   is not unreasonable.")
critical_value <- round((qt(p = 1 -alpha,df,lower.
   tail = T)),3)
cat("critical value is",critical_value)
t.test(online,classroom,alternative = "greater")
t <- t.test(online,classroom,alternative = "greater"
   )$statistic
if(t > critical_value){
  print("reject the null hypothesis")
}else{
```

```
print(”cannot reject the null hypothesis so there is
     insufficient evidence to indicate that the
    online course grades are higher than the
    conventional course grades at the 5 % level of
    significance.”)
  }
```

R code Exa 10.6 P Value For Student Grades

```
#values are referred to example 10.5 in textbook
online <-c (32,37,35,28,41,44,35,31,34)
classroom <-c (35,31,29,25,34,40,27,32,31)
p_value <- t.test(online,classroom,alternative = ”
    greater”)$p.value
cat(”p−value =”,p_value,”is geater than 0.05,most
    researchers would report the result as not
    significant.”)
```

R code Exa 10.7 Lower confidence bound

```
online <- c(32,37,35,28,41,44,35,31,34)
classroom <- c(35,31,29,25,34,40,27,32,31)
t.test(online,classroom,alternative = ”greater”)
lower <- t.test(online,classroom,alternative = ”
    greater”)$conf.int[1]
cat(”lower confidence bound is”,lower)
cat(”since the difference of equal means is included
     in the confidence interval so it is possible
    that two means are equal so there is insufficient
     evidence to indicate that the online average is
    higher than the classroom average.”)
#the results may slightly vary due to rounding off
    values in textbook.
```

R code Exa 10.8 Tire problem

```
tire_A <- c(10.6,9.8,12.3,9.7,8.8)
tire_B <- c(10.2,9.4,11.8,9.1,8.3)
alpha <- 0.05
critical_value <- qt( 1 - alpha/2,df = 4)
k <- t.test(tire_A,tire_B, paired = TRUE,
    alternative = "two.sided")
test_statistics <- k[1]
cat(test_statistics > critical_value)
cat("test statistics from t.test() is larger than
    critical value hence there is difference in the
    mean wear for tire type A and B")
```

R code Exa 10.9 95 percent confidence interval

```
tire_A <- c(10.6,9.8,12.3,9.7,8.8)
tire_B <- c(10.2,9.4,11.8,9.1,8.3)
t.test(tire_A,tire_B, paired = TRUE, alternative = "
    two.sided")
```

R code Exa 10.11 Cement manufacturer

```
sigma_square <- 100
n <- 10
xbar <- 312
s_square <- 195
x_square <- ((n -1) * s_square)/sigma_square
alpha <- 0.05
```

```
df <- n -1
x_square_0.05 <- qchisq(1 - alpha, df)
if(x_square > x_square_0.05){
  print("reject the null hypothesis and the range of concrete strength measurements exceeds the manufacturer's claim")
}else{
  print("accept the null hypothesis")
}
```

R code Exa 10.12 90 percent confidence interval

```
sigma_square <- 4
n <- 3
df <- n - 1
measurements <- c(4.1,5.2,10.2)
s_square <- var(measurements)
s_square
x_square <- ((n - 1) * s_square) / sigma_square
cat("p-value is greater than 0.10 so accept the null hypothesis")
alpha <- 0.10      #at 90% confidence interval alpha = 0.10
x_square_0.95 <- qchisq(alpha/2,df)
x_square_0.05 <- qchisq(1-alpha/2,df)
x_square_0.05
lower_value <- round((n - 1) * s_square/x_square_0.05,2)
upper_value <- round((n - 1) * s_square/x_square_0.95,2)
cat("the interval is from",lower_value,"to",upper_value)
```

R code Exa 10.13 F Value

```
f1 <- round(qf(0.95,6,9),2)          #1 - 0.05 = 0.95
f2 <- round(qf(0.95,5,10),2)
f3 <- round(qf(0.99,6,9),2)          #1 - 0.01 = 0.99
cat("value of f in case 1 is",f1)
cat("value of f in case 2 is",f2)
cat("value of f in case 3 is",f3)
```

R code Exa 10.14 Hypothesis Test For Equality of two Population Variances

```
variance1 <- 7.14;
variance2 <- 3.21;
n1 <- 10;
n2 <- 8;
df1 <- n1 - 1
df2 <- n2 - 1
alpha <- 0.05
rejection_region <- round(qf(1 - (alpha / 2), df1,
   df2),2)
rejection_region
test_statistics <- variance1/variance2;
cat("the calculated value of test statistics is",
   test_statistics)
if(test_statistics > rejection_region){
  print("reject the null hypothesis")
}else{
  print("Cannot reject null hypothesis and there is
     sufficient evidence to indicate a difference in
      the population variance")
}
```

R code Exa 10.15 Confidence Interval Estimate

```
variance1 <- 7.14;
variance2 <- 3.21;
n1 <- 10;
n2 <- 8;
df1 <- n1 - 1
df2 <- n2 - 1
alpha <- 0.05
f97 <- round(qf( 0.95, df1=9, df2=7),2)
f79 <- round(qf(.95, df1=7, df2=9),2)
lower_value <- round((variance1 / variance2) * (1 / f97),2)
upper_value <- round((variance1 / variance2) * (f79),2)
cat("the interval is between ",lower_value,"and", upper_value)
```

R code Exa 10.16 Impurities in the batch of chemical

```
x1bar <- 302
x2bar <- 3.0
s1_square <- 1.04
s2_square <- 0.51
n1 <- n2 <- 25
f <- (s1_square / s2_square)
df1 <- df2 <- n1 - 1
f_0.050 <- qf(1 - 0.050,df1,df2)
f_0.025 <- qf(1 - 0.025,df1,df2)
if(f > f_0.050 & f < f_0.025){
  print("p-value lies between 0.025 and 0.05 and hence at 5% level null hypothesis is rejected")
}else{
  print("null hypothesis is accepted")
}
```

Chapter 11

THE ANALYSIS OF VARIANCE

R code Exa 11.4 Analysis of Variance Table

```
no_breakfast<-c(8,7,9,13,10)
light_breakfast<-c(14,16,12,17,11)
full_breakfast<-c(10,12,16,15,12)

k <- 3;
n1 <- n2 <- n3 <- 5
n <- 15;


sum_square_X <- sum(no_breakfast) + sum(light_
   breakfast) + sum(full_breakfast)
cm <- (sum_square_X * sum_square_X)/n
cm
x <- sum(no_breakfast)
y <- sum(light_breakfast)
z <- sum(full_breakfast)
total_ss <- (8*8 + 7*7 + 9*9 + 13*13 + 10*10 + 14*14
    + 16*16 + 12*12 + 17*17 + 11*11 + 10*10 + 12*12
   + 16*16 + 15*15 + 12*12 ) - cm
```

```
total_ss
degree_of_freedom <- n-1
degree_of_freedom
sst <- ((x*x + y*y + z*z)/5) - cm
sst
df <- k - 1
sse  <- total_ss -sst
sse
deg_of_freedom <- n - k
deg_of_freedom

combined<-data.frame(cbind(no_breakfast,light_
   breakfast,full_breakfast))
stacked<-stack(combined)
stacked
Anova_Results<-aov(values ~ ind,data = stacked)
summary(Anova_Results)
```

R code Exa 11.5 F test

```
no_breakfast<-c(8,7,9,13,10)
light_breakfast<-c(14,16,12,17,11)
full_breakfast<-c(10,12,16,15,12)

k <- 3;
n1 <- n2 <- n3 <- 5
n <- 15;


sum_square_X <- sum(no_breakfast) + sum(light_
   breakfast) + sum(full_breakfast)
cm <- (sum_square_X * sum_square_X)/n
cm
x <- sum(no_breakfast)
y <- sum(light_breakfast)
```

```
z <- sum(full_breakfast)
total_ss <- (8*8 + 7*7 + 9*9 + 13*13 + 10*10 + 14*14
     + 16*16 + 12*12 + 17*17 + 11*11 + 10*10 + 12*12
    + 16*16 + 15*15 + 12*12 ) - cm
total_ss
degree_of_freedom <- n-1
degree_of_freedom
sst <- ((x*x + y*y + z*z)/5) - cm
sst
df1 <- k - 1
df1
sse  <- total_ss -sst
sse
df2 <- n - k
df2

combined<-data.frame(cbind(no_breakfast,light_
    breakfast,full_breakfast))
stacked<-stack(combined)
stacked
Anova_Results<-aov(values ~ ind,data = stacked)
summary(Anova_Results)
mse <- sse / (n - k)
mse
mst <- sst / (k - 1)
mst
f <- mst / mse
f
alpha <- 0.05
f_0.05 <- qf(1 - alpha, df1, df2)
if(f > f_0.05){
  print("reject h0, there is sufficient evidence to
      indicate that at least one of the three average
       attention spans is different from at least one
       of the others")
}
```

R code Exa 11.6 CONFIDENCE INTERVALS FOR A SINGLE TREATMENT MEAN AND THE DIFFERENCE BETWEEN TWO TREATMENT MEANS

```r
#values are referred to example 11_4in the textbook
no_breakfast<-c(8,7,9,13,10)
light_breakfast<-c(14,16,12,17,11)
full_breakfast<-c(10,12,16,15,12)
x1_bar <- mean(no_breakfast)
x2_bar <- mean(light_breakfast)
x3_bar <- mean(full_breakfast)

k <- 3;
n1 <- n2 <- n3 <- 5
n <- 15;


total_ss <- (8*8 + 7*7 + 9*9 + 13*13 + 10*10 + 14*14
    + 16*16 + 12*12 + 17*17 + 11*11 + 10*10 + 12*12
   + 16*16 + 15*15 + 12*12 ) - cm
total_ss
degree_of_freedom <- n-k
degree_of_freedom
sst <- ((x*x + y*y + z*z)/5) - cm
sst
sse  <- total_ss -sst
sse
mse <- sse/(n-k)
s2 <- mse
s <- sqrt(s2)
alpha <- 0.05/2;
t_value <- qt(1-alpha,degree_of_freedom)
t_value
left1 <- round(x1_bar - (t_value * (s/sqrt(n1))),2)
```

```
right1 <- round(x1_bar + (t_value * (s/sqrt(n1))),2)
cat("confidence interval for no breakfast are between",left1,"and",right1,"minutes")
left2 <- round((x2_bar - x3_bar) - t_value*(sqrt(s2*((1/n2) + (1/n3)))),2)
right2 <- round((x2_bar - x3_bar) + t_value*(sqrt(s2*((1/n2) + (1/n3)))),2)
cat("confidence interval for the difference in the average attention spans for light versus full breakfast eaters are between",left2,"and",right2,"minutes")
```

R code Exa 11.7 Tukey method for paired comparison

```
no_breakfast<-c(8,7,9,13,10)
light_breakfast<-c(14,16,12,17,11)
full_breakfast<-c(10,12,16,15,12)

k <- 3;
n1 <- n2 <- n3 <- 5
n <- 15;

sum_square_X <- sum(no_breakfast) + sum(light_breakfast) + sum(full_breakfast)
cm <- (sum_square_X * sum_square_X)/n
cm
x <- sum(no_breakfast)
y <- sum(light_breakfast)
z <- sum(full_breakfast)
total_ss <- (8*8 + 7*7 + 9*9 + 13*13 + 10*10 + 14*14 + 16*16 + 12*12 + 17*17 + 11*11 + 10*10 + 12*12 + 16*16 + 15*15 + 12*12 ) - cm
total_ss
sst <- ((x*x + y*y + z*z)/5) - cm
sst
```

```
sse  <- total_ss -sst
sse
df <- n - k
df
mse <- sse / (n - k)
mse
mst <- sst / (k - 1)
mst
f <- mst / mse
f
s <- sqrt(mse)
s
alpha <- 0.05
w <- (qtukey(1 - alpha, k, df)) * (s / sqrt(n1))
w
cat("the difference between no break fast and light
   breakfas exceeds w = ",w,"  so no breakfast and
   light breakfast are declared significantly
   different ")
```

R code Exa 11.8 Randomized Block Design

```
b <- 3
k <- 4
low <- c(27,24,31,23)
middle <- c(68,76,65,67)
high <- c(308,326,312,300)
df <- data.frame(low, middle, high)
data <-c((as.matrix(df)))
data
f <- c("a","b","c","d")
company <- gl(k,1, b*k,factor(f))
usage <- gl(b, k, k * b)
model <- aov(data ~ usage + company)
model
```

```
summary(model)
```

R code Exa 11.9 Evidence Indication

```
b <- 3
k <- 4
low <- c(27,24,31,23)
middle <- c(68,76,65,67)
high <- c(308,326,312,300)
df <- data.frame(low, middle, high)
data <-c((as.matrix(df)))
data
f <- c("a","b","c","d")
company <- gl(k,1, b*k,factor(f))
usage <- gl(b, k, k * b)
model <- aov(data ~ usage + company)
model
summary(model)
p.value <- summary(model)[[1]][["Pr(>F)"]][[2]]
p.value
cat("since the p-value from anova test p.value=",p.value,"is too large to allow rejection of null hypothesis.")
cat("hence there is insufficient evidence to indicate a difference in the averge monthly costs for the four companies.")
```

R code Exa 11.10 cell phone cost

```
A <- c(27,68,308)
B <- c(24,76,326)
C <- c(31,65,312)
D <- c(23,67,300)
```

```
t1 <- sum(A)
t2 <- sum(B)
t3 <- sum(C)
t4 <- sum(D)
b1 <- 105
b2 <- 276
b3 <- 1246
k <- 4;
n1 <- n2 <- n3 <- n4 <- 3
n <- 12;
b <- 3

sum_square_X <- sum(A) + sum(B) + sum(C) + sum(D)
sum_square_X
cm <- (sum_square_X * sum_square_X)/n
cm

l <- c(A,B,C,D)
l
sum_of_all <-sum( c(l * l))
sum_of_all
total_ss <- sum_of_all - cm
total_ss
sst <- (( t1 * t1 + t2 * t2 + t3 * t3 + t4 * t4)/ n1
   )- cm
sst
ssb <- ((b1 * b1 + b2 * b2 + b3 * b3) / k ) - cm
ssb
sse  <- total_ss - sst - ssb
sse
t2_bar <- t2 / n2
t3_bar <- t3 / n3
mse <- sse / ((b - 1) * (k - 1))
mse
alpha <- 0.05
df <- (b - 1)  * (k - 1)
lower_range <- round((t2_bar - t3_bar) - qt(1 -
   alpha/2,df) * sqrt(mse * (2 / b)),2)
```

```
upper_range <-round((t2_bar - t3_bar) + qt(1 - alpha/2,df) * sqrt(mse * (2 / b)),2)
cat("the difference between two average costs is estimated as betwen",lower_range,"and",upper_range,".")
```

R code Exa 11.12 Two Way ANOVA

```
observation <- c(571,480,470,610,474,430,625,540,450,480,625,630,516,600,680,465,
supervisor <- c(1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2)
shift <- c("day","swing","night","day","swing","night","day","swing","night","day","swing","night","day","swing","night","day","swing","night")
model <- aov(observation ~ supervisor * shift)
model
summary(model)
```

Chapter 12

LINEAR REGRESSION AND CORRELATION

R code Exa 12.1 Least Squares Prediction Line

```
x <- c(39,43,21,64,57,47,28,75,34,52)
y <- c(65,78,52,82,92,89,73,98,56,75)
df <- data.frame(x, y)
lm(y ~ x, df)
cat("least square regression line is: ycap = 40.7842
    + 0.7656x")
```

R code Exa 12.2 Hypothesis Test for Linear Relationship

```
x <- c(39,43,21,64,57,47,28,75,34,52)
y <- c(65,78,52,82,92,89,73,98,56,75)
x_square <- c(x**2)
y_square <- c(y**2)
xy <- c(x*y)
n <- 10;
sum_x <- sum(x)
```

```
sum_y <- sum(y)
sum_x_square <- sum(x_square)
sum_y_square <- sum(y_square)
sum_xy <- sum(xy)
s_xx <- sum_x_square - (sum_x * sum_x)/n
s_yy <- sum_y_square - (sum_y * sum_y)/n
s_xy <- sum_xy - (sum_x *sum_y)/n
y_bar <- sum(y)/n
x_bar <- sum(x)/n
b <- (s_xy/s_xx)
a <- y_bar - b*x_bar
a
b
total_ss <- s_yy
ssr <- (s_xy * s_xy)/s_xx
sse <- total_ss - ssr
mse <- sse / (n - 2)
mse
sse
test_statistics <- round(((b - 0)/sqrt(mse / s_xx))
   ,2)
cat("test statistics is",test_statistics)
alpha <- 0.05              # 5% significance level
range <- round(qt(1 - alpha/2, df = 8),3)
range
cat("rejection region is greater then",range,"or","
   less than",-range)
cat("there is significant linear relationship")
```

R code Exa 12.3 Confidence Interval Estimate

```
x <- c(39,43,21,64,57,47,28,75,34,52)
y <- c(65,78,52,82,92,89,73,98,56,75)
df <- data.frame(x, y)
cat("interval is from",left,"to",right)
```

```
m <- lm(y ~ x, df)
m
confint(m,level = 0.95)
print(paste0("The 95% confidence interval is ",
    sprintf("[ %4.3f, %4.3f ]", confint(m,level =
    0.95)[2],confint(m,level = 0.95)[4])));
#the answers may slightly vary due to rounding off
    values.
```

R code Exa 12.4 Average Calculus Grade

```
x <- c(39,43,21,64,57,47,28,75,34,52)
y <- c(65,78,52,82,92,89,73,98,56,75)
x0 <- 50
df <- data.frame(x, y)
lmresult <- lm(y ~ x, data = df)
lmresult
df2 <- data.frame(x = 50)
df2
ycap <- predict(lmresult, df2, interval = "
    confidence")[1]
ycap
lower <- predict(lmresult, df2, interval = "
    confidence")[2]
upper <- predict(lmresult, df2, interval = "
    confidence")[3]
cat("the average calculus grade for students who
    score 50 on the achievement test will lie between
    ",lower,"and",upper)
```

R code Exa 12.5 Student Achievement test

```
x <- c(39,43,21,64,57,47,28,75,34,52)
```

```
y <- c(65,78,52,82,92,89,73,98,56,75)
x0 <- 50
df <- data.frame(x, y)
lmresult <- lm(y ~ x, data = df)
lmresult
df2 <- data.frame(x =  50)
df2
ycap <- predict(lmresult, df2, interval = "
    prediction")[1]
ycap
lower <- predict(lmresult, df2, interval = "
    prediction")[2]
upper <- predict(lmresult, df2, interval = "
    prediction")[3]
cat("the 95% confidence interval is from",lower,"to"
    ,upper)
```

R code Exa 12.6 Grade Achievement test

```
x <- c(39,43,21,64,57,47,28,75,34,52)
y <- c(65,78,52,82,92,89,73,98,56,75)
x0 <- 0
df <- data.frame(x, y)
lmresult <- lm(y ~ x, data = df)
lmresult
df2 <- data.frame(x = 0)
df2
ycap <- predict(lmresult, df2, interval = "
    confidence")[1]
ycap
lower <- predict(lmresult, df2, interval = "
    confidence")[2]
upper <- predict(lmresult, df2, interval = "
    confidence")[3]
cat("the 95% confidence interval is from",lower,"to"
```

```
,upper)
cat(", this interval does not contain the value alpha = 0 hence y intercept cannot be 0.")
cat("data does not support the hypothesis of 0 intercept")
```

R code Exa 12.7 Correlation Coefficient

```
x <- c(73,71,75,72,72,75,67,69,71,69)
y <- c(185,175,200,210,190,195,150,170,180,175)
r <- cor(x, y)
cat("correlation coefficient is",r)
```

R code Exa 12.8 Correlation Coefficient

```
#refer to example 12.7 for height and weight values
height <- c(73,71,75,72,72,75,67,69,71,69)
weight <- c(185,175,200,210,190,195,150,170,180,175)
cor.test(height, weight, method = "pearson", alternative = "two.sided")
print(cor.test(height, weight, method = "pearson", alternative = "two.sided")[3])
cat("correlation is declared significant at 1% level")
```

Chapter 13

Multiple Regression Analysis

R code Exa 13.2 Multiple Regression Analysis

```
y <- c(169.0,218.5,216.5,225.5,229.9,235.0,239.9,247.9,260.0,269.9,234.
x1 <- c(6,10,10,11,13,13,13,17,19,18,13,18,17,20,21)
x2 <- c(1,1,1,1,1,2,1,2,2,1,1,1,2,2,2)
x3 <- c(2,2,3,3,3,3,3,3,3,3,4,4,4,4,4)
x4 <- c(1,2,2,2,1.7,2.5,2,2.5,2,2,2,2,3,3,3)
df = data.frame(y, x1, x2, x3, x4)
df
model <- lm(y~x1+x2+x3+x4, data = df)
model
#The results may slightly vary due to rounding off values
```

R code Exa 13.3 Productivity of Retail Grocery Outlets

```
y <- c(4.08,3.40,3.51,3.09,2.92,1.94,4.11,3.16,3.75,3.60)
```

```
x <- c
  (21.0,12.0,25.2,10.4,30.9,6.8,19.6,14.5,25.0,19.1)
plot(x,y)
cat("The relationship appears to be curvilinear and
  a quadratic model will provide more accurate
  estimations and predictions")
```

R code Exa 13.4 Productivity of Retail Grocery Outlets

```
y <- c
  (4.08,3.40,3.51,3.09,2.92,1.94,4.11,3.16,3.75,3.60)
x <- c
  (21.0,12.0,25.2,10.4,30.9,6.8,19.6,14.5,25.0,19.1)
model <- lm(y ~ x + I(x^2))
model
summary(model)
cat("quadratic model provide accurate prediction")
```

R code Exa 13.6 Regression Analysis

```
salary <- c(60710,NA
  ,63160,63210,64140,65760,65590,59510,60440,61340,61760,62750,6320
  NA)
x1 <- c(1,2,3,3,4,5,5,1,2,3,3,4,5,5)
x2 <- c(1,1,1,1,1,1,1,0,0,0,0,0,0,0)
df <- data.frame(salary, x1, x2)
df
model <- lm(salary ~ x1 * x2, data = df)
summary(model)
model
```

```
model1 <- aov(model, data = df)
model1
summary(model1)
par(mfrow=c(1,2))
plot(model, pch=16, which=1)
plot(model, pch=16, which=2)
#the textbook's f-value from anova table is wrong and rest all of the values are correct
```

R code Exa 13.7 Sufficient Evidence Test

```
salary <- c(60710,NA,63160,63210,64140,65760,65590,59510,60440,61340,61760,62750,63200
  NA)
x1 <- c(1,2,3,3,4,5,5,1,2,3,3,4,5,5)
x2 <- c(1,1,1,1,1,1,1,0,0,0,0,0,0,0)
df <- data.frame(salary, x1, x2)
df
model <- aov(salary ~ x1 * x2, data = df)
model
summary(model)
p_value <- summary(model)[[1]][["Pr(>F)"]][[3]]
p_value
cat("the p-value =",p_value,"is twice of what it would be for one- tailed test so null hypothesis is rejected ")
cat("So there is sufficient evidence to indicate that the annual rate of increase in men's faculty salaries exceeds the rate for women")
```

R code Exa 13.8 Real Estate Data

```r
y <- c
  (169.0,218.5,216.5,225.5,229.9,235.0,239.9,247.9,260.0,269.9,234.
x1 <- c(6,10,10,11,13,13,13,17,19,18,13,18,17,20,21)
x2 <- c(1,1,1,1,1,2,1,2,2,1,1,1,2,2,2)
x3 <- c(2,2,3,3,3,3,3,3,3,3,4,4,4,4,4)
x4 <- c(1,2,2,2,1.7,2.5,2,2.5,2,2,2,2,3,3,3)
k <- 4
r <- 1
df = data.frame(y, x1, x2, x3, x4)
df
model <- lm(y~x1+x2+x3+x4, data = df)
model
summary(model)
aov(model)
summary(aov(model))
sr <- (summary(aov(model)))
model1 <- lm(y ~ x1)
model1
aov(model1)
s <- (summary(aov(model1)))
s
sse1 <- s[[1]][[2]][[2]]
sse2 <- sr[[1]][[2]][[5]]
sse2
mse2 <- sr[[1]][[3]][[5]]
mse2
f <- round(((sse1 - sse2) / ((k - r))/mse2),1)
cat("test statistics is",f)
alpha <- 0.05
critical_value_f <- qf(1 - alpha, df1 = 3, df2 = 10)
critical_value_f
if(f > critical_value_f){
  print("reject the null hypothesis")
}else{
  print("cannot reject null hypothesis")
}
#The results may slightly vary due to rounding off
```

values

Chapter 14

Analysis of Categorical Data

R code Exa 14.1 Rat problem

```
df1 <- data.frame(green <- 20, red <- 39, blue <-
    31)
chisq <- chisq.test(df1)
n <- 90
p_value <- chisq$p.value
p_value
cat("since the p-value is less than 0.05 so the null
     hypothesis is rejected")
cat("So there is sufficient evidence to indicate
    that the rat has a preference for one of the
    three doors")
```

R code Exa 14.2 Blood phenotype

```
o1 <- 89
o2 <- 18
o3 <- 12
o4 <- 81
```

```r
e1 <- 82
e2 <- 20
e3 <- 8
e4 <- 90
p1 <- 0.41
p2 <- 0.10
p3 <- 0.04
p4 <- 0.45
k <- 4
df <- k -1
alpha <- 0.100
x_square_observed <- ((o1 - e1) * (o1 - e1) / e1) + ((o2 - e2) * (o2 - e2) / e2) + ((o3 - e3) * (o3 - e3) / e3) + ((o4 - e4) * (o4 - e4) / e4)
x_square_observed
x_square_0.100 <-qchisq(1 - alpha, df)   #x-square at alpha = 0.100 using chi- square
x_square_0.100
if(x_square_observed < x_square_0.100){
  print("p-value is greater than 0.100, do not have sufficient evidence to reject null h0")
  }else{
  print("reject the h0")
}
```

R code Exa 14.4 Furniture defect

```r
shift1 <- c(15,21,45,13)
shift2 <- c(26,31,34,5)
shift3 <- c(33,17,49,20)
data <- data.frame(shift1, shift2, shift3)
data
chisq.test(data)
p_value <- chisq.test(data)$p.value
cat(p_value < 0.05)
```

```
cat("since p−value is less than 0.005 so null
    hypothesis can be rejected and there is
    sufficient evidene that the proportion of defect
    types vary shift to shift")
#the answers may slightly vary due to rounding off
    values
```

R code Exa 14.5 Flu Vaccine

```
survey <- matrix(c(24,9,13,289,100,565),ncol = 3,
    byrow = TRUE)
colnames(survey) <- c("No Vaccine","One Shot","Two
    Shots")
rownames(survey) <- c("Flu","No Flu")
survey <- as.table(survey)
survey
chisq.test(survey,correct = F)
cat("null hypothesis is rejected so there is
    sufficient evidence to indicate a relationship
    between treatment and influence of flu")
```

R code Exa 14.7 Survey of Voter

```
voter <- matrix(c(76,53,59,48,124,147,141,152),ncol
    = 4, byrow = TRUE)
colnames(voter) <- c("1","2","3","4")
rownames(voter) <- c("Favor A","Do Not Favor A")
voter <- as.table(voter)
voter
chisq.test(voter,correct = F)
cat("X−squared value is too large so null hypothesis
     is rejected")
```

Chapter 15

Nonparametric Statistics

R code Exa 15.1 Euglossine bees

```
species1 <- c(235,225,190,188);
species2 <- c(180,169,180,185,178,182);
n1 <-length(species1);
n2 <- length(species2);
data <- c(species1,species2);
rank1 <- rank(sort(data))
t1 <- 7+8+9+10;
t1_ <- n1 * (n1 + n2 + 1) - t1
t1_                                     #for n1=4 and n2=6, the critical value of T at alpha = 0.05 is 12
if(t1_ <= 12){
  print("Reject the Null Hypothesis")
}else{
    print("Accept the Null Hypothesis")
  }
```

R code Exa 15.2 Kraft papers

```
standard1 <- c(1.21,1.43,1.35,1.51,1.39,1.17,1.48,1.42,1.29,1.40)
treated2 <- c(1.49,1.37,1.67,1.50,1.31,1.29,1.52,1.37,1.44,1.53)
n1 <- length(standard1)
n2 <- length(treated2)
alpha = 0.05
x <- c(standard1,treated2)
ranksum<-function(x,start,end){
  return(sum(x[start:end]))
}
rank <- rank(x)
t1 <- ranksum(rank,1,n1)
t2 <- n1 * (n1 + n2 + 1) - t1
if(t1 <= 82){                          #critical value of T at n1=n2=10 at alpha = 0.05 is 82
  print("Reject the hypothesis")
}else{
  print("Insufficient evidence to conclude that the treated kraft paper is stronger than the standard paper")
}
muo_t <- (n1 * (n1 * n2 + 1))/2
sigma_sqr_t <- ((n1 * n2) *(n1 + n2 +1))/12
sigma_t <- sqrt(sigma_sqr_t)
z <- (t1 - muo_t)/sigma_t;
p_value <- 0.5 - 0.4292
p_value
if(p_value <= alpha){
  print("Reject the hypothesis")
}else{
  print("Cannot conclude the treated kraft paper is stronger than the standard paper")
}
```

R code Exa 15.3 Defective Electrical Fuses

```
line_A <- c(170,164,140,184,174,142,191,169,161,200)
    ;
line_B <- c(201,179,159,195,177,170,183,179,170,212)
    ;
n = 10;
x = 1
p_value = 2 * round(pbinom(x,n,0.5),3)              #
    hypothesized value of p is 0.5
p_value
cat("Reject the hypothesis at 5% level ")
```

R code Exa 15.4 Employee accident rates

```
x <- 63;
n <- 100;
alpha <- 0.05;
z <- (x - 0.5 * n)/(0.5 * sqrt(n));
z_expected <- qnorm(1-alpha/2)
if(z == z_expected ){
print("Accept the null hypothesis")
}else{
  print("Reject the null hypothesis and the data
      provide sufficient evidence")
}
```

R code Exa 15.5 Densities of cakes

```
x_A <- c(.135,.102,.098,.141,.131,.144)
x_B <- c(.129,.120,.112,.152,.135,.163)
alpha <- 0.10
t0 <- round(qt(1 - alpha/2,5),1)
boxplot(x_A - x_B,horizontal = TRUE,xlab = "Differences")
wilcox.test(x_A,x_B,paired = T)
t_positive <- wilcox.test(x_A,x_B,paired = T)$statistic
t0
t_positive
if(t_positive <= t0){
  print("Reject the null hypothesis so two population frequency distribution of cake densities differ")
}else{
  print("Accept the hypothesis")
}
```

R code Exa 15.6 Kruskal Wallis H test

```
one <- c(65,87,73,79,81,69)
two <- c(75,69,83,81,72,79,90)
three <- c(59,78,67,62,83,76)
four <- c(94,89,80,88)
k <- 4
kruskal.test(list(one, two, three, four))
h <- kruskal.test(list(one, two, three, four))$statistic
h
df <- k - 1
alpha <- 0.05
x_square_0.05 <- qchisq(1 - alpha, df)
x_square_0.05
if(h >= x_square_0.05){
```

```
  print("sufficent evidence to indicate differences")
}else{
  print("there is insufficient evidence to indicate differences in the distributions of achievement test scores for the four teaching techniques")
}
#the answers may slightly vary due to rounding off values
```

R code Exa 15.8 Friedman Fr test

```
a <- c(.6,.7,.9,.5)
B <- c(.9,1.1,1.3,.7)
c <- c(.8,.7,1.0,.8)
d <- c(.7,.8,1.0,.6)
e <- c(.5,.5,.7,.4)
f <- c(.6,.5,.8,.6)
g <- c(a,B,c,d,e,f)
rank_a <- c(2.5,3.5,3,2)
rank_b <- c(6,6,6,5)
rank_c <- c(5,3.5,4.5,6)
rank_d <- c(4,5,4.5,3.5)
rank_e <- c(1,1.5,1,1)
rank_f <- c(2.5,1.5,2,3.5)
t1 <- sum(rank_a)
t2 <- sum(rank_b)
t3 <- sum(rank_c)
t4 <- sum(rank_d)
t5 <- sum(rank_e)
t6 <- sum(rank_f)
k <- 6
b <- 4
fr <-  12 / (b * k * (k + 1)) * (t1 * t1 + t2 * t2 + t3 * t3 + t4 * t4 + t5 * t5 + t6 * t6 ) - (3 * b
```

```
* (k + 1))
fr
df <-  k - 1
alpha <- 0.05
x_square <- qchisq(1 - alpha, 5)    #sampling distribution of fr
if(fr > x_square){
  print("reject null hypothesis, so the distribution of reaction times differ in location  for at least two stimuli")
}else{
  print("accpept the null hypothesis ")
}
```

R code Exa 15.9 P value

```
a <- c(.6,.7,.9,.5)
B <- c(.9,1.1,1.3,.7)
c <- c(.8,.7,1.0,.8)
d <- c(.7,.8,1.0,.6)
e <- c(.5,.5,.7,.4)
f <- c(.6,.5,.8,.6)
data2 <- matrix(c
  (.6,.9,.8,.7,.5,.6,.7,1.1,.7,.8,.5,.5,.9,1.3,1.0,1.0,.7,.8,.5,.7,
  ,nrow = 4, ncol = 6,byrow = T)
dimnames(data2) = list(c(1,2,3,4), c("a","b","c","d","e","f"))
friedman.test(data2)
p_value <- friedman.test(data2)$p.value
cat(" Approximate p-value = ",p_value,",which is slightly less than 0.005")
```

R code Exa 15.10 Spearman rank correlation coefficient

```r
xi <- c(7,4,2,6,1,3,8,5)
yi <- c(1,5,3,4,8,7,2,6)
di <- c(xi - yi)
n <- 8;
di_square <- c(di * di)
total = sum(di_square)
total
rs <- 1 - (6 * total)/((n) * (n * n - 1 ))
cat("rs is ",rs)
```

R code Exa 15.11 Hypothesis Test of no association

```r
xi <- c(7,4,2,6,1,3,8,5)
yi <- c(1,5,3,4,8,7,2,6)
alpha <- 0.05
cor.test(xi, yi)
rs <- cor.test(xi, yi)$estimate
cat("rs is",rs)
p_value <- cor.test(xi, yi)$p.value
if(p_value <= alpha){
  print("null hypothesis is rejected")
}else{
  print("null hypothesis is accepted")
}
```