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

May 29, 2020

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

6

8

9

Chapter 1

DESCRIBING DATA WITH GRAPHS

R code Exa 1.3 Pie and Bar Chart

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

R code Exa 1.4 Pareto Chart of Candies Problem

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

R code Exa 1.5 Bar and Pie Chart

```
1  category<-c("Millitary personnel","Operation and
       maintenance","Procurement","Research and
       development","Military construction","Other")
2  amount<-c(127.5,188.1,82.3,65.7,5.3,5.5)
3  label<-paste(category,amount)
4  label<-paste(label,"%",sep=" ")
5  par(mfrow=c(1,2))
6  pie(amount,labels = label,clockwise = TRUE)
7  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

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

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

R code Exa 1.7 Stem and Leaf Plot

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

R code Exa 1.8 Stem and Leaf Plot for Weights

```
1  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,
2  stem(weights)
```

R code Exa 1.10 Dotplot

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

R code Exa 1.11 Relative Frequency Histogram

```
1 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)

2 h<-hist(visits, plot=F)
3 h$counts <- h$counts / sum(h$counts)
4 plot(h, freq=TRUE, ylab="Relative Frequency")
```

Chapter 2

DESCRIBING DATA WITH NUMERICAL MEASURES

R code Exa 2.1 Dotplot and Sample Mean

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

R code Exa 2.2 Median

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

R code Exa 2.3 Median for Set

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

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

R code Exa 2.5 Variance and Standard Deviation

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

R code Exa 2.6 Tchebysheff theorem

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

R code Exa 2.7 Empirical Rule

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

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

R code Exa 2.9 Range Approximation

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

R code Exa 2.10 Range Approximation

```
1 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

2 x <- range(data)
3 range <- x[2]-x[1]
4 sd <- range /4
5 cat("standard deviation from range approximation is"
      ,sd)
```

R code Exa 2.11 Z Score

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

R code Exa 2.13 Lower and Upper Quartiles

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

R code Exa 2.14 Box Plot and Outlier

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

17

Chapter 3

DESCRIBING BIVARIATE DATA

R code Exa 3.1 Side by Side Bar Chart

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

R code Exa 3.2 Comparative Charts

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

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

R code Exa 3.3 Scatterplot

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

R code Exa 3.4 Scatterplot for Data

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

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

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

```
1  x<-c
      (1360,1940,1750,1550,1790,1750,2230,1600,1450,1870,2210,1480)

2  y<-c
      (278.5,375.7,339.5,329.8,295.6,310.3,460.5,305.2,288.6,365.7,425.

3  plot(x,y)
4  cat("Plot represents the linear pattern in data")
```

R code Exa 3.6 Correlation Coefficient

```
1  x<-c
      (1360,1940,1750,1550,1790,1750,2230,1600,1450,1870,2210,1480)

2  y<-c
      (278.5,375.7,339.5,329.8,295.6,310.3,460.5,305.2,288.6,365.7,425.

3  correlation_Coefficient <- round(cor(x, y),4)
4  cat("correlation coefficient of x and y is",
      correlation_Coefficient)
```

R code Exa 3.7 Line and Dataplot

```
1  #install package("ggplot2")
2  #install library("ggplot2")
3  x<-c(2,3,4,5,6,7)
4  y<-c(6.00,7.50,8.00,12.00,13.00,15.50)
5  install.packages("ggplot2")
6  DF<-data.frame(x,y)
7  library(ggplot2)
8  #read library("ggplot2")
9  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

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

R code Exa 4.6 Probabilities of Blood Phenotypes

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

R code Exa 4.7 Probability of Candies Problem

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

R code Exa 4.8 Probability Using Counting Rules

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

R code Exa 4.9 Candy Dish

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

R code Exa 4.10 Simple Events of Coins

```
1 coin_ways <- 2
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

```
1 Routes_A_B <- 3;
2 Routes_B_C <- 4;
3 Routes_C_D <- 3;
4 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

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

R code Exa 4.13 Permutaion of Tests

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

R code Exa 4.14 Counting Rule for Combinations

```
1 suppliers <- 5
2 choose <- 3
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

```
1  Total_ways <- 10;
2  Two_out_of_best <- choose(3, 2);
3  one_out_of_not_best <- choose(2, 1);
4  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

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

R code Exa 4.18 College Education Expenses Problem

```
1   Child_Too_High <- 0.35;
2   Child_Right_Amount <- 0.08;
3   Child_Too_Little <- 0.01;
4   No_Child_Too_High <- 0.25;
5   No_Child_Right_Amount <- 0.20;
6   No_Child_Too_Little <- 0.11;
7   too_high <- 0.60;
8   right_ammount <- 0.28;
9   too_little <- 0.12;
10  child_college <- Child_Too_High + Child_Too_Little +
       Child_Right_Amount;
```

```
11 cat("probability that respondent has a child in
       college is",child_college)
12 cat("probability that respondent does not have a
       child in college is",1-child_college)
13 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

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

R code Exa 4.20 Independent Events

```
1 s <- c("hh","ht","th","tt")
2 prob_a <- 1/2
3 prob_b <- 1/2
4 prob_a_and_b <- 1/4
5 if((prob_a * prob_b) == prob_a_and_b){
6   print("events must be independent")
7 }else{
8     print("events are not independent")
9   }
```

R code Exa 4.21 Multiplication Rule for Independent Events

```
1  Child_Too <- 0.35;
2  Child_Right_Amount <- 0.08;
3  Child_Too_Little <- 0.01;
4  No_Child_Too_High <- 0.25;
5  No_Child_Right_Amount <- 0.20;
6  No_Child_Too_Little <- 0.11;
7  too_high <- 0.60;
8  right_amount <- 0.28;
9  too_little <- 0.12;
10 child_college <- 0.44;
11 No_child_college <- 0.56;
12 x = too_high * child_college
13 cat("x and child_college values are not same so both
       events are dependent ")
```

R code Exa 4.22 Cards Problem

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

R code Exa 4.23 Law of Total Probability

28

```
1  pg1 <- 0.09
2  pg2 <- 0.20
3  pg3 <- 0.31
4  pg4 <- 0.23
5  pg5 <- 0.17
6  p_a_g1 <- 0.26
7  p_a_g2 <- 0.20
8  p_a_g3 <- 0.13
9  p_a_g4 <- 0.18
10 p_a_g5 <- 0.14
11 pA <- (pg1 * p_a_g1) + (pg2 * p_a_g2) +(pg3 * p_a_g3
      ) +(pg4 * p_a_g4) +(pg5 * p_a_g5)
12 cat("the required probability is",pA)
```

R code Exa 4.24 Sneakers Problem

```
1  pg1 <- 0.09
2  pg2 <- 0.20
3  pg3 <- 0.31
4  pg4 <- 0.23
5  pg5 <- 0.17
6  p_a_g1 <- 0.26
7  p_a_g2 <- 0.20
8  p_a_g3 <- 0.13
9  p_a_g4 <- 0.18
10 p_a_g5 <- 0.14
11 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))
12 cat("the required probability is",required_
      probability)
```

R code Exa 4.26 Mean Variance and Standard Deviation

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

R code Exa 4.27 Expected Gain

```
1  gain<-c(-20,23980)
2  prob_gain<-c((7998/8000),(2/8000))
3  prob_gain
4  expected_gain <- weighted.mean(gain, prob_gain)
5  #expected_gain is in dollar
6  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

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

R code Exa 5.4 Probabilities for free throws

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

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

R code Exa 5.5 Probabilities for Successes

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

R code Exa 5.6 Probability of Survivors

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

R code Exa 5.7 Multiple Choice Test

```
1 prob_correct <- 0.2
2 prob_incorrect <- 1 - prob_correct
3 n <- 100
4 mu0 <- n * prob_correct
5 sigma <- sqrt(n * prob_correct * prob_incorrect)
6 cat("a large proportion of score will lie within two
      standard deviations of the mean, or from",(mu0 -
      2 * sigma),"to",(mu0 + 2 * sigma),"." )
7 cat("allmost all the score will lie within three
      standard deviations of the mean, or from",(mu0 -
      3 * sigma),"to",(mu0 + 3 * sigma),"." )
8 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

```
1 mean <- 2;
2 accidents_in_case_one <- 0;
3 case_one <- round((dpois(accidents_in_case_one,mean)
    ),6)
4 cat("probability of no accident on this section of
    highway during a 1-week period is",case_one)
5 case_two_mean <- 2*mean;
6 case_two_mean
7 case_two <- round((dpois(0,case_two_mean)+ dpois(1,
    case_two_mean) + dpois(2,case_two_mean) + dpois
    (3,case_two_mean)),6)
8 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

```
1 total_men <- 5000;
2 prob <- 0.001;
3 claims <- 4;
4 mean <- total_men * prob
5 exact_prob <- round((dpois(claims,mean)),3)
6 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

```
1 total <- 1000;
2 prob <- 0.001;
3 mean <- total * prob;
4 none_defective <- dpois(0,mean)
5 cat("probability that none is defective is",none_
     defective)
6 three_defective <- dpois(3,mean)
7 cat("probability that three is defective",three_
     defective)
8 four_defective <- dpois(4,mean)
9 cat("probability that four are defective",four_
     defective)
10 #the asnwer may slightly vary due to rounding off
```

R code Exa 5.11 Probability Distribution Mean Variance

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

```r
5  cat("probability distribution of no bottle of
       spoiled wine is",prob_zero)
6  prob_one <- (choose(spoiled_wine,1) * choose(total_
       bottles - spoiled_wine,(sample - 1)))/choose(
       total_bottles,sample)
7  cat("probability distribution of  one bottle of
       spoiled wine in sample is",prob_one)
8  prob_two <- (choose(spoiled_wine,2) * choose(total_
       bottles - spoiled_wine,(sample - 2)))/choose(
       total_bottles,sample)
9  cat("probability distribution of  two bottle of
       spoiled wine in sample is",prob_two)
10 prob_three <- (choose(spoiled_wine,3) * choose(total
       _bottles - spoiled_wine,(sample - 3)))/choose(
       total_bottles,sample)
11 cat("probability distribution of  three bottle of
       spoiled wine in sample is",prob_three)
12 mean = sample * (spoiled_wine/total_bottles)
13 cat("mean is",mean)
14 variance <-  sample * (spoiled_wine/total_bottles)*
       (9/total_bottles)*((total_bottles-sample)/11)
15 cat("variance is",variance)
16 #"The answer may slightly vary due to  rounding off
       values"
```

R code Exa 5.12 Hypergeometric Probability Distribution

```r
1  total_items <- 20;
2  sample_items <- 5;
3  defective <- 4;
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)
5   cat("probability of get accepted is",prob_accept_lot
    )
```

Chapter 6

THE NORMAL PROBABILITY DISTRIBUTION

R code Exa 6.1 Probability

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

R code Exa 6.2 Probability of waiting time

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

R code Exa 6.3 Normal Probability Distribution

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

R code Exa 6.4 Normal Probability Distribution

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

R code Exa 6.5 Normal Probability Distribution

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

R code Exa 6.6 Normal Probability Distribution

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

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

R code Exa 6.7 z value

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

R code Exa 6.8 Normal Distributed Random Variable

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

R code Exa 6.9 Gasoline use for Compact car

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

R code Exa 6.10 Gasoline Use Rate

```
1 mean <- 25.5;
2 standard_deviation <- 4.5;
3 percent <- 95 / 100;
4 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

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

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

R code Exa 6.12 Electric Fuse

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

R code Exa 6.13 Soft Drinks Brand

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

```
9  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

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

R code Exa 7.4 Alzheimer disease

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

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

R code Exa 7.5 Sample Distribution of Sample Mean

```
1  no_of_bottles <- 10;
2  mean <- 12.1;
3  x_value <- 12;
4  standard_Deviation <- 0.2;
5  standard_error <- standard_Deviation/sqrt(no_of_
       bottles)
6  required_prob <- round((pnorm(x_value, mean,
       standard_error)),4)
7  cat("The required probability is",required_prob)
```

R code Exa 7.6 Sample Distribution of Sample Proportion

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

R code Exa 7.7 Probability for Sample Proportion

```
1 # textbook refered to example 7.6 for values
2 sample_space <- 500;
3 proportion <- 0.55;
4 observed_value <- 0.60;
5 standard_error <- round(sqrt((proportion * (1-
      proportion))/sample_space),4)
6 cat("standard error is",standard_error)
7 req_prob <- round(1 - pnorm(observed_value,
      proportion, standard_error),4)
8 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

```
1 #install package ("qicharts")
2 #install library ("qicharts")
3 install.packages("qicharts")
4 library(qicharts)
5 n <- 4
6 bearing1 <- c
      (0.992,1.015,0.988,0.996,1.015,1.000,0.989,0.994,1.018,0.997,1.020

7 bearing2 <- c
      (1.007,0.984,0.993,1.020,1.006,0.982,1.009,1.010,1.016,1.005,0.986

8 bearing3 <- c
      (1.016,0.976,1.011,1.004,1.002,1.005,1.019,1.009,0.990,0.989,1.001
```

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

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

R code Exa 7.9 Control Chart

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

6 n <- 400
7 defects <- c(proportion * n)
8 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

```
1 sample <- 50;
2 standard_deviation <- 105;
3 margin_of_error <- 1.96 * (standard_deviation/sqrt(
    sample))
4 margin_of_error
5 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

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

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

R code Exa 8.6 Construction of Confidence Interval

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

R code Exa 8.7 99 percent Confidence Interval

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

R code Exa 8.8 Large Sample Confidence Interval

```r
1  sample_size <- 985;
2  vote_for_republican <- 592;
3  point_estimate <- vote_for_republican/sample_size;
4  standard_error <- sqrt((point_estimate * (1 - point_
     estimate))/sample_size)
5  standard_error
6  point_estimate
7  coin_interval <- 90/100;
8  z <- qt((1+coin_interval)/2,df=sample_size-1)
9  value <- z * standard_error
10 left <- round((point_estimate - value),3);
11 right <- round((point_estimate + value),3);
12 cat("90% confidence interval is from ",left, "to" ,
     right)
```

R code Exa 8.9 Difference Between Two Sample Mean

```r
1  miles_cover_by_type1 <- 26400;
2  miles_cover_by_type2 <- 25100;
3  type1_sample <- 100;
4  type2_sample <- 100;
5  variance1 <- 1440000;
6  variance2 <- 1960000;
7  point_estimate <- miles_cover_by_type1 - miles_cover
     _by_type2;
8  standard_error <- sqrt(((variance1)/type1_sample) +
     (variance2)/type2_sample);
9  confidence_interval_percent <- 0.99;
10 z_value <- round((qnorm(0.995)),2)
11 value <- z_value * standard_error;
12 left <- point_estimate - value;
13 right <- point_estimate + value;
14 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

```
1 sample_size_men <- 50;
2 sample_size_women <- 50;
3 sample_mean_men <- 756;
4 sample_mean_women <- 762;
5 sample_standard_Deviation_men <- 35;
6 sample_standard_Deviation_women <- 30;
7 z_value <- 1.96
8 #we know from z table that for 95% confidence
      interval the z value is 1.96;
9 point_estimate <- sample_mean_men -sample_mean_women
      ;
10 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))
11 value <- z_value * standard_error;
12 left <- round((point_estimate - value),2);
13 right <- round((point_estimate + value),2);
14 cat("The 95% confidence interval is from",left,"to",
      right)
```

R code Exa 8.11 Bond proposal problem

```
1 p1_cap <- 0.76
2 p2_cap <- 0.65
3 q1_cap <- 0.24
4 q2_cap <- 0.35
5 n1 <- 50
6 n2 <- 100
```

```
 7  p1_cap - p2_cap
 8  standard_error <- round(sqrt(((p1_cap * q1_cap) / n1
       ) + ((p2_cap * q2_cap) / n2 )),4)
 9  standard_error
10  alpha <- 0.01        #at 99% confidence interval
       alpha = 0.01
11  z_0.005 <- qnorm( 1 - alpha/2)
12  lower_value <- (p1_cap - p2_cap) - z_0.005 *
       standard_error
13  upper_value <- (p1_cap - p2_cap) + z_0.005 *
       standard_error
14  cat("the interval is from",lower_value,"to",upper_
       value)
15  n <- 150
16  point_estimation <- 103/n
17  margin_of_error <- 1.96 * sqrt((point_estimation *
       (1 - point_estimation) / n))
18  margin_of_error1 <- - 1.96 * sqrt((point_estimation
       * (1 - point_estimation) / n))
19  cat("margin of error is",margin_of_error,"and",
       margin_of_error1,".")
20  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

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

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

R code Exa 8.13 Plastic Pipe

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

R code Exa 8.14 Workers in Training groups

```
1 range <- 8
2 sigma1 <- sigma2 <- sigma <- range / 4
3 alpha <- 1 - 0.95  # at 0.95 confidence coefficient
4 z_0.05 <- qnorm(1 - alpha/2)
5 z_0.05
6 n <- (z_0.05 * sqrt(8)) ** 2
7 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

```r
1 mean <- 14
2 s <- 2
3 n <- 100
4 se <- s / sqrt(n)
5 x_bar <- 15
6 z <- (x_bar - mean) / se
7 z
8 p_value <- 2 * pnorm(-abs(z))
9 cat("p-value is approximately zero")
```

R code Exa 9.4 Appropriate Hypothesis

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

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

R code Exa 9.5 Appropriate Hypothesis

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

R code Exa 9.6 Calculation P Value

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

R code Exa 9.7 Daily Sodium Intake

```
1  sample_mean <- 3400;                # sample mean
2  hypothesized_value <-  3300;                  #
      hypothesized value
3  sigma <- 1100;            #  standard deviation
4  sample_size <- 100;                  # sample size
5  z_value <- (sample_mean - hypothesized_value)/(sigma
      /sqrt(sample_size))
6  z_value                       # test statistic
7  alpha <- 0.05
8  z.alpha <- qnorm(1 - alpha)
9  z.alpha                 # critical value
10 cat("The value of z is",z_value)
11 cat("critical value is",z.alpha)
12 p_value <- 1 - pnorm(z_value)
13 cat("p value is",p_value)
14 cat("null hyppothesis is not rejected")
15 cat("not enough evidence ")
16 #the answer may slightly vary due to rounding off
      values
```

R code Exa 9.8 Beta and power of beta

```
1  sample_mean <- 871;                    # sample mean
2  hypothesized_value <-  880;                     #
     hypothesized value
3  sigma <- 21;              #  standard deviation
4  sample_size <- 50;                     # sample size
5  alpha <- .05
6  z.alpha <- round((qnorm(1 - alpha/2)),2)
7  z.alpha                  # critical value
8  lower <- round(hypothesized_value - z.alpha * ((
     sigma/sqrt(sample_size))),2)
9  lower
10 upper <- round(hypothesized_value + z.alpha * ((
     sigma/sqrt(sample_size))),2)
11 upper
12 mu <- 870
13 z1 <- round(((lower - mu)/(sigma/sqrt(sample_size)))
     ,2)
14 z1
15 z2 <- round(((upper - mu)/(sigma/sqrt(sample_size)))
     ,2)
16 z2
17 beta <- round( 1 - pnorm( z1),4)
18 power_of_test <- 1 - beta
19 cat("beta is",beta)
20 cat("power of test is",power_of_test)
```

R code Exa 9.9 Car Ownership Affect

```
1  non_owners <- 100;
```

```r
2  owners <- 100;
3  average_non_owner <- 2.70;
4  average_owner <- 2.54;
5  variance_non_owner <- 0.36;
6  variance_owner <- 0.40;
7  d_nod <- 0;
8  point_estimate <-  average_non_owner - average_owner
      ;
9  standard_error <- sqrt(((variance_non_owner)/non_
      owners) + (variance_owner)/owners);
10 z = round((((point_estimate - d_nod)/standard_error)
      ,2);
11 cat("value of z is",z)
12 alpha <- .05
13 z.alpha <- round((qnorm(1 - alpha/2)),2)
14 z.alpha                    # critical value
15 cat("The value of z is",z)
16 cat("critical value is",z.alpha)
17 p_value <- round(((1-pnorm(z)) + pnorm(-z)),4)
18 cat("p value is",p_value)
19 cat("null hypothesis cannot  be rejected and there
      is insufficient evidence")
```

R code Exa 9.10 Hypothesis Testing and Confidence Interval

```r
1  # In textbook values are refered to example 9.9
2  non_owners <- 100;
3  owners <- 100;
4  average_non_owner <- 2.70;
5  average_owner <- 2.54;
6  variance_non_owner <- 0.36;
7  variance_owner <- 0.40;
8  d_nod <- 0;
9  point_estimate <-  average_non_owner - average_owner
      ;
```

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

R code Exa 9.11 Hypothesis Testing and P Value

```
1 hypothesized_value <- 0.2;
2 observed_value <- 15/100;
3 standard_error <- sqrt((hypothesized_value * (1-
     hypothesized_value))/100)
4 z_value <- (observed_value - hypothesized_value)/
     standard_error;
5 p_value <- pnorm(z_value)
6 cat("pvalue is",p_value)
7 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

```
1 admitted_men <- 52;
2 admitted_women <- 23;
3 sample_men <- 1000;
4 sample_women <- 1000;
```

```r
 5  p1_value <- admitted_men/sample_men;
 6  p2_value <- admitted_women/sample_women;
 7  pooled_estimate <- (admitted_men + admitted_women)/(
        sample_men + sample_women);
 8  standard_error <- sqrt(pooled_estimate * (1 - pooled
        _estimate) * ((1/1000) + (1/1000)))
 9  test_statistics <- (p1_value - p2_value)/standard_
        error
10  test_statistics
11  alpha <- 0.05;
12  k <- abs(qnorm(alpha))
13  lower_bound <- (p1_value  - p2_value) - k * sqrt((p1
        _value * (1 - p1_value)/sample_men) + (p2_value *
         (1 - p2_value)/sample_women))
14  cat("Lowest likely value for the difference is",
        lower_bound)
15  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

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

R code Exa 10.3 Average Weight of Diamonds

```
1  weights <- c(0.46,0.61,0.52,0.48,0.57,0.54)
2  t.test(weights, mu = 0.5, alternative = "greater",
      conf.level = 0.95)
3  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

```r
1 data <- c(310,311,412,368,447,376,303,410,365,350)
2 mean <- mean(data)
3 s <- sd(data)
4 mu0 <- 400
5 t.test(data,mu = mu0)
6 p.value <- t.test(data,mu = mu0)$p.value
7 p.value
8 cat(p.value < 0.05)
9 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

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

```
15  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.")
16    }
```

R code Exa 10.6 P Value For Student Grades

```
1  #values are referred to example 10.5 in textbook
2  online <-c (32,37,35,28,41,44,35,31,34)
3  classroom <-c (35,31,29,25,34,40,27,32,31)
4  p_value <- t.test(online,classroom,alternative = "
      greater")$p.value
5  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

```
1  online <- c(32,37,35,28,41,44,35,31,34)
2  classroom <- c(35,31,29,25,34,40,27,32,31)
3  t.test(online,classroom,alternative = "greater")
4  lower <- t.test(online,classroom,alternative = "
      greater")$conf.int[1]
5  cat("lower confidence bound is",lower)
6  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.")
7  #the results may slightly vary due to rounding off
      values in textbook.
```

R code Exa 10.8 Tire problem

```
1 tire_A <- c(10.6,9.8,12.3,9.7,8.8)
2 tire_B <- c(10.2,9.4,11.8,9.1,8.3)
3 alpha <- 0.05
4 critical_value <- qt( 1 - alpha/2,df = 4)
5 k <- t.test(tire_A,tire_B, paired = TRUE,
      alternative = "two.sided")
6 test_statistics <- k[1]
7 cat(test_statistics > critical_value)
8 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

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

R code Exa 10.11 Cement manufacturer

```
1 sigma_square <- 100
2 n <- 10
3 xbar <- 312
4 s_square <- 195
5 x_square <- ((n -1) * s_square)/sigma_square
6 alpha <- 0.05
```

63

```
 7  df <- n -1
 8  x_square_0.05 <- qchisq(1 - alpha, df)
 9  if(x_square > x_square_0.05){
10    print("reject the null hypothesis and the range of
          concrete strength measurements exceeds the
        manufacturer's claim")
11  }else{
12    print("accept the null hypothesis")
13  }
```

R code Exa 10.12 90 percent confidence interval

```
 1  sigma_square <- 4
 2  n <- 3
 3  df <- n - 1
 4  measurements <- c(4.1,5.2,10.2)
 5  s_square <- var(measurements)
 6  s_square
 7  x_square <- ((n - 1) * s_square) / sigma_square
 8  cat("p-value is greater than 0.10 so accept the null
        hypothesis")
 9  alpha <- 0.10          #at 90% confidence interval alpha
        = 0.10
10  x_square_0.95 <- qchisq(alpha/2,df)
11  x_square_0.05 <- qchisq(1-alpha/2,df)
12  x_square_0.05
13  lower_value <- round((n - 1) * s_square/x_square_
        0.05,2)
14  upper_value <- round((n - 1) * s_square/x_square_
        0.95,2)
15  cat("the interval is from",lower_value,"to",upper_
        value)
```

R code Exa 10.13 F Value

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

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

```
1  variance1 <- 7.14;
2  variance2 <- 3.21;
3  n1 <- 10;
4  n2 <- 8;
5  df1 <- n1 - 1
6  df2 <- n2 - 1
7  alpha <- 0.05
8  rejection_region <- round(qf(1 - (alpha / 2), df1,
      df2),2)
9  rejection_region
10 test_statistics <- variance1/variance2;
11 cat("the calculated value of test statistics is",
      test_statistics)
12 if(test_statistics > rejection_region){
13   print("reject the null hypothesis")
14 }else{
15   print("Cannot reject null hypothesis and there is
         sufficient evidence to indicate a difference in
          the population variance")
16 }
```

R code Exa 10.15 Confidence Interval Estimate

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

R code Exa 10.16 Impurities in the batch of chemical

```
1  x1bar <- 302
2  x2bar <- 3.0
3  s1_square <- 1.04
4  s2_square <- 0.51
5  n1 <- n2 <- 25
6  f <- (s1_square / s2_square)
7  df1 <- df2 <- n1 - 1
8  f_0.050 <- qf(1 - 0.050,df1,df2)
9  f_0.025 <- qf(1 - 0.025,df1,df2)
10 if(f > f_0.050 & f < f_0.025){
11   print("p-value lies between 0.025 and 0.05 and
        hence at 5% level null hypothesis is rejected")
12 }else{
13   print("null hypothesis is accepted")
14 }
```

Chapter 11

THE ANALYSIS OF VARIANCE

R code Exa 11.4 Analysis of Variance Table

```
1  no_breakfast<-c(8,7,9,13,10)
2  light_breakfast<-c(14,16,12,17,11)
3  full_breakfast<-c(10,12,16,15,12)
4
5  k <- 3;
6  n1 <- n2 <- n3 <- 5
7  n <- 15;
8
9
10 sum_square_X <- sum(no_breakfast) + sum(light_
      breakfast) + sum(full_breakfast)
11 cm <- (sum_square_X * sum_square_X)/n
12 cm
13 x <- sum(no_breakfast)
14 y <- sum(light_breakfast)
15 z <- sum(full_breakfast)
16 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
```

```
17  total_ss
18  degree_of_freedom <- n-1
19  degree_of_freedom
20  sst <- ((x*x + y*y + z*z)/5) - cm
21  sst
22  df <- k - 1
23  sse  <- total_ss -sst
24  sse
25  deg_of_freedom <- n - k
26  deg_of_freedom
27
28  combined<-data.frame(cbind(no_breakfast,light_
       breakfast,full_breakfast))
29  stacked<-stack(combined)
30  stacked
31  Anova_Results<-aov(values ~ ind,data = stacked)
32  summary(Anova_Results)
```

R code Exa 11.5 F test

```
1  no_breakfast<-c(8,7,9,13,10)
2  light_breakfast<-c(14,16,12,17,11)
3  full_breakfast<-c(10,12,16,15,12)
4
5  k <- 3;
6  n1 <- n2 <- n3 <- 5
7  n <- 15;
8
9
10 sum_square_X <- sum(no_breakfast) + sum(light_
      breakfast) + sum(full_breakfast)
11 cm <- (sum_square_X * sum_square_X)/n
12 cm
13 x <- sum(no_breakfast)
14 y <- sum(light_breakfast)
```

```
15  z <- sum(full_breakfast)
16  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
17  total_ss
18  degree_of_freedom <- n-1
19  degree_of_freedom
20  sst <- ((x*x + y*y + z*z)/5) - cm
21  sst
22  df1 <- k - 1
23  df1
24  sse  <- total_ss -sst
25  sse
26  df2 <- n - k
27  df2
28
29  combined<-data.frame(cbind(no_breakfast,light_
        breakfast,full_breakfast))
30  stacked<-stack(combined)
31  stacked
32  Anova_Results<-aov(values ~ ind,data = stacked)
33  summary(Anova_Results)
34  mse <- sse / (n - k)
35  mse
36  mst <- sst / (k - 1)
37  mst
38  f <- mst / mse
39  f
40  alpha <- 0.05
41  f_0.05 <- qf(1 - alpha, df1, df2)
42  if(f > f_0.05){
43    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")
44  }
```

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

```r
1  #values are referred to example 11_4in the textbook
2  no_breakfast<-c(8,7,9,13,10)
3  light_breakfast<-c(14,16,12,17,11)
4  full_breakfast<-c(10,12,16,15,12)
5  x1_bar <- mean(no_breakfast)
6  x2_bar <- mean(light_breakfast)
7  x3_bar <- mean(full_breakfast)
8
9  k <- 3;
10 n1 <- n2 <- n3 <- 5
11 n <- 15;
12
13
14 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
15 total_ss
16 degree_of_freedom <- n-k
17 degree_of_freedom
18 sst <- ((x*x + y*y + z*z)/5) - cm
19 sst
20 sse  <- total_ss -sst
21 sse
22 mse <- sse/(n-k)
23 s2 <- mse
24 s <- sqrt(s2)
25 alpha <- 0.05/2;
26 t_value <- qt(1-alpha,degree_of_freedom)
27 t_value
28 left1 <- round(x1_bar - (t_value * (s/sqrt(n1))),2)
```

```
29  right1 <- round(x1_bar + (t_value * (s/sqrt(n1))),2)
30  cat("confidence interval for no breakfast are
        between",left1,"and",right1,"minutes")
31  left2 <- round((x2_bar - x3_bar) - t_value*(sqrt(s2*
        ((1/n2) + (1/n3)))),2)
32  right2 <- round((x2_bar - x3_bar) + t_value*(sqrt(s2
        *((1/n2) + (1/n3)))),2)
33  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

```
1  no_breakfast<-c(8,7,9,13,10)
2  light_breakfast<-c(14,16,12,17,11)
3  full_breakfast<-c(10,12,16,15,12)
4
5  k <- 3;
6  n1 <- n2 <- n3 <- 5
7  n <- 15;
8
9  sum_square_X <- sum(no_breakfast) + sum(light_
        breakfast) + sum(full_breakfast)
10 cm <- (sum_square_X * sum_square_X)/n
11 cm
12 x <- sum(no_breakfast)
13 y <- sum(light_breakfast)
14 z <- sum(full_breakfast)
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
16 total_ss
17 sst <- ((x*x + y*y + z*z)/5) - cm
18 sst
```

```
19 sse   <- total_ss -sst
20 sse
21 df <- n - k
22 df
23 mse <- sse / (n - k)
24 mse
25 mst <- sst / (k - 1)
26 mst
27 f <- mst / mse
28 f
29 s <- sqrt(mse)
30 s
31 alpha <- 0.05
32 w <- (qtukey(1 - alpha, k, df)) * (s / sqrt(n1))
33 w
34 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

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

```
14 summary(model)
```

R code Exa 11.9 Evidence Indication

```
1 b <- 3
2 k <- 4
3 low <- c(27,24,31,23)
4 middle <- c(68,76,65,67)
5 high <- c(308,326,312,300)
6 df <- data.frame(low, middle, high)
7 data <-c((as.matrix(df)))
8 data
9 f <- c("a","b","c","d")
10 company <- gl(k,1, b*k,factor(f))
11 usage <- gl(b, k, k * b)
12 model <- aov(data ~ usage + company)
13 model
14 summary(model)
15 p.value <- summary(model)[[1]][["Pr(>F)"]][[2]]
16 p.value
17 cat("since the p-value from anova test p.value=",p.
       value,"is too large to allow rejection of null
       hypothesis.")
18 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

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

```r
 5  t1 <- sum(A)
 6  t2 <- sum(B)
 7  t3 <- sum(C)
 8  t4 <- sum(D)
 9  b1 <- 105
10  b2 <- 276
11  b3 <- 1246
12  k <- 4;
13  n1 <- n2 <- n3 <- n4 <- 3
14  n <- 12;
15  b <- 3
16
17  sum_square_X <- sum(A) + sum(B) + sum(C) + sum(D)
18  sum_square_X
19  cm <- (sum_square_X * sum_square_X)/n
20  cm
21
22  l <- c(A,B,C,D)
23  l
24  sum_of_all <-sum( c(l * l))
25  sum_of_all
26  total_ss <- sum_of_all - cm
27  total_ss
28  sst <- (( t1 * t1 + t2 * t2 + t3 * t3 + t4 * t4)/ n1
        )- cm
29  sst
30  ssb <- ((b1 * b1 + b2 * b2 + b3 * b3) / k ) - cm
31  ssb
32  sse  <- total_ss - sst - ssb
33  sse
34  t2_bar <- t2 / n2
35  t3_bar <- t3 / n3
36  mse <- sse / ((b - 1) * (k - 1))
37  mse
38  alpha <- 0.05
39  df <- (b - 1)  * (k - 1)
40  lower_range <- round((t2_bar - t3_bar) - qt(1 -
        alpha/2,df) * sqrt(mse * (2 / b)),2)
```

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

R code Exa 11.12 Two Way ANOVA

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

Chapter 12

LINEAR REGRESSION AND CORRELATION

R code Exa 12.1 Least Squares Prediction Line

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

R code Exa 12.2 Hypothesis Test for Linear Relationship

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

```
 8  sum_y <- sum(y)
 9  sum_x_square <- sum(x_square)
10  sum_y_square <- sum(y_square)
11  sum_xy <- sum(xy)
12  s_xx <- sum_x_square - (sum_x * sum_x)/n
13  s_yy <- sum_y_square - (sum_y * sum_y)/n
14  s_xy <- sum_xy - (sum_x *sum_y)/n
15  y_bar <- sum(y)/n
16  x_bar <- sum(x)/n
17  b <- (s_xy/s_xx)
18  a <- y_bar - b*x_bar
19  a
20  b
21  total_ss <- s_yy
22  ssr <- (s_xy * s_xy)/s_xx
23  sse <- total_ss - ssr
24  mse <- sse / (n - 2)
25  mse
26  sse
27  test_statistics <- round(((b - 0)/sqrt(mse / s_xx))
       ,2)
28  cat("test statistics is",test_statistics)
29  alpha <- 0.05              # 5% significance level
30  range <- round(qt(1 - alpha/2, df = 8),3)
31  range
32  cat("rejection region is greater then",range,"or","
       less than",-range)
33  cat("there is significant linear relationship")
```

R code Exa 12.3 Confidence Interval Estimate

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

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

R code Exa 12.4 Average Calculus Grade

```
1  x <- c(39,43,21,64,57,47,28,75,34,52)
2  y <- c(65,78,52,82,92,89,73,98,56,75)
3  x0 <- 50
4  df <- data.frame(x, y)
5  lmresult <- lm(y ~ x, data = df)
6  lmresult
7  df2 <- data.frame(x = 50)
8  df2
9  ycap <- predict(lmresult, df2, interval = "
       confidence")[1]
10 ycap
11 lower <- predict(lmresult, df2, interval = "
       confidence")[2]
12 upper <- predict(lmresult, df2, interval = "
       confidence")[3]
13 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

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

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

R code Exa 12.6 Grade Achievement test

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

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

R code Exa 12.7 Correlation Coefficient

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

R code Exa 12.8 Correlation Coefficient

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

Chapter 13

Multiple Regression Analysis

R code Exa 13.2 Multiple Regression Analysis

```
1  y <- c
      (169.0,218.5,216.5,225.5,229.9,235.0,239.9,247.9,260.0,269.9,234.9

2  x1 <- c(6,10,10,11,13,13,13,17,19,18,13,18,17,20,21)
3  x2 <- c(1,1,1,1,1,2,1,2,2,1,1,1,2,2,2)
4  x3 <- c(2,2,3,3,3,3,3,3,3,3,4,4,4,4,4)
5  x4 <- c(1,2,2,2,1.7,2.5,2,2.5,2,2,2,2,3,3,3)
6  df = data.frame(y, x1, x2, x3, x4)
7  df
8  model <- lm(y~x1+x2+x3+x4, data = df)
9  model
10 #The results may slightly vary due to rounding off
      values
```

R code Exa 13.3 Productivity of Retail Grocery Outlets

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

```
2  x <- c
      (21.0 ,12.0 ,25.2 ,10.4 ,30.9 ,6.8 ,19.6 ,14.5 ,25.0 ,19.1)

3  plot(x,y)
4  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

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

2  x <- c
      (21.0 ,12.0 ,25.2 ,10.4 ,30.9 ,6.8 ,19.6 ,14.5 ,25.0 ,19.1)

3  model <- lm(y ~ x + I(x^2))
4  model
5  summary(model)
6  cat("quadratic model provide accurate prediction")
```

R code Exa 13.6 Regression Analysis

```
1  salary <- c(60710 ,NA
      ,63160 ,63210 ,64140 ,65760 ,65590 ,59510 ,60440 ,61340 ,61760 ,62750 ,63200
      NA)
2  x1 <- c(1,2,3,3,4,5,5,1,2,3,3,4,5,5)
3  x2 <- c(1,1,1,1,1,1,1,0,0,0,0,0,0,0)
4  df <- data.frame(salary , x1, x2)
5  df
6  model <- lm(salary ~ x1 * x2, data = df)
7  summary(model)
8  model
```

83

```
 9  model1 <- aov(model, data = df)
10  model1
11  summary(model1)
12  par(mfrow=c(1,2))
13  plot(model, pch=16, which=1)
14  plot(model, pch=16, which=2)
15  #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

```
 1  salary <- c(60710,NA
       ,63160,63210,64140,65760,65590,59510,60440,61340,61760,62750,63200(
       NA)
 2  x1 <- c(1,2,3,3,4,5,5,1,2,3,3,4,5,5)
 3  x2 <- c(1,1,1,1,1,1,1,0,0,0,0,0,0,0)
 4  df <- data.frame(salary, x1, x2)
 5  df
 6  model <- aov(salary ~ x1 * x2, data = df)
 7  model
 8  summary(model)
 9  p_value <- summary(model)[[1]][["Pr(>F)"]][[3]]
10  p_value
11  cat("the p-value =",p_value,"is twice of what it
       would be for one- tailed test so null hypothesis
       is rejected ")
12  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
1  y <- c
       (169.0,218.5,216.5,225.5,229.9,235.0,239.9,247.9,260.0,269.9,234.9
2  x1 <- c(6,10,10,11,13,13,13,17,19,18,13,18,17,20,21)
3  x2 <- c(1,1,1,1,1,2,1,2,2,1,1,1,2,2,2)
4  x3 <- c(2,2,3,3,3,3,3,3,3,3,4,4,4,4,4)
5  x4 <- c(1,2,2,2,1.7,2.5,2,2.5,2,2,2,2,3,3,3)
6  k <- 4
7  r <- 1
8  df = data.frame(y, x1, x2, x3, x4)
9  df
10 model <- lm(y~x1+x2+x3+x4, data = df)
11 model
12 summary(model)
13 aov(model)
14 summary(aov(model))
15 sr <- (summary(aov(model)))
16 model1 <- lm(y ~ x1)
17 model1
18 aov(model1)
19 s <- (summary(aov(model1)))
20 s
21 sse1 <- s[[1]][[2]][[2]]
22 sse2 <- sr[[1]][[2]][[5]]
23 sse2
24 mse2 <- sr[[1]][[3]][[5]]
25 mse2
26 f <- round(((sse1 - sse2) / ((k - r))/mse2),1)
27 cat("test statistics is",f)
28 alpha <- 0.05
29 critical_value_f <- qf(1 - alpha, df1 = 3, df2 = 10)
30 critical_value_f
31 if(f > critical_value_f){
32   print("reject the null hypothesis")
33 }else{
34   print("cannot reject null hypothesis")
35 }
36 #The results may slightly vary due to rounding off
```

values

Chapter 14

Analysis of Categorical Data

R code Exa 14.1 Rat problem

```
1 df1 <- data.frame(green <- 20, red <- 39, blue <-
     31)
2 chisq <- chisq.test(df1)
3 n <- 90
4 p_value <- chisq$p.value
5 p_value
6 cat("since the p-value is less than 0.05 so the null
      hypothesis is rejected")
7 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

```
1 o1 <- 89
2 o2 <- 18
3 o3 <- 12
4 o4 <- 81
```

```r
 5  e1 <- 82
 6  e2 <- 20
 7  e3 <- 8
 8  e4 <- 90
 9  p1 <- 0.41
10  p2 <- 0.10
11  p3 <- 0.04
12  p4 <- 0.45
13  k <- 4
14  df <- k -1
15  alpha <- 0.100
16  x_square_observed <- ((o1 - e1) * (o1 - e1) / e1) +
        ((o2 - e2) * (o2 - e2) / e2) + ((o3 - e3) * (o3 -
        e3) / e3) + ((o4 - e4) * (o4 - e4) / e4)
17  x_square_observed
18  x_square_0.100 <-qchisq(1 - alpha, df)   #x-square
        at alpha = 0.100 using chi- square
19  x_square_0.100
20  if(x_square_observed < x_square_0.100){
21    print("p-value is greater than 0.100, do not have
          sufficient evidence to reject null h0")
22    }else{
23    print("reject the h0")
24  }
```

R code Exa 14.4 Furniture defect

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

```
9  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")
10 #the answers may slightly vary due to rounding off
       values
```

R code Exa 14.5 Flu Vaccine

```
1  survey <- matrix(c(24,9,13,289,100,565),ncol = 3,
       byrow = TRUE)
2  colnames(survey) <- c("No Vaccine","One Shot","Two
       Shots")
3  rownames(survey) <- c("Flu","No Flu")
4  survey <- as.table(survey)
5  survey
6  chisq.test(survey,correct = F)
7  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

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

Chapter 15

Nonparametric Statistics

R code Exa 15.1 Euglossine bees

```
1  species1 <- c(235,225,190,188);
2  species2 <- c(180,169,180,185,178,182);
3  n1 <-length(species1);
4  n2 <- length(species2);
5  data <- c(species1,species2);
6  rank1 <- rank(sort(data))
7  t1 <- 7+8+9+10;
8  t1_ <- n1 * (n1 + n2 + 1) - t1
9  t1_                                      #for n1=4
      and n2=6, the critical value of T at alpha = 0.05
       is 12
10 if(t1_ <= 12){
11   print("Reject the Null Hypothesis")
12 }else{
13     print("Accept the Null Hypothesis")
14   }
```

R code Exa 15.2 Kraft papers

```r
 1  standard1 <- c
        (1.21 ,1.43 ,1.35 ,1.51 ,1.39 ,1.17 ,1.48 ,1.42 ,1.29 ,1.40)

 2  treated2 <- c
        (1.49 ,1.37 ,1.67 ,1.50 ,1.31 ,1.29 ,1.52 ,1.37 ,1.44 ,1.53)

 3  n1 <- length ( standard1 )
 4  n2 <- length ( treated2 )
 5  alpha = 0.05
 6  x <- c( standard1 , treated2 )
 7  ranksum <-function (x, start ,end ){
 8    return ( sum(x[start : end ]))
 9  }
10  rank <- rank(x)
11  t1 <- ranksum (rank ,1, n1)
12  t2 <- n1 * (n1 + n2 + 1) - t1
13  if(t1 <= 82){                            #critical  value
        of  T  at  n1=n2=10  at  alpha  =  0.05  is  82
14    print (" Reject  the  hypothesis ")
15  } else {
16    print (" Insufficient  evidence  to  conclude  that  the
        treated  kraft  paper  is  stronger  than  the
        standard  paper ")
17  }
18  muo_t <- (n1 * (n1 * n2 + 1))/2
19  sigma_sqr_t <- ((n1 * n2) *(n1 + n2 +1))/12
20  sigma_t <- sqrt ( sigma_sqr_t )
21  z <- (t1 - muo_t )/ sigma_t ;
22  p_value <- 0.5 - 0.4292
23  p_value
24  if( p_value <= alpha ){
25    print (" Reject  the  hypothesis ")
26  } else {
27    print (" Cannot  conclude  the  treated  kraft  paper  is
        stronger  than  the  standard  paper ")
28  }
```

R code Exa 15.3 Defective Electrical Fuses

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

R code Exa 15.4 Employee accident rates

```
1 x <- 63;
2 n <- 100;
3 alpha <- 0.05;
4 z <- (x - 0.5 * n)/(0.5 * sqrt(n));
5 z_expected <- qnorm(1-alpha/2)
6 if(z == z_expected ){
7 print("Accept the null hypothesis")
8 }else{
9   print("Reject the null hypothesis and the data
        provide sufficient evidence")
10 }
```

R code Exa 15.5 Densities of cakes

```r
1  x_A <- c(.135,.102,.098,.141,.131,.144)
2  x_B <- c(.129,.120,.112,.152,.135,.163)
3  alpha <- 0.10
4  t0 <- round(qt(1 - alpha/2,5),1)
5  boxplot(x_A - x_B,horizontal = TRUE,xlab = "
      Differences")
6  wilcox.test(x_A,x_B,paired = T)
7  t_positive <- wilcox.test(x_A,x_B,paired = T)$
      statistic
8  t0
9  t_positive
10 if(t_positive <= t0){
11   print("Reject the null hypothesis so two
         population frequency distribution of cake
         densities differ")
12 }else{
13   print("Accept the hypothesis")
14 }
```

R code Exa 15.6 Kruskal Wallis H test

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

```
14    print("sufficent evidence to indicate differences"
         )
15 }else{
16    print("there is insufficient evidence to indicate
         differences in the distributions of achievement
          test scores for the four teaching techniques")
17 }
18 #the answers may slightly vary due to rounding off
      values
```

R code Exa 15.8 Friedman Fr test

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

```
          * (k + 1))
23  fr
24  df <-  k - 1
25  alpha <- 0.05
26  x_square <- qchisq(1 - alpha, 5)     #sampling
       distribution of fr
27  if(fr > x_square){
28    print("reject null hypothesis, so the distribution
           of reaction times differ in location  for at
         least two stimuli")
29  }else{
30    print("accpept the null hypothesis ")
31  }
```

R code Exa 15.9 P value

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

R code Exa 15.10 Spearman rank correlation coefficient

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

R code Exa 15.11 Hypothesis Test of no association

```
1   xi <- c(7,4,2,6,1,3,8,5)
2   yi <- c(1,5,3,4,8,7,2,6)
3   alpha <- 0.05
4   cor.test(xi, yi)
5   rs <- cor.test(xi, yi)$estimate
6   cat("rs is",rs)
7   p_value <- cor.test(xi, yi)$p.value
8   if(p_value <= alpha){
9     print("null hypothesis is rejected")
10  }else{
11    print("null hypothesis is accepted")
12  }
```