R Textbook Companion for Statistics - Concepts and Applications by Nabendu Pal and Sahadeb Sarkar[1]

Created by
M. Sai Anand
Bachelor of Technology
Computer Science and Engineering
Kalasalingam Academy of Research and Education
Cross-Checked by
R TBC Team

June 8, 2021

[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: Statistics - Concepts and Applications

Author: Nabendu Pal and Sahadeb Sarkar

Publisher: Phi Learning Private Limited, Sonepat, Harayana.

Edition: 2

Year: 2013

ISBN: 9788120334458

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 2

Organizing a Raw Dataset

R code Exa 2.1 Relative frequency of dataset on voters preference

```
#Page Number: 9

dataset<- read.csv("dataset.csv", header = T)
print(dataset)
A <- dataset[ dataset$Vote == 'A', ]
AF_A <-length(A$Population)
B <- dataset[ dataset$Vote == 'B', ]
AF_B <-length(B$Population)
C <- dataset[ dataset$Vote == 'C', ]
AF_C <-length(C$Population)
Total_Freq <- length(dataset$Population)
RF_A <- AF_A / Total_Freq
print(RF_A)
RF_B <- AF_B / Total_Freq
print(RF_B)
RF_C <- AF_C / Total_Freq
print(RF_C)
Total_RF = RF_A+RF_B+RF_C
print(Total_RF)
```

R code Exa 2.2 Relative frequency of dataset on gender painkiller study

```
#Page Number: 10

Tot_F = 0.60* 500
Tot_M = 500 - Tot_F
F1 = 0.45 * 300
F2 = 300 - F1
MED_1 = 0.50 * 500
MED_2 = 500 - MED_1
M1 = MED_1 - F1
M2 = Tot_M - M1
req <- data.frame(GENDER = c("M" ,"F"), Medicine_1 =
    c(M1, F1), Medicine_2 = c(M2, F2))
RF_M1 = M1 / 500
RF_M2 = M2 / 500
RF_F1 = F1 / 500
RF_F2 = F2 / 500
genfreq <- data.frame( Catagory = c('M1','M2','F1','
   F2' ), Frequency = c(M1,M2,F1,F2), Relative_
   Frequency = c(RF_M1,RF_M2,RF_F1,RF_F2)  )
print(genfreq)
cat(paste("Total Frequency:",sum(genfreq$Frequency),
   "\nTotal Relative Frequency:",sum(genfreq$
   Relative_Frequency)))
```

R code Exa 2.3 Relative frequency of dataset on number of individuals treated at ER per day

```
#Page Number: 12

dataset <- read.csv("data.csv", header = T)
```

```
minc1 <- min(dataset$C1)
minc2 <- min(dataset$C2)
minc3 <- min(dataset$C3)
minc4 <- min(dataset$C4)
minc5 <- min(dataset$C5)
minc6 <- min(dataset$C6)
minc7 <- min(dataset$C7)
minc8 <- min(dataset$C8)


maxc1 <- max(dataset$C1)
maxc2 <- max(dataset$C2)
maxc3 <- max(dataset$C3)
maxc4 <- max(dataset$C4)
maxc5 <- max(dataset$C5)
maxc6 <- max(dataset$C6)
maxc7 <- max(dataset$C7)
maxc8 <- max(dataset$C8)

Column_Min = c(minc1,minc2,minc3,minc4,minc5,minc6,
   minc7,minc8)
print(Column_Min)

Column_Max = c(maxc1,maxc2,maxc3,maxc4,maxc5,maxc6,
   maxc7,maxc8)
print(Column_Max)

Up_Lim <- max(dataset) + 0.5
Low_Lim <- min(dataset) - 0.5
L <- Up_Lim - Low_Lim


k <- 5
l <- L/k


breaks = seq( Low_Lim, Up_Lim, by= l )
```

```


frequency.cut<- cut(x<- c(40,8,26,21,36,35,42,32,46,30,42,53,28,22,40,6,17,28,19,38,13,25,
,breaks)

dataset.freq <- table(frequency.cut)

cbind(dataset.freq)


Relative_Freq <- dataset.freq/ 40
print(Relative_Freq)
cat("Total Frequency:",sum(dataset.freq),"\nTotal Relative Frequency:",sum(Relative_Freq))
```

R code Exa 2.4 Relative frequency of dataset on for the oak tree circumference

```
#Page Number: 13

circum_table <- data.frame(Circumference = c('10-24','25-39', '40-54', '55-69' ), No._Of_Tree = c(8,22,53,17))
mod_class_table <- data.frame(Class_modified = c('9.5-24.5','24.5-39.5', '39.5-54.5', '54.5-69.5'))
REL_Freq <- circum_table$No._Of_Tree / 100
Modfreq_Table <- data.frame(mod_class_table, Frequency = circum_table$No._Of_Tree, Relative_Frequency = REL_Freq)
final <- data.frame(Class = mod_class_table, Frequency = circum_table$No._Of_Tree, Rel.Frequency = REL_Freq)
print(final)
```

```
cat("Total Frequency:",sum(circum_table$No._Of_Tree),"\nTotal Relative Frequency:",sum(REL_Freq))
```

Chapter 3

Pictorial Representation of a Dataset

R code Exa 3.1 Stem leaf display of commuting distance data

```
#Page Number: 33

data <- data.frame(dist = c(13,7,12,6,34,14,47,25,45,2,13,26,10,8,1,14,41,10,3,21,8,13,28,24)
)
stem(data$dist)
```

R code Exa 3.2 Time plot of burglaries between 1981 and 1990

```
#Page Number: 34

library(ggplot2)
data <- data.frame(YEAR = c('1981','1982','1983','1984','1985','1986','1987','1988','1989','1990'), Theft = c(24,18,16,27,30,34,37,38,40,39) )
ggplot(data, aes(YEAR, Theft, group = 1)) +
```

```
  geom_point(shape = 0) +
  geom_line()+
  xlab("Year")+ ylab("Theft")
```

Chapter 4

Summarizing a Raw Dataset

R code Exa 4.1 Computation of mode for distinct entry

```
#Page Number: 55

library(plyr)

dataset <- data.frame(Distinct_entry = c
    (8,9,0,9,1,4,3,7))
n <- length(dataset$Distinct_entry)


y <- count(dataset,'Distinct_entry')
print(y)


z <- table(dataset)
cat("Mode is",names(z)[which(z==max(z))] )
```

R code Exa 4.2 Computation of mode for dataset

```
#Page Number: 56

library(plyr)

dataset <- data.frame(Distinct_entry = c
    (-1,-2,3,0,3,-1,4,2,7,5))
n <- length(dataset$Distinct_entry)


y <- count(dataset,'Distinct_entry')
print(y)


z <- table(dataset)
cat("Mode is",names(z)[which(z==max(z))] )
```

R code Exa 4.3 Median of the given dataset

```
#Page Number:56

data <- c(2,-1,0,4,7,7,8,0,11)


ascend_order <- sort(data)


M <- median(ascend_order)
print(M)
```

R code Exa 4.4 Median of the given dataset

```
#Page Number: 56

```

```
data <- c(4,7,9,1,3,2,1,1,7,0)


ascend_order <- sort(data)

M <- median(ascend_order)
print(M)
```

R code Exa 4.5 Median of the 2 datasets

```
#Page Number: 56

data1 <- c(0,1,1,1,10)
data2 <- c(-27,1,1,2,1000)


ascend_order1 <- sort(data1)
ascend_order2 <- sort(data2)


M1 <- median(ascend_order1)
M2 <- median(ascend_order2)
print(M1)
print(M2)
```

R code Exa 4.6 Median of the 2 datasets

```
#Page Number: 56

data1 <- c(0,1,1,1,1,1,2)
data2 <- c(-100,-100,-100,1,100,100,200)


```

```
ascend_order1 <- sort(data1)
ascend_order2 <- sort(data2)


M1 <- median(ascend_order1)
M2 <- median(ascend_order2)
print(M1)
print(M2)
```

R code Exa 4.7 Computation of mean for the dataset

```
#Page Number : 57

data <- c(5,-1,0,2,4)
mean(data)
```

R code Exa 4.8.i Construction of datasets with same medians but different means

```
#Page Number: 57

dataset1 <- c(5,6,6,9,9)
dataset2 <- c(3,6,6,6,9)

cat("Mean of dataset 1:",mean(dataset1),"\nMean of dataset 2:",mean(dataset2))
cat("Median of dataset 1:",median(dataset1),"\nMedian of dataset 2:",median(dataset2))
```

R code Exa 4.8.ii Construction of datasets with same means but different medians

```r
#Page Number: 57

dataset1 <- c(1,1,3,9,11)
dataset2 <- c(1,1,1,11,11)


cat("Mean of dataset 1:",mean(dataset1),"\nMean of dataset 2:",mean(dataset2))
cat("Median of dataset 1:",median(dataset1),"\nMedian of dataset 2:",median(dataset2))
```

R code Exa 4.8.iii Construction of datasets with same medians but different modes

```r
#Page Number: 57

dataset1 <- c(1,1,1,2,2)
dataset2 <- c(1,1,3,4,5)


cat("Median of dataset 1:",median(dataset1),"\nMedian of dataset 2:",median(dataset2))


x <- table(dataset1)
y <- table(dataset2)

cat("Mode of dataset 1:",names(x)[which(x==max(x))],"\nMode of dataset 2:",names(y)[which(y==max(y))])
```

R code Exa 4.8.iv Construction of datasets with same means medians and modes

```
#Page Number: 57

dataset1 <- c(1,2,2,2,3)
dataset2 <- c(0,2,2,2,4)


x <- table(dataset1)
cat("Mean:",mean(dataset1),"Median:",median(dataset1
    ),"Mode:",as.numeric(names(x)[which(x==max(x))]))

y <- table(dataset2)
cat("Mean:",mean(dataset2),"Median:",median(dataset2
    ),"Mode:",as.numeric(names(y)[which(y==max(y))]))
```

R code Exa 4.9 Range length of elementary school children dataset

```
#Page Number: 57

data <- c(72,113,97,91,105,88,119)

x <- min(data)
y <- max(data)

RL <- y-x
print(paste('The Range Length is:', RL))
```

R code Exa 4.10 Range and IQR of given dataset

```
#Page Number: 58

data <- c(5,7,-1,0,2,9,9,4,3,11)

x <- min(data)
```

```
y <- max(data)
cat('The range is:',x,",",y)

IQR_data <- quantile(data,type = 2)
cat('The IQR is:',IQR_data[2],",",IQR_data[4] )
```

R code Exa 4.11 Range and IQR of given dataset

```
#Page Number: 58

data <- c(5,7,-1,0,2,9,9,4,3,11,6)

x <- min(data)
y <- max(data)
RL <- y - x
print(paste('The range is:', RL))


IQR_data <- IQR(sort(data), type = 1)
print(paste('The IQRL is:',IQR_data ))
```

R code Exa 4.12 Variance of the given dataset

```
#Page Number: 59

data <- c(-2,3,0,2,2)

y <- var(data)
print(paste('The Variance is :',y ))
```

R code Exa 4.13 Box whisker plots for males and females working in telecom sector

```
#Page Number: 60

data <- read.csv("data.csv", header = T)


boxplot(data$MALES,main = "Box-whisker plot for Male
   ", ylim = c(10,80) )
boxplot(data$FEMALES,main = "Box-whisker plot for
   Female", ylim = c(10,80) )
```

Chapter 5

Summarizing an Organized Dataset

R code Exa 5.1 Computation of central tendency and variability for the grouped dataset on the number of patients treated per day at the emergency room

```
#Page Number: 68

data <-c(40,8,26,21,36,35,42,32,46,30,42,53,28,22,40,6,17,28,19,38,13,25,
n <- length(data)
interval <- seq(5.5, 60.5 , by = 11)
class <- cut(data, interval)
frequency <- cbind(table(class))

RF <- frequency / n
CRF <- cumsum(frequency) / n


midpoints <- function(x, dp=2){
  lower <- as.numeric(gsub(",.*","",gsub("\\(|\\[|\\)|\\]","", x)))
```

```
  upper <- as.numeric(gsub(".*,","",gsub("\\(|\\[|\\)|\\]","", x)))
  return(round(lower+(upper-lower)/2, dp))
}

Midpoint <- sort(unique(midpoints(class)))
Midfreq <- Midpoint * RF
xbar <- sum(Midfreq)
k <- sum((Midpoint - xbar)^2 * frequency)
variance <- k / (n-1)
sd <- sqrt(variance)
cat("Variance :",round(variance,2),"& Standard Deviation:",round(sd,2))

Q1 <- interval[2] + ((Midpoint[1] / RF[2]) * (0.25 - CRF[1] ))
Q2 <- interval[3] + ((Midpoint[1] / RF[3]) * (0.5 - CRF[2] ))
Q3 <- interval[3] + ((Midpoint[1] / RF[3]) * (0.75 - CRF[2] ))
cat("Q1=",round(Q1,1),"Q2=",round(Q2,2),"Q3=",round(Q3,2))
```

R code Exa 5.2 Computation of variability for the randomly generated grouped dataset with the given class and frequency on the population of all working mothers in a large city

```
#Page Number: 72

random_input_data <- c(6,11,21,39,45,9,18,29,36,41,2,18,27,7,19,21,1,13,29,4,20,28,7,1,
n <- length(random_input_data)

interval <- seq( from = 0.5, to = 50.5, by = 10)
```

```
class <- cut(random_input_data, interval)
frequency <- cbind(table(class))


RF <- frequency / n



midpoints <- function(x, dp=2){
  lower <- as.numeric(gsub(",.*","",gsub("
     \\(|\\[|\\)|\\]","", x)))
  upper <- as.numeric(gsub(".*,","",gsub("
     \\(|\\[|\\)|\\]","", x)))
  return(round(lower+(upper-lower)/2, dp))
}

a <-midpoints(class)

Midpoint <- unique(a)

Mid_Freq <- Midpoint * frequency

b <- sum(Mid_Freq)

m <-  b/n

S <- (Midpoint - m)^2 * frequency
Var <- (sum(S)) / (n-1)
print(paste('Variance is:',round(Var,2)))
SD <- sqrt(Var)
print(paste('Standard Deviation is:',round(SD,2)))
```

Chapter 6

Concepts of Probability

R code Exa 6.1 Probability of getting head in tossing a fair coin

```
#Page Number: 81 & 83(cont.)

library(prob)
a <- tosscoin(1)
c <- length(a[a$toss1 == 'H']) / length(a$toss1)
print(paste('The probability is :',c))
```

R code Exa 6.3 Probability of getting face value less than 3 in rolling a die

```
#Page Number: 83

library(prob)
a <- rolldie(1)
c <- length(a[a$X1 < 3,]) / length(a$X1)
print(round(c,4))
```

R code Exa 6.6 Probability of getting a queen card from a deck of 52 cards

```
#Page Number: 83

library(prob)
a <- cards()
b <- subset(a, rank == 'Q')
c <- length(b$rank) /length(a$rank)
print(paste(' The probability is :',round(c,3)))
```

R code Exa 6.7 Probability of getting 2 heads in tossing a fair coin twice in a row

```
#Page Number: 83

library(prob)
S <- tosscoin(2,makespace = T)
P <- S[1,3]
print(P)
```

R code Exa 6.8 Probability of getting 2 heads in tossing a fair coin thrice in a row

```
#Page Number: 83

library(prob)

S <- tosscoin(3, makespace = T)
H1 <- S[2,'probs']
H2 <- S[3,'probs']
H3 <- S[5,'probs']
Pro <- H1+H2+H3
print(paste('The probability is:',Pro))
```

R code Exa 6.9.a Probability of getting red marbles from a box of red and blue marbles when there are 30 red and 40 blue marbles

```
#Page Number: 84

library(prob)
S <- rep(c("Red", "Blue"), times = c(30,40))
P <- urnsamples(S, size = 1, replace = F, ordered =
    F)


R <- P[P$out == 'Red',]

P_R = length(R) / length(P$out)
print(round(P_R,4))

#The answer provided in the textbook is wrong.
```

R code Exa 6.9.b Probability of getting red marbles from a box of red and blue marbles when composition of box is unknown

```
#Page Number: 84

unknown_comp <- c(3,3,8,24.5,53,112,165,226)
n <- c(5,10,20,50,100,200,300,400)

r_n <- unknown_comp / n
print(r_n)
```

R code Exa 6.10 Getting 4 or more face values from the unbalanced die with known sample points

```
#Page Number: 85

library(prob)
library(MASS)
S <- probspace(rolldie(1),probs = c(1/12,3/12,1/12,3
   /12,1/12,3/12))

P_G1 <- fractions( S[4,"probs"] )
P_G2 <- fractions( S[5,"probs"] )
P_G3 <- fractions( S[6,"probs"] )
P_G4 <- P_G1+P_G2+P_G3
print(paste('The Probability is:',P_G4))
```

R code Exa 6.11.a Probability of students failed in maths test from a class of 120 students

```
#Page Number: 85
library(MASS)
Total <- 120
Maths101 <- 65
French101 <- 75
Both <- 35

mathpassprob <- fractions( Maths101 / Total )
mathfailprob <- 1 - mathpassprob
print(paste('Probability:',mathfailprob))

#The answer may vary due to difference in
   representation
```

R code Exa 6.11.b Probability of students passed in both test from a class of 120 students

```r
#Page Number: 85

library(MASS)
Total <- 120
Maths101 <- 65
French101 <- 75
Both <- 35

bothpassprob <- fractions( Both / Total )
print(paste('Probability would be:',bothpassprob))

#The answer may vary due to difference in representation
```

R code Exa 6.11.c Probability of students failed in both test from a class of 120 students

```r
#Page Number: 85

library(MASS)
Total <- 120
Maths101 <- 65
French101 <- 75
Both <- 35

mathpassprob <- fractions( Maths101 / Total )
frenchpassprob <- fractions( French101 / Total )
bothpassprob <- fractions( Both / Total )
allpassprob <- mathpassprob + frenchpassprob - bothpassprob
allfailprob <- 1 - allpassprob
print(paste('Probability :',allfailprob ))
```

```
#The answer may vary due to difference in representation
```

R code Exa 6.12 Probability of getting a card that is either red or a king from a deck of 52 cards

```
#Page Number: 86

library(prob)
library(MASS)

card <- cards()


diamond <- subset(card, suit == 'Diamond')
heart <- subset(card, suit == 'Heart')
E1 <- length(diamond$suit) + length(heart$suit)
probE1 <- fractions(E1 / length(card$suit))


E2 <-subset(card, rank == 'K')
probE2 <- fractions( length(E2$rank) / length(card$suit))


diaking <- intersect(diamond, E2)
heartking <-intersect(heart, E2)
intersec <- length(diaking$suit) + length(heartking$suit)
probintersec <- fractions(intersec / length(card$suit))

probunion <- probE1 + probE2 - probintersec
print(paste('The probability is:',probunion))
```

```
#The answer may vary due to difference in
    representation.
```

R code Exa 6.13 Probability of getting atleast 1 head in tossing a fair coin four times in a row

```
#Page Number: 86

library(prob)
library(MASS)

S <- iidspace(c("H","T"), ntrials = 4)

E <- S[16,'probs']

probathead <- 1 - fractions(E)
print(paste('The probability is:',probathead))
```

R code Exa 6.14 Probability of drawing a card that is King and knowing that there is no Queen in deck of 52 cards

```
#Page Number: 86
library(prob)
library(MASS)
card <- cards(makespace = T)

condprob <- fractions( Prob(card,rank == 'K' ,given
    = rank != 'Q') )
print(paste(" The Probability is:", condprob))
```

R code Exa 6.15 Probability of a student passed in maths test given that he or she is passed in atleat 1 test from a class of 120 students

```
#Page Number: 87

library(prob)
library(MASS)

Total <- 120
Maths101 <- 65
French101 <- 75
Both <- 35

A <- Maths101
P_A <- fractions( A / Total)

mathpassprob <- fractions( Maths101 / Total )
frenchpassprob <- fractions( French101 / Total )
bothpassprob <- fractions( Both / Total )
P_B <- mathpassprob + frenchpassprob - bothpassprob

condprob = P_A / P_B
print(paste("The probability is:", condprob))
#The answer may vary due to difference in
    representation.
```

R code Exa 6.16 Independent events of cards being drawn at random that are either red or a king from deck of 52 cards

```
#Page Number: 88

library(prob)
library(MASS)

card <- cards()
```

```
diamond <- subset(card, suit == 'Diamond')
heart <- subset(card, suit == 'Heart')


A <- length(diamond$suit) + length(heart$suit)
probA <- fractions(A / length(card$suit))

B <-subset(card, rank == 'K')
probB <- fractions( length(B$rank) / length(card$
    suit))

cat(paste("P(A)=",probA,"\nP(B)=",probB))

diaking <- intersect(diamond, B)
heartking <-intersect(heart, B)
intersec <- length(diaking$suit) + length(heartking$
    suit)

probintersec <- fractions(intersec / length(card$
    suit))
cat(paste(probintersec))
```

R code Exa 6.17 Probability of a ship departing on time

```
#Page Number: 88

library(prob)

P_A <- 0.81

P_B <- 0.77

condAB <- 0.91

```

```
bayesBA <- (condAB * P_B) / P_A

print(paste('The probability is:',round(bayesBA,3)))
```

R code Exa 6.18 Probability of gettting a blue marble during first draw given that the second draw gave a blue marble from 2 boxes using bayes theorem

```
#Page Number: 89

library(prob)

S1 <- rep(c("Red", "Blue"), times = c(4,3))
Box1 <- urnsamples(S1, size = 1, replace = F, ordered = F)
dfBox1 <- c(Box1)
totalred1 <- sum(dfBox1$out == "Red")
totalblue1 <- sum(dfBox1$out == "Blue")
prob_totalred1 <- totalred1 / length(dfBox1$out)
prob_totalblue1 <- totalblue1 / length(dfBox1$out)

S2 <- rep(c("Red", "Blue"), times = c(10,20))
Box2 <- urnsamples(S2, size = 1, replace = F, ordered = F)
dfBox2 <- c(Box2)
totalred2 <- sum(dfBox2$out == "Red")
totalblue2 <- sum(dfBox2$out == "Blue")

f_red_totalred1 <- totalred1 - 1
f_red_totalred2 <- totalred2 + 1
red_totalmarble.in.1 <- length(dfBox1$out) - 1
red_totalmarble.in.2 <- length(dfBox2$out) + 1
f_prob_totalred1 <- f_red_totalred1 / red_totalmarble.in.1
f_prob_totalred2 <- f_red_totalred2 / red_
```

```
totalmarble.in.2
f_prob_totalblue1 <- totalblue1 / (length(dfBox1$out) - 1)
f_prob_totalblue2 <- totalblue2 / (length(dfBox2$out) + 1  )

s_blue_totalblue1 <- totalblue1 - 1
s_blue_totalblue2 <- totalblue2 + 1
blue_totalmarble.in.1 <- length(dfBox1$out) - 1
blue_totalmarble.in.2 <- length(dfBox2$out) + 1
s_prob_totalblue1 <- s_blue_totalblue1 / blue_totalmarble.in.1
s_prob_totalblue2 <- s_blue_totalblue2 / blue_totalmarble.in.2

bayes <- (s_prob_totalblue2 * prob_totalblue1 ) /( ( s_prob_totalblue2 * prob_totalblue1 )  + ( f_prob_totalblue2 * prob_totalred1 ))
print(paste("The probability is :",round(bayes,2)))
```

Chapter 7

Random Variables

R code Exa 7.1 Random variable operations on a rolling die

```
#Page Number: 93, 94(contd.), 95(contd.), 97(contd.), 99(contd.) .

library(prob)
library(MASS)
die <- rolldie(1,makespace = T)
rx <- die$X1
print(rx)

X_1 <-  die[1,2]
X_2 <-  die[2,2]
X_3 <-  die[3,2]
X_4 <- die[4,2]
X_5 <-  die[5,2]
X_6 <-  die[6,2]

print(paste('P(X = 1) =',fractions(X_1)))
print(paste('P(X = 2) =',fractions(X_2)))
print(paste('P(X = 3) =',fractions(X_3)))
print(paste('P(X = 4) =',fractions(X_4)))
print(paste('P(X = 5) =',fractions(X_5)))
```

```
print(paste('P(X = 6) =',fractions(X_6)))

E_X <-  (die[1,1] * X_1 ) + ( die[2,1] * X_2 ) + ( die[3,1] * X_3 ) + ( die[4,1] * X_4 ) + ( die[5,1] * X_5 ) + ( die[6,1] * X_6 )
print(E_X)

var <- ( ((die[1,1]-E_X)^2) * X_1 ) + ( ((die[2,1]-E_X)^2) * X_2 ) + ( ((die[3,1]-E_X)^2) * X_3 ) + ( ((die[4,1]-E_X)^2) * X_4 ) + ( ((die[5,1]-E_X)^2) * X_5 ) + ( ((die[6,1]-E_X)^2) * X_6 )
print(paste('Variance = ',round(var,4)))
sd <- sqrt(var)
print(paste('Standard Deviation = ',round(sd,4)))
```

R code Exa 7.2 Random variable operations on tossing a coin twice

```
#Page Number: 93, 94(contd.), 95(contd.), 97(contd.), 99(contd.) .

library(prob)
library(MASS)

toss <- iidspace(c("H","T"), ntrials = 2)

X_0 <- toss[4,'probs']
X_1 <- toss[2,'probs'] + toss[3,'probs']
X_2 <- toss[1,'probs']
print(paste('P(X = 0) =',fractions(X_0)))
print(paste('P(X = 1) =',fractions(X_1)))
print(paste('P(X = 2) =',fractions(X_2)))

E_X <- (0 * X_0) + (1 * X_1) + (2 * X_2)
print(E_X)

```

```
var <- (((0 - E_X)^2)*(X_0)) + (((1 - E_X)^2)*(X_1)) + (((2 - E_X)^2)*(X_2))
print(paste('Variance = ',var))
sd <- sqrt(var)
print(paste('Standard Deviation',round(sd,4)))
```

R code Exa 7.3 Random variable operations in tossing a fair coin twice and then rolling a balanced die

```
#Page Number: 94, 95(contd.), 97(contd.), 99(contd.).

library(MASS)

S <- data.frame(TOSS = c('H','H','H','H','H','H','T','T','T','T','T','T'),DIE = c(1,2,3,4,5,6,1,2,3,4,5,6))

X_0 <- sum(S$TOSS == "T") / length(S$TOSS)
X_1 <- sum(S$TOSS == "H") / length(S$TOSS)
print(paste('P(X = 0) =',fractions(X_0)))
print(paste('P(X = 1) =',fractions(X_1)))

Y_1 <- sum(S$DIE == "1") / length(S$DIE)
Y_2 <- sum(S$DIE == "2") / length(S$DIE)
Y_3 <- sum(S$DIE == "3") / length(S$DIE)
Y_4 <- sum(S$DIE == "4") / length(S$DIE)
Y_5 <- sum(S$DIE == "5") / length(S$DIE)
Y_6 <- sum(S$DIE == "6") / length(S$DIE)

mu_X <- (0 * X_0) + (1 * X_1)
print(mu_X)
mu_Y <- (1 * Y_1) + (2 * Y_2) + (3 * Y_3) + (4 * Y_4) + (5 * Y_5) + (6 * Y_6)
print(mu_Y)
```

```
varX <- (((0- mu_X)^2) * (X_0)) + (((1 - mu_X)^2) * (X_1))
VarY <- (((1 - mu_Y)^2) * (Y_1)) + (((2 - mu_Y)^2) * (Y_2)) +(((3 - mu_Y)^2) * (Y_3)) +(((4 - mu_Y)^2) * (Y_4)) +
        (((5 - mu_Y)^2) * (Y_5)) +(((6 - mu_Y)^2) * (Y_6))
print(paste('Variance:',varX,' & ',round(VarY,4)))

sdX <- sqrt(varX)
sdY <- sqrt (VarY)
print(paste('Standard Deviation:',sdX,'&',round(sdY,4)))

#The answer provided in the textbook is wrong
```

R code Exa 7.4 Random variable operations on a sample of 3 voters

```
#Page Number: 94, 95(contd.), 97(contd.), 99(contd.).

library(prob)
library(MASS)

vote <- iidspace(c("Y","N"), ntrials = 3)
print(vote)

X_0 <- vote[8,"probs"]
X_1 <- vote[4,"probs"] + vote[6,"probs"] + vote[7,"probs"]
X_2 <- vote[2,"probs"] + vote[3,"probs"] + vote[5,"probs"]
X_3 <- vote[1,"probs"]

```

```
print(paste("P(X = 0) =",fractions(X_0)))
print(paste("P(X = 1) =",fractions(X_1)))
print(paste("P(X = 2) =",fractions(X_2)))
print(paste("P(X = 3) =",fractions(X_3)))

E_X <- (0 * X_0) + (1 * X_1) + (2 * X_2) + (3 * X_3)
print(E_X)

var <- (((0 - E_X)^2)*(X_0)) + (((1 - E_X)^2)*(X_1))
    + (((2 - E_X)^2)*(X_2))  + (((3 - E_X)^2)*(X_3))
print(paste('Varience = ',var))
sd <- sqrt(var)
print(paste('Standard Deviation',round(sd,2)))
```

R code Exa 7.6 To find all possible grade pairs for a certain mark obtained

```
#Page Number: 95

library(tidyverse)
grade_combo <- function(x) {
  dataframe <- data.frame(g1 = c("A", "B", "C", "D",
      "F"),
                          g2 = c("A", "B", "C", "D", "F"))
                             %>%
    expand.grid(.) %>%
    data.frame(.) %>%
    mutate(p = case_when(g1 == "A" ~ 4,
                               g1 == "B" ~ 3,
                               g1 == "C" ~ 2,
                               g1 == "D" ~ 1,
                               TRUE ~ 0),
           p2 = case_when(g2 == "A" ~ 4,
                               g2 == "B" ~ 3,
                               g2 == "C" ~ 2,
                               g2 == "D" ~ 1,
```

```
                                    TRUE ~ 0),
            GPA = (p + p2)/2)


  res <- dataframe %>%
    filter(GPA == x)

  return(res = res)
}

grade <- grade_combo(2.5)

print(grade[c(1,2)])
```

Chapter 8

Binomial Experiments

R code Exa 8.1 Probability of answering correctly in a pop quiz

```
#Page Number: 103, 106(contd.) .

library(prob)
library(MASS)
see <- iidspace(c("C","I"), ntrials = 7)
options(max.print = 10000)
print(see)

size <- 7
prob <- 1/2

X_0 <- dbinom(0,size,prob)
X_1 <- dbinom(1,size,prob)
X_2 <- dbinom(2,size,prob)
X_3 <- dbinom(3,size,prob)
X_4 <- dbinom(4,size,prob)
X_5 <- dbinom(5,size,prob)
X_6 <- dbinom(6,size,prob)
X_7 <- dbinom(7,size,prob)

print(paste('P(X = 0) =',round(X_0,3)))
```

```
print(paste('P(X = 1) =',round(X_1,3)))
print(paste('P(X = 2) =',round(X_2,3)))
print(paste('P(X = 3) =',round(X_3,3)))
print(paste('P(X = 4) =',round(X_4,3)))
print(paste('P(X = 5) =',round(X_5,3)))
print(paste('P(X = 6) =',round(X_6,3)))
print(paste('P(X = 7) =',round(X_7,3)))
```

R code Exa 8.2 Probability of finding mortar shell explosion on enemy target

```
#Page Number: 104, 107(contd.).

fprob <- 0.10
size <- 20
sprob <- 1 - fprob

Xles14 <- dbinom(0:14,size = size,prob = sprob)
probXles14 <- sum(Xles14)
more15 <- 1 - probXles14
print(round(more15,3))
```

R code Exa 8.3 Probability of at least 1 sensor detecting gas leak

```
#Page Number: 104, 107(contd.).

sprob <- 0.80
size <- 5
X_0 <- dbinom(0,size, sprob)
res <- 1 - X_0
print(round(res,1))
```

R code Exa 8.4 Probability of fire accident due to arson

```
#Page Number: 104, 108(contd.).

sprob <- 0.25
size <- 10


less4 <- dbinom(0:4,size,sprob )
res <- sum(less4)
print(round(res,3))
```

R code Exa 8.5 Probability of members who are seeking capital punsihment

```
#Page Number: 104, 108(contd.).

probagainst <- 0.70
size <- 12
probfavour <- 1- probagainst

less6 <- dbinom(0:6, size, probfavour)
probless6 <- sum(less6)
more7 <- 1 - probless6
print(round(more7,3))
```

R code Exa 8.6 Tossing a fair coin 15 times

```
#Page Number: 104
```

```

library(prob)

a <- tosscoin(15)
n <- ncol(a)
print(n)

print(levels(a$toss1))

b <- tosscoin(1)
c <- length(b[b$toss1 == 'H']) / length(b$toss1)
print(c)
```

Chapter 9

Normal Curve and Normal Distribution

R code Exa 9.1 Percentage of children facing difficulties with new teaching method

```
#Page Number: 119, 123 (contd.) .

IQlev <- 95
prevm <- 112
prevsd <- 7.8

res<- pnorm(IQlev, mean = prevm, sd = prevsd)
print(round(res,4))
```

R code Exa 9.2.a Probability for a standard normal random variable

```
#Page Number: 122


a <- 1.5
```

```
res <- 0.5 - pnorm(a, lower.tail = F)
print(round(res,4))
```

R code Exa 9.2.b Probability for a standard normal random variable

```
#Page Number: 122

a <- -2.33

res <- 1 - (2 *pnorm(a, lower.tail = TRUE ))
print(round(res,4))
```

R code Exa 9.2.c Probability for a standard normal random variable

```
#Page Number: 122

a <- 1.69

res <- 1 - pnorm(a, lower.tail = FALSE)
print(round(res,4))
```

R code Exa 9.2.d Probability for a standard normal random variable

```
#Page Number: 122.

a <- -1.25

res <- 1 - pnorm(a)
print(round(res,4))
```

R code Exa 9.3 Percentage of day time temperature between 80 to 100

```r
#Page Number: 123

mean <- 89
sd <- 6.9
x1 <- 80
x2 <- 100

a <- (x1 - mean)/sd
b <- (x2 - mean)/sd

res <- pnorm(round(b,2)) - pnorm(round(a,2))
print(round(res,4))
```

Chapter 10

Applications of the Normal Distribution

R code Exa 10.1 Probability of 6 or more people suffering from hypothermia

```
#Page Number: 131

prob <- 0.75
n <- 8

res <- dbinom(6:8,n,prob)
print(sum(round(res,3)))
```

R code Exa 10.2 Probability of 80 or less people suffering from hypothermia

```
#Page Number: 132, 133(contd.).


prob <- 0.75
```

```
n <- 120


mu <- n*prob
sigma <- round(sqrt((n*prob)*(1-prob)),4)

res <- round(pnorm(80 + 0.5, mean = mu, sd = sigma),4)
print(res)
```

R code Exa 10.3.a Probability of 640 or more students passing the exam

```
#Page Number: 134

prob <- 0.57
n <- 950

mu <- n*prob
sigma <- round(sqrt((n*prob)*(1-prob)),4)

res <- pnorm(640 - 0.5, mean = mu, sd = sigma)
print(floor(res))
```

R code Exa 10.3.b Probability of 600 or fewer students passing the exam

```
#Page Number: 134

prob <- 0.57
n <- 950

mu <- n*prob
sigma <- round(sqrt((n*prob)*(1-prob)),4)

```

```
res <- pnorm(600 + 0.5, mean =  mu, sd = sigma)
print(round(res,1))
```

R code Exa 10.3.c Probability of students passing the exam between 585 and 625

```
#Page Number: 134

prob <- 0.57
n <- 950

mu <- n*prob
sigma <- sqrt((n*prob)*(1-prob))

X_grt585 <- ((585 - 0.5 ) - mu) / sigma
X_les625 <- ((625 + 0.5 ) - mu) /sigma

res<- pnorm(X_grt585)
print(round(res,1))
```

R code Exa 10.4 Probability of sample batteries having average life of 52 or less

```
#Page Number: 137

mu_x <- 54
n <- 50
sigma <- 6
sigma_x <- round(sigma / sqrt(n),2)


res<- pnorm(52,mean = mu_x, sd = sigma_x)
print(round(res,4))
```

R code Exa 10.5.a Distribution of sample average of bacteria count per ML of milk

```
#Page Number: 137

mu_x <- 2550
n <- 42
sigma <- 300

sigma_x <- round(sigma / sqrt(n),2)

print(sigma_x)
```

R code Exa 10.5.b Probability of sample having bacteria between 2400 to 2700

```
#Page Number: 137

mu_x <- 2550
n <- 42
sigma <- 300
sigma_x <- round(sigma / sqrt(n),2)


xless <- ((2400 - mu_x) / sigma_x)
xgrtr <- ((2700 - mu_x) / sigma_x)

res <- pnorm(xgrtr) - pnorm(xless)
print(round(res,4))
```

R code Exa 10.6 Probability of sample average of IQ scores not deffering from population mean by more than 1 point

```
#Page Number: 138

mu_x <- 104
n <- 64
sigma <- 5
sigma_x <- sigma / sqrt(n)

xless <- ((103 - mu_x) / sigma_x)
xgrtr <- ((105 - mu_x) / sigma_x)

res <- pnorm(xgrtr) - pnorm(xless)
print(round(res,4))
```

Chapter 11

Estimation of Population Parameters

R code Exa 11.2 Computing a 90 percent confidence interval for mean PE ratio of all utility stocks in NYSE

```
#Page Number: 151

data <- data.frame( "Utility Comp." = c(1,2,3,4,5) ,
    "P/E Ratio" = c(20.41,12.69,16.15,10.09,13.10) )

conf.interval = function(data, conf) {
   n <- length(data)
   alpha <- 1 - conf
   E <-  qt((alpha/2) + conf, df = n-1)
   xbar <-  mean(data)
   sdx <- sqrt(var(data)/length(data))
   c(xbar - E * sdx, xbar + E * sdx)
  }
print(round(conf.interval(data$P.E.Ratio, 0.90),2))
```

R code Exa 11.3 Computing 95 percent confidence interval for the mean January average temperature in Phoenix

```
#Page Number: 152.

data <- read.csv("Ex11_3.csv")

conf.interval <-  function(data, conf) {
  n <- length(data)
  alpha <- 1 - conf
  E <-  qnorm((alpha/2) + conf)
  xbar <-  mean(data)
  sdx <- sqrt(var(data)/length(data))
  c(xbar - E * sdx, xbar + E * sdx)
}
res <- conf.interval(data$Average.temperature, 0.95)
print(round(res,2))
```

R code Exa 11.4 Computing 90 percent confidence interval for the percentage of US parents whose kids favourite food is pizza

```
#Page Number: 159

n <- 1034
X <- 848

conf.interval <- function(data, conf, x){
  p <- x/data
  alpha <- 1 - conf
  SE <- sqrt(p*(1 - p)/data)
  E <- qnorm((alpha / 2) + conf)*SE
  c(p - E, p + E )
}
res <- conf.interval(n,0.90,X)
print(round(res,4))
```

Chapter 12

Hypothesis Testing for a Single Population

R code Exa 12.2 Hypothesis testing on claim of mean household income in a business district

```
#Page Number: 176, 190(contd.).


tstatvalue <-function(n,x,p){
pbar <- round(x/n,4)
del <- (sqrt(n)*(pbar -p)) / sqrt(p*(1-p))
del
}
tstat <-round(tstatvalue(72,15,0.25),4)
print(paste("Test Statistic:",tstat))

res <-  round(pnorm(-abs(tstat)),4)
print(paste("P-value :",res))
```

R code Exa 12.3 Hypothesis testing on claims of nutritionist on the rise of potato consumption

```
#Page Number: 179, 190(contd.).

n <- 46
xbar <- 53.4
mu <- 48.3
sd <- 4.12

alpha = .05
t.alpha = round(qt(1-alpha, df=n-1),4)

tstat <- round((sqrt(n)*(xbar - mu)) / sd,4)
print(paste("The test statistic",tstat))

res <-  round(pnorm(-abs(tstat)),4)
print(paste("P-value :",res))
```

R code Exa 12.4 Hypothesis testing on claims based on students councils on time spent in library

```
#Page Number: 183, 190(contd.).

hour <- c(10.4,9.1,13.6,6.4,7.9,8.6,11.8,12.6)
mu <- 12.5
xbar <- mean(hour)
s <- round(sd(hour),4)
del <- round((sqrt(length(hour))*(xbar-mu)) / s,4)

alpha5 <- 0.05
t.alpha5 <- round(qt(alpha5/2, df = length(hour) -1,
    lower.tail = F),3)

alpha1 <- 0.01
t.alpha1 <- round(qt(alpha1/2, df = length(hour) -1,
    lower.tail = F),3)

```

```
cat("Test Statistics:",abs(del))

n = length(hour)
p = pnorm(-abs(del))
print(paste("P-value:", round(p,3)))

#The answer provided in the textbook is wrong.
```

R code Exa 12.5 Hypothesis testing on claims of management on share of fruits is 50 percent

```
#Page Number: 185, 191(contd.).

data <- read.csv("EX12_5.csv")
n <- length(data$Diameters)
p <- 0.50
large <- data$Diameters > 3.0
pbar <- sum(large)/ n
del <- (sqrt(n)*(pbar -p)) / sqrt(p*(1-p))

zt <- round(qnorm(0.05),3)
print(paste("The test statistic",round(del,4)))

res <-  round(pnorm(-abs(del)),4)
print(paste("P-value :",res))
print(paste("Combined Area, P-Value:",res+res))
```

R code Exa 12.6 Hypothesis testing on claims on weight of the newly minted coins

```
#Page Number: 187, 192(contd.).


```

```
data <- data.frame("Weight" = c(2.99,3.01,3.03,3.05,3.07,3.09,3.11,3.13,3.15,3.17,3.19), "Frequency" = c(1,4,4,4,7,17,24,17,13,6,2))
n <- 100
sig <- 0.05
interval <- seq( from = 2.98, to = 3.22, by = 0.02)
subint <- cut(data$Weight,interval)
redata <- data.frame("Mid" = c(2.99,3.01,3.03,3.05, 3.07, 3.09, 3.11, 3.13, 3.15, 3.17, 3.19,3.21), "Frequency" = c(1,4,4,4,7,17,24,17,13,6,2,1) )
rf <- redata$Frequency/ sum(redata$Frequency)
xbar <- sum(redata$Mid * rf)
s <- sum(((redata$Mid-xbar)^2)*redata$Frequency)
s2 <- s / (sum(redata$Frequency)-1)


del <- round(((n-1)*s2) / (sig)^2,4)
chi <- round(qchisq(0.10, df = n-1, lower.tail = F),4)

print(paste("The test statistic:",del))

res <- pchisq(del,df = n-1,lower.tail = F)
print(paste("P-value :",round(res,3)))
```

R code Exa 12.7 Verifying Normality for the given dataset

```
#Page Number: 194

hour <- c(10.4,9.1,13.6,6.4,7.9,8.6,11.8,12.6)
xbar <- mean(hour)
s <-round(sd(hour),2)
Z.i <- sort(round(as.vector((hour - xbar)/ s),2))
B.i <- round(sort(pnorm(Z.i)),4)
A.i <- round(1:length(hour) / length(hour),4)
```

```
A.i.min1 <- round(0:(length(hour)-1) / length(hour),4)
C.i <- abs(round(A.i - B.i,4))
C.i.str <- abs(round(A.i.min1 - B.i,4))

dt <- data.frame("Zi" = Z.i, "Bi" = B.i, "Ai" = A.i, "Ai_minus_1" = A.i.min1, "Ci" = C.i, "Ci_str" = C.i.str)
print(dt)

del_ks <- round(max(C.i),4)
print(del_ks)
```

Chapter 13

Hypothesis Testing to Compare Two Populations

R code Exa 13.2 Testing the claims on painkiller which are identical in terms of mean time required for body absorption

```
#Page Number: 212, 224(contd.).

n1 <- 10
n2 <- 10
xbar1 <- 20.2
xbar2 <- 17.9
s1 <- 8.1
s2 <- 7.3

poolsample <- round( sqrt(((n1-1)*(s1)^2 + (n2-1)*(
   s2)^2 ) / (n1+n2-2) ),4)
cat("Pooled Sample :",poolsample)

tstat <- round((xbar1-xbar2) /( poolsample* sqrt((1/
   n1) + (1/n2))),4)
cat("Test Statistic value:",tstat)

alpha <- 0.05
```

```
t.alpha <- round(qt(alpha/2, df = n1+n2-2, lower.
    tail = F),3)
cat(tstat,"<",t.alpha)

res <- 2 * pt(-abs(tstat),df= n1+n2-2)
cat("The P-Value:",round(res,2))
```

R code Exa 13.3 Testing the difference between the mean stopping distances of two types of tyres

```
#Page Number: 213, 224(contd.).

n1 <- 10
n2 <- 12
xbar1 <- 51
xbar2 <- 55
s1 <- 8
s2 <- 3

tstat <- abs(round((xbar1 - xbar2) / sqrt(((s1)^2/n1
    ) + ((s2)^2/n2)),3))
q <-round( ((((s1)^2/n1)+((s2)^2/n2))^2) / ((s1)^4 /
     ((n1)^2*(n1-1)) + (s2)^4 / ((n2)^2*(n2-1))),0)
cat("Test Statistic value:",tstat)
cat("q:",q)

alpha <- 0.05
t.alpha <- round(qt(alpha/2, df = q,lower.tail = F)
    ,3)
cat(tstat,"<",t.alpha)

res <- 2 * pt(-abs(tstat),df= q)
cat("The P-Value:",round(res,2))
```

R code Exa 13.4 Testing on effects of a new hyperyension drug which changes the body temperature

```
#Page Number: 215, 225(contd.).

n <- 7
bdrug <- c(98.9,100.2,97.9,99.4,98.1,98.5,101.8)
adrug <- c(98.6,99.7,97.6,99.9,98.2,97.9,99.2)
diff <- bdrug - adrug
dbar <- round(sum(diff) / n,4)
sD <- sqrt (sum((diff - dbar)^2) / (n-1))

tstat <- round((sqrt(n)*dbar) / (sD),4)
cat("Test Statistic value:",tstat)

alpha <- 0.05
t.alpha <- round(qt(alpha/2,df = n-1,lower.tail = F)
    ,3)
cat(tstat,"<",t.alpha)

res <- 2 * pt(-abs(tstat),df= n-1)
cat("The P-Value:",round(res,2))
```

R code Exa 13.5 Testing the beliefs on the methods which is superior than other in treating patients with broken arm

```
#Page Number: 218, 225(contd.).

n1 <- 145
n2 <- 111
x1 <- 21
x2 <- 9
```

```
pt.est1 <- round(x1/n1,4)
pt.est2 <- round(x2/n2,4)
pcap <- round((x1+x2) / (n1+n2),4)

tstat <- round((pt.est1 - pt.est2) / sqrt( pcap*(1-pcap)*((1/n1)+(1/n2)) ),4)
cat("Test Statistic value:",tstat)

zt <- round(qnorm(0.05,lower.tail = F),3)
cat(tstat,"<",zt)

res <-  pnorm(-tstat)
cat("The P-Value:",round(res,4))
```

R code Exa 13.6 Testing on data on ball bearing from 2 machines indicating that the machines are identical

```
#Page Number: 220, 225(contd).

n1 <- 10
n2 <- 10
xbar1 <- 1.96
xbar2 <- 2.01
s1 <- 0.031
s2 <- 0.050

k1 <- n1-1
k2 <- n2-1
tstat_F<- (s1)^2 / (s2)^2
cat("Test Statistic value:",tstat_F)

alpha <- 0.05
f.test_r <- round( qf(alpha/2,df1 = k1,df2 = k2,lower.tail = F),2)
f.test_l <- round(qf(alpha/2,df1 = k1,df2 = k2),3)
```

```
cat(f.test_l,"<",tstat_F,"<",f.test_r)

right_side <- round(1/tstat_F,3)
res = 2 * round(pf(1/right_side,df1 = k1,df2 = k2,
    lower.tail = T),1)

cat("The P-Value:",res)
```

Chapter 14

Bivariate Quantitative Data Correlation and Regression

R code Exa 14.1 Computing the operations of correlation and regression on population consisting of all married couples

```
#Page Number: 241, 246(contd.), 251(contd.), 255(
   contd.), 257(contd.).

data <- data.frame("Couple No." = c(1:12), "Husband
   Age" = c(30,29,36,72,37,36,51,48,37,50,51,36),"
   Wife Age" = c
   (27,20,34,67,35,37,50,46,36,42,46,35))
n <- length(data$Couple.No.)

hus_interval <- seq( from = 28.5, to = 72.5, by =
   11)
hus_class <- cut(data$Husband.Age, hus_interval)
hus_frequency <- cbind(table(hus_class))
hus.Rel_Freq <- (hus_frequency/n)*100
df1 <- data.frame("Group" = unique(hus_class), "
   Frequency" = hus_frequency, "Relative.Frequency"
   = round(hus.Rel_Freq,2))
print(df1)
```

```
wif_interval <- seq( from = 19.5, to = 67.5, by = 12)
wif_class <- cut(data$Wife.Age, wif_interval)
wif_frequency <- cbind(table(wif_class))
wif.Rel_Freq <- (wif_frequency/n)*100
df2 <- data.frame("Group" = unique(wif_class), "Frequency" = wif_frequency, "Relative.Frequency" = round(wif.Rel_Freq,2))
print(df2)



xbar <- sum(data$Husband.Age) / n
ybar <- round(sum(data$Wife.Age) / n,2)
S_x <- round(sd(data$Husband.Age),4)
S_y <- round(sd(data$Wife.Age),4)
sample.average <- data.frame( "Couple" = c("Husband","Wife"), "Mean" = c(xbar,ybar), "SD" = c(S_x,S_y))
print(sample.average)

plot(data$Husband.Age,data$Wife.Age)
cor_coef <- round(cor(data$Husband.Age,data$Wife.Age),4)
cat("The Correlation Coefficient is",cor_coef)

tstat.r <- round((cor_coef*(sqrt(n-2))) / sqrt(1-(cor_coef)^2),4)
cat("Test Statistic value:",tstat.r)

alpha <- 0.05
critical.value <-round(qt(alpha,df = n-2,lower.tail = F),3)
cat(tstat.r,">",critical.value)

Z <- round(qnorm(alpha/2,lower.tail = F),2)
C <- round(exp((2/sqrt(n-3))*(Z)),3)
```

```
D <- round((1+cor_coef)/(1-cor_coef),4)
A <- C*D
B <- D/C
lower <- round((B-1)/(B+1),4)
upper <- round((A-1)/(A+1),4)
cat("95% confidence interval is",lower,",",upper)

reg.fun <- function(data1,data2){
  beta1 <- round(cor(data1,data2)*(sd(data2)/sd(data1)),4)
  beta0 <- round(mean(data2)-(beta1*mean(data1)),3)
  Y.fit <- round(beta0 + beta1*(data1),3)
  return(data2 - Y.fit)
}
sumation.of.Y_fit <- reg.fun(data$Husband.Age,data$Wife.Age)
Std_error <- round(sqrt(sum((sumation.of.Y_fit)^2) / (n-2)),4)
print(paste("Standard Error:",Std_error))

x <- 62
beta1 <- round(cor_coef*(S_y/S_x),4)
beta0 <- round(ybar-(beta1*xbar),3)
yfit <- beta0 + beta1*(62)
cat("Husband's Age is 62 years,prediction for wife's age:",yfit)

E <- round(critical.value * Std_error * sqrt(1 + (1/n)+((x-xbar)^2)/((n-1) * (S_x)^2)),2)
cat("Error Margin:",E)
cat("90% prediction:",yfit-E,",",yfit+E )
```

R code Exa 14.2 Computing the operations of correlation and regression on population of sales of hot chocolate and temperature for a snack shop

```r
#Page Number: 244, 251(contd.), 256(contd.), 257(contd.).

data <- data.frame("Match No." = c(1:9),"Daytime Temp." = c(51,60,65,71,39,32,81,76,66),"Hot Chocolate Sale"= c(187,210,137,136,241,262,110,143,152))
n <- length(data$Match.No.)
plot(x = data$Daytime.Temp., y = data$Hot.Chocolate.Sale, xlim = c(0,90), ylim = c(0,270))

cor_coef <- round(cor(data$Daytime.Temp.,data$Hot.Chocolate.Sale),4)
cat("The Correlation Coefficient is",cor_coef)

tstat.r <- round((cor_coef*(sqrt(n-2))) / sqrt(1-(cor_coef)^2),4)
cat("Test Statistic value:",tstat.r)
alpha <- 0.05
critical.value <-round(qt(alpha,df = n-2,),3)
cat(tstat.r,"<",critical.value)

xbar <- mean(data$Daytime.Temp.)
ybar <- mean(data$Hot.Chocolate.Sale)
s_x <- sd(data$Daytime.Temp.)
s_y <- sd(data$Hot.Chocolate.Sale)
beta1 <- round((cor_coef * (s_y/s_x)),4)
beta0 <- round(ybar-(beta1*xbar),4)
cat("Regression line:",beta0,"-(",-beta1,")X")
```

Chapter 15

Bivariate Catagorical Data Contingency Table

R code Exa 15.1 Testing independence of gender wise usage of 2 types of painkillers

```
#Page Number: 274

data <- matrix(c(115,85,135,165),nrow = 2,ncol = 2, byrow = T)
data <- rbind(data, data[1,]+data[2,])
data <- cbind(data, data[,1]+data[,2])
rownames(data) <- c("Male", "Female","Column Total")
colnames(data) <- c("Med-1","Med-2","Row Total")
n <- data[3,3]
data

E11 <- data[1,3]*data[3,1] / n
E12 <- data[1,3] - E11
E21 <- data[3,2] - E11
E22 <- data[2,3] - E21

tstat.c <- ((data[1,1] - E11)^2 / E11) + ((data[1,2] - E12)^2 / E12) + ((data[2,1] - E21)^2 / E21) +
```

```
    ((data[2,2] - E22)^2 / E22)
cat("test statistic:",tstat.c)

alpha <- 0.05
a <- 2
b <- 2
k <- (a-1)*(b-1)

critical.val <- round(qchisq(alpha,k,lower.tail = F)
    ,3)
cat(tstat.c,">",critical.val)
```

R code Exa 15.2 Testing independence on whether smoking causes hypertension on dataset of randomly selected set of patients

```
#Page Number: 278

data <- matrix(c(20,38,28,50,27,18),nrow = 2,ncol =
    3,byrow = T)
data <- rbind(data, data[1,]+data[2,])
data <- cbind(data, data[,1]+data[,2]+data[,3])
rownames(data) <- c("Hypertension", "No hypertension
    ","Column Total")
colnames(data) <- c("Non Smoker","Moderate","Heavy",
    "Row Total")
n <- data[3,4]
print(data)

E11 <- round(data[1,4]*data[3,1] / n,2)
E12 <- round(data[1,4]*data[3,2] / n,2)
E13 <- round(data[1,4]*data[3,3] / n,2)
E21 <- round(data[2,4]*data[3,1] / n,2)
E22 <- round(data[2,4]*data[3,2] / n,2)
E23 <- round(data[2,4]*data[3,3] / n,2)

```

```
tstat.c <- round(((data[1,1] - E11)^2 / E11) + ((data[1,2] - E12)^2 / E12)+((data[1,3] - E13)^2 / E13) + ((data[2,1] - E21)^2 / E21) + ((data[2,2] - E22)^2 / E22) +((data[2,3] - E23)^2 / E23),4)
cat("test statistic:",tstat.c)

alpha <- 0.05
a <- 2
b <- 3
k <- (a-1)*(b-1)

critical.val <- round(qchisq(alpha,k,lower.tail = F),3)
print(critical.val)
```

R code Exa 15.3 Testing independence of accidents involved in accordence with BAC levels and the region

```
#Page Number: 280

data <- matrix(c(5281,525,3064,3693,336,1949,5337,519,3213,10967,979,6722),nrow = 4,ncol = 3,byrow = T)
data <- rbind(data, data[1,]+data[2,]+data[3,]+data[4,])
data <- cbind(data, data[,1]+data[,2]+data[,3])
rownames(data) <- c("West","North-East","Mid-West","South","Column Total")
colnames(data) <- c("No Alcohol","Low Alcohol","High Alcohol","Row Total")
n <- data[5,4]
data

E11 <- round(data[1,4]*data[5,1] / n,2)
E12 <- round(data[1,4]*data[5,2] / n,2)
```

```
E13 <- round(data[1,4]*data[5,3] / n,2)

E21 <- round(data[2,4]*data[5,1] / n,2)
E22 <- round(data[2,4]*data[5,2] / n,2)
E23 <- round(data[2,4]*data[5,3] / n,2)

E31 <- round(data[3,4]*data[5,1] / n,2)
E32 <- round(data[3,4]*data[5,2] / n,2)
E33 <- round(data[3,4]*data[5,3] / n,2)

E41 <- round(data[4,4]*data[5,1] / n,2)
E42 <- round(data[4,4]*data[5,2] / n,2)
E43 <- round(data[4,4]*data[5,3] / n,2)

tstat.c <- round(((data[1,1] - E11)^2 / E11) + ((data[1,2] - E12)^2 / E12)+((data[1,3] - E13)^2 / E13) + ((data[2,1] - E21)^2 / E21) + ((data[2,2] - E22)^2 / E22) +((data[2,3] - E23)^2 / E23) +((data[3,1] - E31)^2 / E31) + ((data[3,2] - E32)^2 / E32)+((data[3,3] - E33)^2 / E33)+((data[4,1] - E41)^2 / E41) + ((data[4,2] - E42)^2 / E42)+((data[4,3] - E43)^2 / E43),4)
cat("test statistic:",tstat.c)

alpha <- 0.05
a <- 4
b <- 3
k <- (a-1)*(b-1)
critical.val <- round(qchisq(alpha,k,lower.tail = F),3)
cat(tstat.c,">",critical.val)
```

Chapter 16

Multinomial Experiments Goodness of Fit Test

R code Exa 16.1 Correlation coefficient describing the probability distribution of the outcome when 15 marbles are drawn at random from a box of 100 where 60 is red and 40 is blue

```
#Page Number: 289

library(ggplot2)

x <- 0:15
y <- 15 - x
xpy <- data.frame(X = x, Y = y)
coords = paste(x,y,sep=",")

ggplot(xpy, aes(x, y)) +
  geom_point() +
  geom_line() +
  geom_label(aes(x+.5,y+0.5,label=coords))
```

R code Exa 16.3 Finding the probability of rolling a die and getting multiple faces

```
#Page Number: 292

trials <- 15
times <- 0:5
probab <- c(rep(1/6,6))
options(scipen = 999)
val <- round(dmultinom(times,size = trials,prob =
    probab),6)
cat("Probability:",val)
```

R code Exa 16.4 Finding the probabilty of getting five 6s by rolling a die 15 times

```
#Page Number: 292

trials <- 15
times <- c(5,10)
probab <- c(rep(1/6,2))
res <- sum(round(dbinom(times,size = trials,prob =
    probab),4))
cat("Probability:",res)
```

R code Exa 16.5 Testing the hypothesis of having 60 percent red marbles from a box having unknown composition of red and blue marbles

```
#Page Number: 293

n <- 15
X <- 11
prob <- 0.60
```

```
pcap <- round(X/n,4)
tstat <- round(sqrt(n)*(pcap-prob) / sqrt(prob*(1-
   prob)),4)
cat("Test Statistic:",tstat)

alpha <- 0.05
zt <- round(qnorm(alpha/2,lower.tail = F),2)
cat(tstat,"<",zt)
```

R code Exa 16.6 Determining the uniformity in composition of types of marble in a box

```
#Page Number: 294

n <- 60
x1 <- 22
x2 <- 17
x3 <- 21
k <- 3
p <- 1/3
np10  = np20 = np30 <- n*p

tstat <- round(((x1 - np10)^2 / np10) +((x2 - np20)
   ^2 / np20)+((x3 - np30)^2 / np30),3)
cat("Test Statistic:",tstat)

alpha <- 0.05
chi <- round(qchisq(alpha,df = k-1,lower.tail = F )
   ,2)
cat(tstat,"<",chi)
```

R code Exa 16.7 Testing the hypothesis of ratio of the offsprings in a biology experiment

```r
#Page Number: 296

n <- 432
k <- 4
O1 <- 253
O2 <- 69
O3 <- 87
O4 <- 23

alpha <- 0.05

p1 <- 9/16
p2=p3 <- 3/16
p4 <- 1/16

E1 <- n*p1
E2 <- n*p2
E3 <- n*p3
E4 <- n*p4

tstat <- round(((O1-E1)^2 / E1)+((O2-E2)^2 / E2)+((O3-E3)^2 / E3)+((O4-E4)^2 / E4),4)
cat("Test Statistic:",tstat)

chi <- round(qchisq(alpha,k-1,lower.tail = F),2)
cat(tstat,"<",chi)
```

R code Exa 16.8 Goodness of fit test of survey data on a city daytime traffic

```r
#Page Number: 297

data <- data.frame("Traffic Catagory" = c("PPV","LCV","HCV"), "Sur 1" = c(436,391,297),"Sur 2" = c(520,401,319),"Sur 3" = c(376,281,191))
```

```
alpha <- 0.05
k <- 3

n1 <- sum(data$Sur.1)
n2 <- sum(data$Sur.2)
n3 <- sum(data$Sur.3)

p1 <- 0.40
p2 <- 0.35
p3 <- 0.25

O1 <- data$Sur.1
O2 <- data$Sur.2
O3 <- data$Sur.3


E1 <- n1*c(p1,p2,p3)
E2 <- n2*c(p1,p2,p3)
E3 <- n3*c(p1,p2,p3)

tstat1 <- sum(((O1 - E1)^2) /E1)
tstat2 <- sum(((O2 - E2)^2) /E2)
tstat3 <- sum(((O3 - E3)^2) /E3)
tot.tstat <- round(tstat1+tstat2+tstat3,4)
cat("Test Statistic:",tot.tstat)

chi <- round(qchisq(alpha,3*(k-1),lower.tail = F),3)
cat(tot.tstat,"<",chi)
```

Chapter 17

Hypothesis Testing to Compare Multiple Populations

R code Exa 17.1 Computing the population mean variance and applying 1 way anova test on experimental data of number of seconds taken to help in presence of other persons

```
#Page Number: 308

ze <- c(25,30,20,32,NA)
ze1 <- na.exclude(ze)
tw <- c(30,33,29,40,36)
fo <- c(32,39,35,41,44)
k <- 3
n1 <- length(ze1)
n2 <- length(tw)
n3 <- length(fo)

nT <-n1+n2+n3
s1 <- round(sd(ze,na.rm = T),3)
s2 <- round(sd(tw),3)
s3 <- round(sd(fo),3)

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

```
n3-1)*(s3)^2)) / (nT-k)),3)
cat("Pooled Sample SD:",sp)

c <- round(1+(1/(3*(k-1))) * ((1/(n1-1)+1/(n2-1)+1/(n3-1)) - (1/(nT-k))),4)
cat("Constant:",c)

tstat <- round((2/c) *(((nT-k)*log(sp))-(((n1-1)*log(s1)) +((n2-1)*log(s2))+((n3-1)*log(s3)))),4)
cat("Test Statistics:",tstat)

alpha <- 0.05
chi <- round(qchisq(alpha,(k),lower.tail = F),2)
cat(tstat,"<",chi)


xbar1 <- mean(ze1)
xbar2 <- mean(tw)
xbar3 <- mean(fo)
xbar <- round((n1*xbar1+n2*xbar2+n3*xbar3)/nT,4)
SSP <- round(n1*(xbar1-xbar)^2+n2*(xbar2-xbar)^2+n3*(xbar3-xbar)^2,2)
SSE <-  round((nT - k)*(sp)^2,2)
TSS <- SSP + SSE
MSP <- SSP / (k-1)
MSE <- round(SSE / (nT-k),2)
Anova1 <- round(MSP / MSE,2)

tab <- data.frame("TSS"=TSS,"SSP"=SSP,"SSE"=SSE,"MSP"=MSP,"MSE"=MSE,"TStat"=Anova1)
print(tab)

ft <- qf(alpha,k-1,nT-k,lower.tail = F)
cat(Anova1,">",ft)
```

R code Exa 17.2 Classification of two way ANOVA and Barletts test on data studying weight gains by rats influenced by the source of protein

```r
#Page Number: 310

data <- data.frame("Diet-1"= c(90,94,86,63,81,89,63,82,63,81),"Diet-2"= c(87,70,95,71,75,84,62,72,81,93),"Diet-3"= c(83,82,85,63,72,85,64,80,82,83),"Diet-4"= c(107,102,102,93,111,128,56,97,80,103),"Diet-5"= c(96,72,76,70,79,89,70,91,63,102),"Diet-6"= c(111,100,102,93,101,104,72,92,87,112))

concat = c(t(as.matrix(data)))
f = c("Diet-1","Diet-2","Diet-3","Diet-4","Diet-5","Diet-6")
k <- 6
n <- 10
data

pop <-gl(k, 1, n*k, factor(f))
block = gl(n, k, k*n)
anov2 = aov(concat ~ pop + block)
summary(anov2)


nT <- 60
s1 <- round(sd(data$Diet.1),4)
s2 <- round(sd(data$Diet.2),4)
s3 <- round(sd(data$Diet.3),4)
s4 <- round(sd(data$Diet.4),4)
s5 <- round(sd(data$Diet.5),4)
s6<-  round(sd(data$Diet.6),4)

sp <- round(sqrt((((n-1)*(s1)^2)+((n-1)*(s2)^2)+((n-1)*(s3)^2)+((n-1)*(s4)^2)+((n-1)*(s5)^2)+((n-1)*(s6)^2)) / (nT-k)),4)
cat("Pooled Sample SD:",sp)
```

```
c <- round(1 + (1/(3*(k-1))) * (((1/(n-1))*6) - (1/(nT-k))),4)
cat("Constant:",c)

tstat <- round((2/c) *(((nT-k)*log(sp)) - (((n-1)*log(s1)+(n-1)*log(s2)+(n-1)*log(s3)+(n-1)*log(s4)+(n-1)*log(s5)+(n-1)*log(s6)))),4)
cat("Test Statistics:",tstat)

alpha <- 0.05
chi <- qchisq(alpha,(k),lower.tail = F)
cat(tstat,"<",round(chi,2))

#The answer may vary due to difference in representation.
```

Chapter 18

Quality Management Using Statistics

R code Exa 18.1 To test whether the manufacturing process is in control by construction of x chart and s chart from the dataset of JetMate

```
#Page Number: 336

library(qcc)

data <- read.csv("18_1.csv")
chart <- qcc.groups(data = data$JET,sample = data$GROUP)


plot.chart.R <- qcc(chart,type = "xbar",std.dev = "UWAVE-R")
summary(plot.chart.R)
plot.chart.SD <- qcc(chart,type = "xbar",std.dev = "UWAVE-SD")
summary(plot.chart.SD)


plot.chart_R <- qcc(chart,type = "R")
```

```
summary(plot.chart_R)

plot.chart_S <- qcc(chart,type = "S")
summary(plot.chart_S)
# The answer may slightly vary due to rounding off values.
# The answer may vary due to difference in representation.
```

R code Exa 18.2 To test whether the bottling process is in control by construction of x chart and r chart and s chart from the dataset of BottlePack company

```
#Page Number: 338

library(qcc)

data <- data.frame("Sam"=c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5),rep(6,5),rep(7,5)),"Obs"=c(1.014,1.013,1.001,1.017,1.019,1.007,0.992,0.986,1.004,0.985,1.00
)
data
chart <- qcc.groups(data = data$Obs,sample = data$Sam)


plot.chart_R <- qcc(chart,type = "R")
summary(plot.chart_R)

plot.chart.SD <- qcc(chart,type = "xbar",std.dev = "UWAVE-SD")
summary(plot.chart.SD)


```

```
plot.chart_S <- qcc(chart,type = "S")
summary(plot.chart_S)

#The answer may vary due to difference in
   representation.
```

R code Exa 18.3 To test whether the manufacturing process is in control by construction of s chart from the dataset of Blackstone company

```
#Page Number: 342

library(qcc)

data <- data.frame("Sam"= c(rep(1,3),rep(2,3),rep
   (3,3),rep(4,3),rep(5,3),rep(6,3),rep(7,3),rep
   (8,3),rep(9,3),rep(10,3)),"Obs"= c
   (40,33,30,34,21,28,27,32,36,18,28,24,40,40,38,30,20,18,30,28,31,3
   )
chart <- qcc.groups(data = data$Obs,sample = data$
   Sam)

plot.chart_S <- qcc(chart,type = "S")
summary(plot.chart_S)
```

R code Exa 18.4 To test whether the manufacturing process is in control by construction of p chart from the dataset of ChipItGood company

```
#Page Number: 344

library(qcc)

data <- c
   (6,10,8,5,7,4,12,15,8,15,20,4,10,14,7,5,8,6)
```

```
n <- 100
plot.p <- qcc(data, type ="p",sizes = n)
summary(plot.p)

#The answer may vary due to difference in
   representation.
```