R Textbook Companion for Statistics for Business and Economics by Anderson, Sweeney, and Williams[1]

Created by
Shivam Agrawal
B.Tech.
Computer Science and Engineering
Techno India NJR Institute of Technology
Cross-Checked by
R TBC Team

May 29, 2020

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

Book Description

Title: Statistics for Business and Economics

Author: Anderson, Sweeney, and Williams

Publisher: Cengage Learning, USA

Edition: 11

Year: 2011

ISBN: 9780538471886

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

Descriptive Statistics Tabular and Graphical Presentations

R code Exa 2.1a Summarizing Categorical Data Part 1

```
                                                    # Page no. :
                                                        33

# Frequency Distribution (Categorical Data)

# Dataset

soft_drink_names <- c("Coke Classic", "Diet Coke", "
    Pepsi", "Diet Coke", "Coke Classic","Coke Classic
    ",
                            "Dr. Pepper", "Diet Coke", "
                                Pepsi", "Pepsi" , "Coke
                                Classic", "Dr. Pepper",
                            "Sprite", "Coke Classic", "Diet
                                Coke", "Coke Classic", "
                                Coke Classic", "Sprite",
                            "Coke Classic", "Diet Coke", "
                                Coke Classic", "Diet Coke",
                                "Coke Classic", "Sprite",
```

```
                        "Pepsi", "Coke Classic", "Coke Classic", "Coke Classic", "Pepsi", "Coke Classic",
                        "Sprite", "Dr. Pepper", "Pepsi", "Diet Coke", "Pepsi", "Coke Classic",
                        "Coke Classic", "Coke Classic", "Pepsi", "Dr. Pepper", "Coke Classic", "Diet Coke",
                        "Pepsi", "Pepsi", "Pepsi", "Pepsi", "Coke Classic", "Dr. Pepper", "Pepsi", "Sprite")

soft_drink_table <- data.frame(table(soft_drink_names))

                                        # Page no. : 34

FD <- data.frame(Soft_drinks = soft_drink_table$soft_drink_names,
                 Frequency = soft_drink_table$Freq)    # Frequency Distribution


RF <-  FD$Frequency / sum(FD$Frequency)   # Relative Frequency

FD <- cbind(FD,Relative_frequency = RF)

PF <- FD$Relative_frequency * 100    # Percentage Frequency

FD <- cbind(FD, Percentage_frequency = PF)

View(FD)   # Viewing the Frequency Distribution Table

# Total values of Frequency Distribution
```

```

total_freq <- sum(FD$Frequency)

total_rel_freq <- sum(FD$Relative_frequency)

total_per_freq <- sum(FD$Percentage_frequency)

cat("Total value for frequency is", total_freq, "\n")
cat("Total value for relative frequency is", total_rel_freq, "\n")
cat("Total value for percentage frequency is", total_per_freq)
```

R code Exa 2.1b Summarizing Categorical Data Part 2

```
                                        # Page no. : 35

# Bar Charts and Pie Charts

# Dataset

soft_drink_names <- c("Coke Classic", "Diet Coke", "Pepsi", "Diet Coke", "Coke Classic","Coke Classic",
                      "Dr. Pepper", "Diet Coke", "Pepsi", "Pepsi" , "Coke Classic", "Dr. Pepper",
                      "Sprite", "Coke Classic", "Diet Coke", "Coke Classic", "Coke Classic", "Sprite",
                      "Coke Classic", "Diet Coke", "Coke Classic", "Diet Coke", "Coke Classic", "Sprite",
```

```
                              "Pepsi", "Coke Classic", "Coke Classic", "Coke Classic", "Pepsi", "Coke Classic",
                              "Sprite", "Dr. Pepper", "Pepsi", "Diet Coke", "Pepsi", "Coke Classic",
                              "Coke Classic", "Coke Classic", "Pepsi", "Dr. Pepper", "Coke Classic", "Diet Coke",
                              "Pepsi", "Pepsi", "Pepsi", "Pepsi", "Coke Classic", "Dr. Pepper", "Pepsi", "Sprite")

soft_drink_table <- data.frame(table(soft_drink_names))

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

# Bar Chart

ggplot(soft_drink_table,aes(soft_drink_names,Freq, fill = soft_drink_names))+geom_bar(stat = "identity")+
labs(title="Bar chart" , x = "Soft Drink", y = "Frequency")+ylim(0,20)

# Pie Chart

soft_drink_purchase_slices <- soft_drink_table$Freq
soft_drink_names_labels <- soft_drink_table$soft_drink_names
```

```
soft_drink_purchase_pct <- (soft_drink_table$Freq/
    sum(soft_drink_table$Freq)) * 100

soft_drink_names_labels <- paste(soft_drink_names_
    labels,soft_drink_purchase_pct)
soft_drink_names_labels <- paste(soft_drink_names_
    labels, "%", sep = "")

pie(soft_drink_purchase_slices, labels = soft_drink_
    names_labels,
     col = rainbow(length(soft_drink_names_labels)),
     main = "Pie Chart for Soft Drink Purchase")
```

R code Exa 2.2a Summarizing Quantitative Data

```
                                        # Page no. :
                                          39-40

# Frequency Distribution (Quantitative Data)

audit_data <- c
    (12,15,20,22,14,14,15,27,21,18,19,18,22,33,16,18,17,23,28,13)


no_of_classes <- 5

width <- (max(audit_data) - min(audit_data)) / no_of
    _classes

width <- ceiling(width)   # Rounding up of the value

breaks <- seq(10,34,by = width)

class_range <- cut(audit_data, breaks, right=T)

```

```
frequency <- table(class_range)

frequency_distribution <- data.frame(frequency)

frequency_distribution <- data.frame(class_range = frequency_distribution$class_range,
                                     frequency = frequency_distribution$Freq)

# Note that :- Book answer will differ with my answer though number of classes and width of
# each class is same as in the book!!!

                                                        # Page no. : 41

relative_frequency <- round(frequency_distribution$frequency / sum(frequency_distribution$frequency)
                            ,2)   # Rounding of data to 2 digits

percentage_frequency <- relative_frequency * 100

audit_data_FD <- cbind(frequency_distribution,
                       relative_frequency, percentage_frequency)

View(audit_data_FD)

                                                        # Page no. : 41-42

# Dot Plot and Histogram

```

```
# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

# Dot Plot

dotchart(audit_data, main = "Dot Plot for the Audit
   Time Data", xlab = "Audit Time (days)",
         cex = 0.5)   # cex is for scaling

# Note that: Book dot plot is different from my dot
   plot.

# Histogram

ggplot(audit_data_FD, aes(class_range, frequency,
   fill = class_range))+
  geom_histogram(stat = "identity")+labs(title="
     Histogram for the Audit Time Data" ,
                                          x = "Audit
                                             Time (
                                             days)",
                                             y = "
                                             Frequency
                                             ")+ylim
                                             (0,8)
```

R code Exa 2.3a Scatter Plot and Tradeline

```
                                    # Page no. :
                                       57-58
```

```
# Dataset

week <- c(1,2,3,4,5,6,7,8,9,10)
x <- c(2,5,1,3,4,1,5,3,4,2)
y <- c(50,57,41,54,54,38,63,48,59,46)
data <- data.frame(week,x,y)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

# Scatter Plot

ggplot(data, aes(x,y)) + geom_point() + geom_smooth(
    method = "lm", se = F) +
labs(title = "Scatter Plot and Tradeline for the
    Stereo and Sound Equipment Store",
    x = "Number of Commercials",
    y =
```

```
"
Sales

(
$
100
s
)
"
)
```

R code Exa 2.4a Cumulative Distributions

```
                                        # Page no. : 44

# Cumulative Distributions

audit_time <- c("Less than or equal to 14", "Less than or equal to 19", "Less than or equal to 24",
                "Less than or equal to 29", "Less than or equal to 34")
frequency <- c(4,8,5,2,1)   # Refer to table no. 2.5 page no. 40
cumulative_freq <- cumsum(frequency)

DF <- data.frame(audit_time, frequency, cumulative_freq)

CRF <-  DF$cumulative_freq / sum(DF$frequency)   # Cumulative Relative Frequency

CPF <- CRF * 100   # Cumulative Percentage Frequency
```

```r

DF <- cbind(DF, CRF, CPF)
View(DF)

breaks <- seq(9, 34, by = 5)
cumfreq0 <- c(0, DF$cumulative_freq)

DF2 <- data.frame(breaks, cumfreq0)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF2, aes(breaks, cumfreq0, group = 1)) + geom
    _point() + geom_line() +
  xlim(c(0,35)) + labs(title = "Ogive For the Audit
      Time Data", x ="Audit Time (Days)",
      y ="Cumulative Frequency")
```

R code Exa 2.5a Exploratory Data Analysis The Stem and Leaf Display

```r
                                        # Page no. :
                                           49

# Exploratory Data Analysis : The Stem - and - Leaf
    Display

data <- c(112, 72, 69, 97, 107, 73, 92, 76, 86, 73,
    126, 128, 118, 127, 124, 82, 104, 132, 134, 83,
           92, 108, 96, 100, 92, 115, 76, 91, 102,
              81, 95, 141, 81, 80, 106, 84, 119, 113,
```

```
              98, 75,
            68, 98, 115, 106, 95, 100, 85, 94, 106,
               119)

stem(data)

                                                      # Page no. :
                                                         51

data2 <- c(1565, 1852, 1644, 1766, 1888, 1912, 2044,
    1812, 1790, 1679, 2008, 1852, 1967, 1954,
           1733)
stem(data2)   # Answer is varing from the book
```

Chapter 3

Descriptive Statistics Numerical Measures

R code Exa 3.1a Measures of Location Mean Part 1

```
                                        # Page no. :
                                          87

# Mean

x <- c(46, 54, 42, 46, 32)

sample_mean <- mean(x)

cat("Sample mean for x is ",sample_mean)
```

R code Exa 3.1b Measures of Location Mean Part 2

```
                                        # Page no. :
                                          88

```

```
# Dataset

graduate <- c(1,2,3,4,5,6,7,8,9,10,11,12)
salary <- c
    (3450,3550,3650,3480,3355,3310,3490,3730,3540,3925,3520,3480)


dataset <- data.frame(graduate, salary)

# Mean

mean_salary <- mean(dataset$salary)

cat("Mean monthly starting salary of 12 business
    school graduates is",mean_salary)
```

R code Exa 3.1c Measures of Location Median Part 1

```
                                                # Page no.
                                                    : 89

# Median

x <- c(32, 42, 46, 46, 54)

median <- median(x)

cat("Median value for x is ", median)
```

R code Exa 3.1d Measures of Location Median Part 2

```r
# Page no. : 89

salary <- c(3450,3550,3650,3480,3355,3310,3490,3730,3540,3925,3520,3480)

# Median

median <- median(salary)

cat("The median of salary is ", median)
```

R code Exa 3.1e Measures of Location Percentiles and Quartiles

```r
# Page no. : 90-91

salary <- c(3450,3550,3650,3480,3355,3310,3490,3730,3540,3925,3520,3480)

# 85th and 50th Percentiles

solution <- quantile(salary, probs = c(0.85, 0.5))

cat("Value for 85th and 50th percentile are ", solution[1], ",", solution[2])

# Note that: 85th percentile value is different from the book


# 25th, 50th, 75th Percentiles (First, Second, Third Quartiles)
```

```
values <- quantile(salary, probs = c(0.25, 0.5,
   0.75))

cat("Value for first, second, third quartiles are ",
    values[1], ",", values[2], ",", values[3] )

# Note that: First and Second Quartile values are
   different from the book
```

R code Exa 3.2a Measures of Variability Range and IQR

```
                                              # Page no. :
                                                 96-97

salary <- c
   (3450,3550,3650,3480,3355,3310,3490,3730,3540,3925,3520,3480)

# Range

range <- range(salary)
diff <- range[2] - range[1]

cat("Range is",diff)

# Inter-Quartile Range

IQR <- IQR(salary)

cat("IQR is ",IQR)

# Note that : IQR value of Book is different.
```

R code Exa 3.2b Measures of Variability Variance

```
                                        # Page no. :
                                          97-98

students <- c(46, 54, 42, 46, 32)

# Variance

variance <- var(students)

cat("Variance of students is ",variance)
```

R code Exa 3.2c Measures of Variability Standard Deviation

```
                                        # Page no. :
                                          98-99

salary <- c
   (3450,3550,3650,3480,3355,3310,3490,3730,3540,3925,3520,3480)

# Variance

variance <- var(salary)

# Standard Deviation

sd <- sd(salary)

cat("Variance of salary is ",variance)
cat("Standard Deviation is ",sd)
```

R code Exa 3.3a Z Score

```
                                                              # Page
                                                                 no. :
                                                                  104

# Z−score

students <- c(46,54,42,46,32)

deviation <- students - mean(students)

sample_variance <- var(students)

dataset <- data.frame(students,deviation)

z <- c()

for(i in 1:length(dataset$students)){
  z[i] <- deviation[i]/sqrt(sample_variance)
}

dataset <- cbind(dataset, zScore = z)

View(dataset)
```

R code Exa 3.4a Boxplot

```
                                                            # Page no.
                                                               : 110

```

```
salary <- c
    (3450,3550,3650,3480,3355,3310,3490,3730,3540,3925,3520,3480)

# Install Library if not installed

#install.packages("ggplot2")

# Import Library

library(ggplot2)

# Boxplot

ggplot(data.frame(salary), aes(x = "", y = salary))
    + geom_boxplot(fill = "purple")+
  ggtitle("Boxplot for Salary") + ylab("Salary")
```

R code Exa 3.5a Covariance and Correlation Coefficient

```
                                        # Page no. :
                                           115-116 and 119

# Dataset

x <- c(2,5,1,3,4,1,5,3,4,2)
y <- c(50,57,41,54,54,38,63,48,59,46)

dataset <- data.frame(x,y)

# Coveriance

coveriance <- cov(dataset$x,dataset$y)

cat("Value of covariance is ",coveriance)
```

```r

# Correlation Coefficient

correlation <- cor(dataset$x,dataset$y)

cat("Value of correlation coefficient is ",
    correlation)
```

R code Exa 3.5b Sample Correlation Coefficient

```r
                                          # Page no.
                                            : 121

# Data

x <- c(5,10,15)
y <- c(10,30,50)

# Sample Correlation Coefficient

corr <- cor(x, y)

cat("Sample correlation coefficient of x and y is ",
     corr)
```

R code Exa 3.6a Weighted Mean

```r
                                       # Page no. :
                                          124 - 125

# Data

purchase <- c(1,2,3,4,5)
```

```
cost <- c(3.00,3.40,2.80,2.90,3.25)
pound <- c(1200,500,2750,1000,800)

dataset <- data.frame(purchase,cost,pound)

# Weighted Mean

mean <- weighted.mean(dataset$cost,dataset$pound)

cat("Weigted mean for the dataset is",mean)
```

R code Exa 3.6b Grouped Data Mean and Sample Variance

```
# Page no. : 126 - 127

# Data

audit <- c("10-14", "15-19","20-24","25-29","30-34")
midpoint <- c(12,17,22,27,32)
frequency <- c(4,8,5,2,1)

dataset <- data.frame(audit, midpoint, frequency)

# Mean for Grouped Data

mean <- weighted.mean(dataset$midpoint,dataset$frequency)

cat("Mean for grouped data is",mean)

# Sample Variance for Grouped Data

var <- sum(dataset$frequency*((dataset$midpoint -
```

```
    mean)**2)) / (sum(dataset$frequency) - 1)

# Note that : Grouped sample variance has no inbuild
    function

cat("Sample variance for grouped data is",var)
```

Chapter 4

Introduction to Probability

R code Exa 4.1a Combinations

```
                                                    # Page no. :
                                                       154

# Combinations

# Eg. 1

N <- 5
n <- 2

combinations <- choose(n = N, k = n)

cat("The total combinations are",combinations)

# Eg. 2

N <- 53
n <- 6

combinations <-  choose(n = N, k = n)

```

```
cat("The total combinations are",combinations)
```

R code Exa 4.1b Permutations

```
                                                    # Page no.
                                                       : 155

# Permutations

N <- 5
n <- 2

permutations <- choose(n = N, k = n)*factorial(n)

cat("The total permutations are",permutations)
```

R code Exa 4.1c Assigning Probabilities

```
                                               # Page no. : 155
                                                  - 156

# Assigning Probabilities

x <- c(0,1,2,3,4)
y <- c(2,5,6,4,3)

DF <- data.frame(x,y)

y_sum <- sum(DF$y)

prob <- DF$y / y_sum

DF <- cbind(DF,prob)
```

```
View(DF)
```

R code Exa 4.1d Probabilities Assigning Example

```
                                        # Page no. :
                                          157 - 158

# Probabilities Assigning Example

x <- c(2,2,2,3,3,3,4,4,4)
y <- c(6,7,8,6,7,8,6,7,8)

z = list()

for(i in 1:length(x))
{
  z[i] <- list(c(x[i],y[i]))
}

past_project <- c(6,6,2,4,8,2,2,4,6)

DF <- data.frame(x,y,I(z),past_project)

past_project_sum <- sum(DF$past_project)

p <- DF$past_project / past_project_sum

DF <- cbind(DF,p)

total_probability <- sum(DF$p)

cat("Total probability for the Sample Point is",
   total_probability)

```

```
View(DF)
```

R code Exa 4.2a Probability of an Event

```
                                          # Page no. :
                                             161

# Probability of an Event

# C denotes the event that is completed in 10 months
    or less

C <- c(list(c(2,6)), list(c(2,7)), list(c(2,8)),
   list(c(3,7)), list(c(3,8)), list(c(4,6)))
prob <- c(0.15,0.15,0.05,0.10,0.20,0.05)

dataset <- data.frame(I(C),prob)

event <- sum(dataset$prob)

# P(C) = P(2,6) + P(2,7) + P(2,8) + P(3,7) + P(3,8)
   + P(4,6)

cat("Probability of an event P(C) is",event)

# L denotes the event that is completed in less than
    10 months

L <- c(list(c(2,6)), list(c(2,7)), list(c(3,7)))
prob <- c(0.15,0.15,0.10)

dataset <- data.frame(I(L),prob)

# P(L) = P(2,6) + P(2,7) + P(3,7)

```

```
event2 <- sum(dataset$prob)

cat("Probability of an event P(L) is",event2)

# M denotes the event that is completed in more than
    10 months

M <- c(list(c(3,8)), list(c(4,7)), list(c(4,8)))
prob <- c(0.05,0.10,0.15)

dataset <- data.frame(I(M),prob)

# P(M) = P(3,8) + P(4,7) + P(4,8)

event3 <- sum(dataset$prob)

cat("Probability of an event P(L) is",event3)
```

R code Exa 4.3a Probability Computation using Complement

```
                                                # Page no. :
                                                    165

# Probability Computation using Complement

p_comp_A <- 0.80

p_A <- 1 - p_comp_A

cat("Probability for A is",p_A)
```

R code Exa 4.3b Intersection and Union of Events

```r
                                          # Pahe no. : 167

# Intersection and Union of Events

a <- 5
n <- 50
c <- 6
d <- 2

p_L <- a / n

p_D <- c / n

p_L_and_D <- d / n


cat("Probability for L intersection D is",p_L_and_D)

p_L_or_D <- p_L + p_D - p_L_and_D

cat("Probability for L union D is",p_L_or_D)
```

R code Exa 4.3c Addition Law

```r
                                          # Page no. :
                                            168

# Addition Law

p_S <- 0.30
p_W <- 0.20
p_S_and_W <- 0.12

p_S_or_W <- p_S + p_W - p_S_and_W

```

```
cat("Probability after applying addition law is ",p_
    S_or_W)
```

R code Exa 4.4a Conditional Probability

```
                                         # Page no. :
                                            171 - 174

# Conditional Probability

position <- c("Promoted", "Not Promoted")
gender <- c("Men","Women")
number1 <- c(288, 36)   # Promoted
number2 <- c(672, 204)  # Not Promoted

DF <- data.frame(position, gender, number1, number2)

table <- round(prop.table(DF[,3:4]), 2)
table <- as.matrix(table)
table

position <- c("Promoted", "Not Promoted")
men <- c(table[1,1][[1]], table[1,2][[1]])
women <- c(table[2,1][[1]], table[2,2][[1]])

DF2 <- data.frame(position, men, women)
View(DF2)

columnSums <- apply(DF2[,2:3],2, sum)
rowSums <- apply(DF2[,2:3], 1, sum)

cat("Total Probability For having Men is",
    columnSums[1])
cat("Total Probability For having Women is",
    columnSums[2])
```

```
cat("Total Probability For Promotion", rowSums[1])
cat("Total Probability For Not a Promotion", rowSums
   [2])

cond1 <- DF2$men[1]  / columnSums[1]

cat("Conditional Probability for Men and getting
   Promoted given the Probability
     of Total Men is",cond1)

cond2 <- DF2$women[1] / columnSums[2]

cat("Conditional Probability for Women and getting
   Promoted given the Probability
     of Total Women is",cond2)
```

R code Exa 4.4b Multiplication Law

```
                                                  # Page no. :
                                                     174 - 175

# Multiplication Law

# P(S | D) = 0.75

x <- 0.75

# Event S = 0.84

y <- 0.84

x_and_y <- x * y

cat("After applying multiplication law we get", x_
   and_y)
```

```r

# Multiplication Law for Independent Events

A <- 0.80

B <- 0.80

A_and_B <- A * B

cat("After applying multiplication law for
   independent events A and B we get", A_and_B)
```

R code Exa 4.5a Bayes Theorem Tabular Approach

```r
                                    # Page no. : 182

# Tabular Approach for Bayes' Theorem

events <- c("A1","A2")
prior_probabilities <- c(0.65,0.35)

conditional_probabilities <- c(0.02,0.05)

joint_probabilities <- prior_probabilities *
   conditional_probabilities

total_joint_probability <- sum(joint_probabilities)

posterior_probabilities <- joint_probabilities /
   total_joint_probability

DF <- data.frame(events,prior_probabilities,
   conditional_probabilities
                  ,joint_probabilities,posterior_
                     probabilities)
```

```
View(DF)

cat("Total posterior probability is",sum(DF$posterior_probabilities))
```

Chapter 5

Discrete Probability Distribution

R code Exa 5.1a Discrete Probability Distribution Graph Expected value Variance and Standard Deviation

```
                                    # Page no. : 198
                                        - 199

# Discrete Probability Distribution:

x <- c(0,1,2,3,4,5)
prob_of_x <- c(0.18,0.39,0.24,0.14,0.04,0.01)

dataset <- data.frame(x, prob_of_x)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

```

```
ggplot(dataset, aes(x = x, y = prob_of_x)) + geom_
    bar(stat = "identity", fill = "blue") + labs(
  title = "Graphical representation of the
      Probability Distribution",
  x = "Number of Automobiles", y = "Probability")

                                             # Page no. : 203

# Expected value

exp_values <- dataset$x * dataset$prob_of_x

dataset <- data.frame(cbind(dataset, exp_values))

expected_value <- sum(dataset$exp_values)
cat("Expected value for the given problem is",
    expected_value)

                                             # Page no. : 204 -
                                                 205

# Variance and Standard Deviation

deviation_of_x <- (dataset$x - expected_value)

deviation_square <- (deviation_of_x) ** 2
variance <- sum(dataset$prob_of_x * deviation_square
    )
standard_deviation <- sqrt(variance)

cat("Variance is",variance)
cat("Standard Deviation is",standard_deviation)
```

R code Exa 5.2a Binomial Probability Distribution

```r
                                        # Page no. : 211

# Binomial Probability Distribution

no_of_trials <- 3
no_of_successes <- 2
BPD <- choose(n = no_of_trials, k = no_of_successes)

cat("Answer is",BPD)

no_of_successes <- 3
BPD <- choose(n = no_of_trials, k = no_of_successes)
cat("Answer is",BPD)
```

R code Exa 5.2b Binomial Probability Distribution Eg2

```r
                                                # Page no
                                                  . :
                                                  211

# Binomial Probability Distribution Eg-2

# Data
customer1 <- c("Purchase","Purchase","No Purchase")
customer2 <- c("Purchase","No Purchase","Purchase")
customer3 <- c("No Purchase","Purchase","Purchase")

customer <- data.frame(customer1,customer2,customer3
   )

len <- nrow(customer)  # Trial

x <- 1   # For purchase
y <- 0   # For no purchase

```

```
p <- 0.30
q <- 1 - p
outcome <- c()

# Install Library if not installed

install.packages("Rlab")

# Import Library

library(Rlab)    # For dbern

for(i in 1:len)
{
  if(customer1[i] == "Purchase" && customer2[i] == "Purchase" && customer3[i] == "Purchase")
  {
    outcome[i] = dbern(x,p) * dbern(x,p) * dbern(x,p)
  } else if(customer1[i] == "Purchase" && customer2[i] == "Purchase" && customer3[i] == "No Purchase")
  {
    outcome[i] = dbern(x,p) * dbern(x,p) * dbern(y,p)
  } else if(customer1[i] == "Purchase" && customer2[i] == "No Purchase" && customer3[i] == "Purchase")
  {
    outcome[i] = dbern(x,p) * dbern(y,p) * dbern(x,p)
  } else if(customer1[i] == "Purchase" && customer2[i] == "No Purchase" && customer3[i] == "No Purchase")
  {
    outcome[i] = dbern(x,p) * dbern(y,p) * dbern(y,p)
  } else if(customer1[i] == "No Purchase" &&
```

```
    customer2[i] == "Purchase" && customer3[i] == "Purchase")
  {
    outcome[i] = dbern(y,p) * dbern(x,p) * dbern(x,p)
  } else if(customer1[i] == "No Purchase" && customer2[i] == "Purchase" && customer3[i] == "No Purchase")
  {
    outcome[i] = dbern(y,p) * dbern(x,p) * dbern(y,p)
  } else if(customer1[i] == "No Purchase" && customer2[i] == "No Purchase" && customer3[i] == "Purchase")
  {
    outcome[i] = dbern(y,p) * dbern(y,p) * dbern(x,p)
  } else
  {
    outcome[i] = dbern(y,p) * dbern(y,p) * dbern(y,p)
  }
}

customer <- cbind(customer,outcome)

View(customer)

                                            # Page no. : 212 - 213

# Binomial Probability Function

x <- c(0,1,2,3)
fun <- c()

for (i in 0:length(x)) {
```

```
  fun[i] <- dbinom(x[i],len,p)
}

dataset <- data.frame(x, fun)
View(dataset)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(dataset, aes(x = x, y = fun)) + geom_bar(stat = "identity", fill = "blue") + labs(
  title = "Graphical representation of the Probability Distribution",
  x = "Number of Customers", y = "Probability")

                                        # Page no. : 214 - 215

# Expected Value, Variance and Standard Deviation for Binomial Probability Distribution

expected_value <- len * p
variance <- len * p * q
standard_deviation <- sqrt(variance)

cat("Expected value is",expected_value)
cat("Variance is",variance)
cat("Standard deviation is",standard_deviation)
```

R code Exa 5.3a Poisson Probability Distribution

```
                                                # Page no. :
                                                    218

# Poisson Probability Distribution

x = 5
expected_value <- 10
e <- 2.72  # Exponential value

PPD <- dpois(x, expected_value)  # Poisson
    Probability Distribution

cat("Answer is",PPD)
```

R code Exa 5.4a Hypergeometric Probability Distribution

```
                                          # Page no. : 222
                                              - 223

# Hypergeometric Probability Distribution

# Probability for 1 defective item

N <- 12
n <- 3
r <- 5
x <- 1

HPD <- dhyper(x = x,m = r,n = N-r,k = n)

cat("Answer is",HPD)

# Probability for atleast one defective item
```

```

HPD <- dhyper(x = 0,m = r,n = N-r,k = n)  # Probability for no defective item

cat("Answer is",1 - HPD)  # Probability for atleast one defective item

# Expected value, Variance and Standard Deviation

expected_value <- n * (r / N)
variance <- expected_value * (1 -(r/N)) * ((N-n)/(N-1))
standard_deviation <- sqrt(variance)

cat("Expected value",expected_value)
cat("Variance",variance)
cat("standard deviation",standard_deviation)
```

R code Exa 5.5a Expected Value and Variance

```
                                        # Page no. : 203 - 204

# Expected Value and Variance

x <- c(0,1,2,3,4,5)
prob_of_x <- c(0.18,0.39,0.24,0.14,0.04,0.01)
expected_value <- x * prob_of_x

DF <- data.frame(x, prob_of_x, expected_value)

expected_mean <- sum(DF$expected_value)

cat("Expected mean is",expected_mean)

```

```
deviation <- DF$x - expected_mean
sq_deviation <- deviation ** 2
expected_value2 <- DF$prob_of_x * sq_deviation

DF <- cbind(DF, deviation, sq_deviation, expected_
    value2)
View(DF)

expected_variance <- sum(DF$expected_value2)

cat("Expected variance is", expected_variance)

expected_SD <- sqrt(expected_variance)

cat("Expected SD is", expected_SD)
```

Chapter 6

Continuous Probability Distribution

R code Exa 6.1a Uniform Probability Distribution

```
                                        # Page no. : 234 - 236

# Uniform Probability Distribution

a <- 120
b <- 140
fun_over_x <- 1/20

# Since uniform probability is symmetric we can split it into left and right parts which
# are symmetric in nature

c <- 130  # (120+140)/2 = 130

# Probability of uniform probability  distribution is the area of the figure (rectangle)

# Area for the left symmetric part of the figure
```

```r

area <- punif(c,a,b)
area_full <- 2 * area

cat("Probability is",area_full)

# Expected value, Variance and standard Deviation

expected_value <- (a + b) / 2
variance <- (b - a) ** 2 / 12
standard_deviation <- sqrt(variance)

cat("Expected value is",expected_value)
cat("Variance is",variance)
cat("Standard deviation is",standard_deviation)
```

R code Exa 6.2a Normal Probability Distribution

```r
                                        # Page no. : 246 - 247

# Normal Probability Distribution

mean <- 36500
sigma <- 5000
x <- 40000
probability <- pnorm(40000, mean=36500, sd=5000, lower.tail= F)

cat("The probability of x exceed 40000 is", probability)

                                        # Page no. : 247

probability <- 0.10
```

```
z_value <- round(qnorm(probability), 2)   # Round it to 2 decimal place
x <- (sigma * z_value) + mean

cat("Value of x for not more than 10% of area is selected is",x)
```

R code Exa 6.3a Normal Approximation of Binomial Probabilities

```
                                                  # Page no. : 251 - 252

# Normal Approximation of Binomial Probabilities

n <- 100
p <- 0.1
q <- 1 - p

mu <- n * p
sigma <- sqrt(mu * q)

# P(x = 12)  ==>  P(11.5 <= x <= 12.5)

x1 <- 12.5
x2 <- 11.5
z_value1 <- (x1 - mu) / sigma
z_value2 <- (x2 - mu) / sigma

area1 <- pnorm(z_value1)
area2 <- pnorm(z_value2)

diff <- area1 - area2

cat("The normal approximation to the probability of 12 successes in 100 trials is ",diff)
```

```
# Probability for 13

x <- 13.5
z <- (x - mu) / sigma
ans <- pnorm(z)

cat("Answer is", ans)
```

R code Exa 6.4a Exponential Probability Distribution

```
                                        # Page no. :
                                            255

# Exponential Probability Distribution

mu <- 15
x1 <- 6
x2 <- 18

# P(x <= 6)

EPD <- pexp(x1,1/mu)

# P(x <= 18)

EPD2 <- pexp(x2, 1/mu)

diff <- EPD2 - EPD

cat("The probability that loading a truck will take
    between 6 and 18 minutes is ",diff)

SD <- mu
sigma <- SD ** 2
```

```
cat("Variance is", sigma)
```

Chapter 7

Sampling and Sampling Distribution

R code Exa 7.1a Point Estimator

```
                                            # Page no
                                              . :
                                              274

# Sample Mean and Sample Standard Deviation

annual_salary <- c
   (49094.30,53263.90,49643.50,49894.90,47621.60,55924.00,49092.30,51

                  55109.70,45922.60,57268.40,55688.80,51564.70,5618

                  51932.60,52973.00,45120.90,51753.00,54391.80,5016

                  50979.40,55860.90,57309.10)

program <- c("Yes","Yes","Yes","Yes","No","Yes","Yes
   ","Yes","Yes","Yes","Yes","No","Yes","No","No","
   Yes",
            "No","Yes","Yes","Yes","Yes","Yes","No","
```

```
No","No","No","No","Yes","Yes","No")

dataset <- data.frame(annual_salary, program)

sample_mean <- mean(dataset$annual_salary)
sample_sd <- sd(dataset$annual_salary)

cat("Sample mean of the data is", sample_mean)
cat("Sample satndard deviation is", sample_sd)

# Note that : Book SD is different from our SD

# Sample Proportion

n <- nrow(dataset)
x <- 19

sample_proportion <- x / n

cat("Sample Proportion is",sample_proportion)
```

R code Exa 7.2a Sampling Distribution

```
                                        # Page no. 277
                                          - 278

# Sampling Distribution

mean_annual_salary <- c("49500.00-49999.99","
    50000.00-50499.99","50500.00-50999.99","
    51000.00-51499.99",
                        "51500.00-51999.99","
                          52000.00-52499.99","
                          52500.00-52999.99","
                          53000.00-53499.99",
```

```
                                "53500.00-53999.99")
frequency <- c(2,16,52,101,133,110,54,26,6)
relative_frequency <- c
    (.004,.032,.104,.202,.266,.220,.108,.052,.012)

DF <- data.frame(mean_annual_salary,frequency,
    relative_frequency)

library(ggplot2)

ggplot(DF,aes(mean_annual_salary,relative_frequency)
    ) +
  geom_histogram(stat = "identity", fill = "purple")
      + labs(title = "Relative Frequency Histogram",
                                                x = " Mean Salary",
                                                     y = "frequency"
                                                   )
```

R code Exa 7.3a Sampling Distribution of Sample Mean

```
                                                  # Page no. :
                                                      281

# Sampling Distribution of xbar

sigma <- 4000
N <- 2500
n <- 30
x <- n / N

if(x > 0.05)    # Condition to include finite
    population factor or not (< 5%)
{
  standard_error <- sqrt((N-n)/(N-1)) * (sigma /
      sqrt(n))
```

```
} else{
  standard_error <- sigma / sqrt(n)
}

cat("Standard deviation of sample mean is",standard_error)

                                        # Page no. : 284

# To find probability that xbar is between 51300 and 52300

xbar1 <- 52300
xbar2 <- 51300
mu <- 51800

z1 <- (xbar1 - mu) / standard_error
z2 <- (xbar2 - mu) / standard_error

p1 <- pnorm(z1, lower.tail = T)
p2 <- pnorm(z2, lower.tail = T)

diff <- p1 - p2

cat("Probability that xbar is between 51300 and 52300 is", diff)
```

R code Exa 7.3b Relationship between Sample Size and Sampling Distribution of xbar

```
                                        # Page no. : 285 - 286

# Relationship between Sample Size and Sampling
```

```
   Distribution of xbar

sigma <- 4000
n <- 100
population_mean <- 51800

standard_error <- sigma / sqrt(n)

xbar1 <- 52300
xbar2 <- 51300

z1 <- (xbar1 - population_mean) / standard_error
z2 <- (xbar2 - population_mean) / standard_error

p1 <- pnorm(z1, lower.tail = T)
p2 <- pnorm(z2, lower.tail = T)

diff <- p1 - p2

cat("Probability that xbar is between 51300 and
   52300 with increased sample size is", diff)
```

R code Exa 7.4a Sampling Distribution of Sample Proportion

```
                                          # Page no.
                                             : 290

# Sampling Distribution of pbar

population_proportion <- 0.60
n <- 30
N <- 2500

x <- n / N

```

```
if(x > 0.05)   # Condition to include finite population or not ( < 5%)
{
  standard_deviation <- sqrt((N-n)(N-1)) * sqrt((population_proportion *
                                                  (1 - population_proportion)) / n)
} else{
  standard_deviation <- sqrt((population_proportion * (1 - population_proportion)) / n)
}

cat("Standard deviation for sample proportion is", standard_deviation)
```

R code Exa 7.4b Practical value of the Sampling Distribution of Sample Proportion

```
                                         # Page no. : 291

# Practical value of the Sampling Distribution of pbar
```

```
population_proportion <- 0.60
standard_error <- 0.0894
sample_proportion <- 0.65

z_value <- (sample_proportion - population_
    proportion) / standard_error

prob1 <- pnorm(sample_proportion, population_
    proportion, standard_error, lower.tail =T)

sample_proportion2 <- 0.55

z_value <- (sample_proportion2 - population_
    proportion) / standard_error

prob2 <- pnorm(sample_proportion2, population_
    proportion, standard_error, lower.tail =T)

final_prob <- prob1 - prob2

cat("The final probability is",final_prob)
```

R code Exa 7.4c Practical value of the Sampling Distribution of Sample Proportion Eg2

```
                                    # Page no. : 292 -
                                        293

# Practical value of the Sampling Distribution of
    pbar Eg-2

population_proportion <- 0.60
n <- 100

```

```
standard_error <- sqrt(population_proportion*(1 - population_proportion)/(n))

sample_proportion <- 0.65

z_value <- (sample_proportion - population_proportion) / standard_error

prob1 <- pnorm(sample_proportion, population_proportion, standard_error, lower.tail =T)

sample_proportion2 <- 0.55

z_value <- (sample_proportion2 - population_proportion) / standard_error

prob2 <- pnorm(sample_proportion2, population_proportion, standard_error, lower.tail =T)

final_prob <- prob1 - prob2

cat("The final probability is",final_prob)
```

Chapter 8

Interval Estimation

R code Exa 8.1a Population Mean Sigma Known

```
                                                # Page no. :
                                                   310 - 314
# Population Mean Sigma known

pop_sd <- 20
sample_size <- 100
sample_mean <- 82

standard_error <- pop_sd / sqrt(sample_size)

# 95% confidence interval

margin_of_error <- qnorm(0.975)*standard_error   #
   95% confidence interval --> 1 - 0.025 = 0.975

IE <- sample_mean + c(-margin_of_error, margin_of_
   error)

cat("The margin of error is given by", margin_of_
   error)
cat("The 95% interval estimate is given by", IE)
```

```
# 90% confidence interval

margin_of_error <- qnorm(0.95)*standard_error    # 90% confidence interval --> 1 - 0.05 = 0.95

IE <- sample_mean + c(-margin_of_error, margin_of_error)

cat("The margin of error is given by", margin_of_error)
cat("The 90% interval estimate is given by", IE)

# 99% confidence interval

margin_of_error <- qnorm(0.995)*standard_error    # 99% confidence interval --> 1 - 0.005 = 0.995

IE <- sample_mean + c(-margin_of_error, margin_of_error)

cat("The margin of error is given by", margin_of_error)
cat("The 99% interval estimate is given by", IE)
```

R code Exa 8.2a Population Mean Sigma Unknown

```
                                            # Page no. : 319 - 320


# Population Mean Sigma Unknown

cradit_card_balances <- c(9430,7535,4078,5604,5179,4416,10676,1627,10112,6567,13627,18719,
```

```
                              10544,13659,7061,6245,13021,9719,2200,1074

                              7917,11346,12806,4972,11356,7117,9465,1926

                              6845,10493,615,13627,12557,6232,9691,11448

                              12851,5337,8372,7445,11032,6525,5239,6195,


test <- t.test(cradit_card_balances)
IE1 <- test$conf.int[1]
IE2 <- test$conf.int[2]

cat("The 95% interval estimate is given by", IE1, "
   to", IE2)
```

R code Exa 8.2b Population Mean Sigma Unknown Eg2

```
                                             # Page no. :
                                               321 - 322

# Population Mean Sigma Unknown Eg-2

data <- c
   (52,44,55,44,45,59,50,54,62,46,54,42,60,62,43,42,48,55,57,56)



hist(data, col = "blue", main = "Histogram of
   Training Times", xlab = "Traing Times (days)",
     ylab = "Frequency")

test <- t.test(data)
IE1 <- test$conf.int[1]
```

```
IE2 <- test$conf.int[2]

cat("The 95% interval estimate is given by", IE1, " to", IE2)
```

R code Exa 8.3a Determining the Sample Size

```
                                        # Page no. : 326 - 327

# Determining the Sample Size

margin_of_error <- 2
z_value <- 1.96   # 95% Level of Confidence Interval

sample_standard_deviation <- 9.65

sample_size <- ((z_value)**2) * ((sample_standard_deviation)**2) / (margin_of_error)**2

cat("Sample size is",ceiling(sample_size))
# If Sample Size is not integer then we round up to next higher integer
```

R code Exa 8.4a Population Proportion

```
                                        # Page no. : 329

# Population Proportion

N <- 900
n <- 396
```

```

p <- n / N
q <- 1 - p

# Confidence Interval is 95%

z_value <- qnorm(0.975)   # 95% confidence interval
   —> 1 − 0.025 = 0.975

margin_of_error <- z_value * sqrt((p*q)/N)
IE <- p + c(-margin_of_error, margin_of_error)

cat("The margin of error is given by", margin_of_
   error)
cat("The 95% interval estimate is given by", IE)
```

R code Exa 8.4b Determining the Sample Size

```
                                          # Page no. :
                                             330

# Determining the Sample Size

margin_of_error <- 0.025
z_value <- 1.96   # 95% Level of Confidence Interval

p <- 0.44
q <- 1 - p

sample_size <- ((z_value)**2 * p * q) / (margin_of_
   error)**2

cat("Sample size is",ceiling(sample_size))
# If Sample Size is not integer then we round up to
   next higher integer
```

```
                                                   # Page no. : 331

margin_of_error <- 0.025
z_value <- 1.96   # 95% Level of Confidence Interval

p <- 0.50
q <- 1 - p

sample_size <- ((z_value)**2 * p * q) / (margin_of_error)**2

cat("Sample size is",ceiling(sample_size))
# If Sample Size is not integer then we round up to next higher integer
```

Chapter 9

Hypothesis Testing

R code Exa 9.1a Population Mean Sigma Known One Tailed Test

```
                                            # Page no. :
                                              359 – 360

# Population Mean Sigma Known One Tailed Test

sigma <- 0.18
n <- 36
xbar <- 2.92
mu <- 3

z_value <- (xbar - mu) / (sigma / sqrt(n))
z_value <- round(z_value,2)

alpha <- 0.01

# P Value Approch (Lower–Tail Test)

pval <- pnorm(z_value)

if(pval > alpha)
{
```

```
  cat("Since p-value ",pval ,"is greater than 0.01,
     therefore we will accept null hypothesis")
} else {
  cat("Since p-value ",pval ,"is less than 0.01,
     therefore we will reject null hypothesis and
     accept
      alternative hypothesis.")
}

# Critical Value Approch (Lower-Tail Test)

z_alpha <- qnorm(1 - (alpha))   # Area of 0.01 to
    the left (1 - 0.01 = 0.99)
critical_approch <- -z_alpha

if(z_value <= critical_approch)
{
  cat("Since z-value", z_value ,"is less then or
     equal to", critical_approch , "therefore we
     reject
      the null hypothesis and accept the alternative
         hypothesis.")
} else{
  cat("Since z-value", z_value ,"is more than",
     critical_approch,"therefore we accept the null
     hypothesis.")
}
```

R code Exa 9.1b Population Mean Sigma Known Two Tailed Test

```
                                          # Page no. :
                                             362 -
                                            364

# Population Mean Sigma Known Two Tailed Test
```

```
sample_size <- 50
sample_mean <- 297.6
population_mean <- 295
significance_level <- 0.05    # alpha
population_sd <- 12

                                                  # Page no. : 364

## Critical value approch

z_value <- (sample_mean - population_mean) / (population_sd / sqrt(sample_size))

z_half_alpha <- qnorm(1 - (significance_level/2))
critical_value_1 <- -z_half_alpha
critical_value_2 <- z_half_alpha

if(z_value >= critical_value_2 || z_value <= critical_value_1)
{
  cat("Since z-value", z_value ,"does not lie in the range", critical_value_1 ,"and", critical_value_2
      ,"therefore we reject the null hypothesis and accept the alternative hypothesis.")
} else{
  cat("Since z-value", z_value ,"lies in the range", critical_value_1 ,"and", critical_value_2
      ,"therefore we accept the null hypothesis.")
}

                                                  # Page no. : 363 - 364

## P-value approch

```

```
area_under_curve <- 1 - pnorm(z_value)
pval <- 2 * area_under_curve   # P-value
if(pval > 0.05)
{
  cat("Since p-value ",pval ,"is greater than 0.05,
     therefore we will accept null hypothesis")
} else {
  cat("Since p-value ",pval ,"is less than 0.05,
     therefore we will reject null hypothesis and
     accept
      alternative hypothesis.")
}
```

R code Exa 9.1c Relationship between Interval Estimation and Hypothesis Testing

```
                                        # Page no. :
                                            366 - 367

# Relationship between Interval Estimation and
   Hypothesis Testing

mu <- 295
alpha <- 0.05
n <- 50
xbar <- 297.6
sigma <- 12

z_value <- 1.96     # alpha = 0.05 so alpha/2 = 0.05/
   2 = 0.025 (z-value is for 0.025)

# 95% Confidence Interval

margin_of_error <- z_value * (sigma / sqrt(n))

```

```
IE <- xbar + c(-margin_of_error, margin_of_error)  # Interval Estimate

cat("Interval estimate for 95% Confidence Interval is",IE)

if(mu >= IE[2] || mu <= IE[1])
{
  cat("Null Hypothesis is rejected..")
} else{
  cat("Null Hypothesis cannot be rejected..")
}
```

R code Exa 9.2a Population Mean Sigma Unknown One Tailed Test

```
                                        # Page no. : 371 - 372

# Population Mean Sigma Unknown One Tailed Test

mu0 <- 7
alpha <- 0.05
xbar <- 7.25
s <- 1.052
n <- 60

t_value <- (xbar - mu0) / (s / sqrt(n))

df <- n - 1

# Upper Tail Test

pval <- pt(t_value,df = df,lower.tail = F)   # Book answer is 0.354
```

```
if(pval > alpha)
{
  cat("Since p-value ",pval ,"is greater than 0.05,
     therefore we will accept null hypothesis")
} else {
  cat("Since p-value ",pval ,"is less than 0.05,
     therefore we will reject null hypothesis and
     accept
      alternative hypothesis.")
}
```

R code Exa 9.2b Population Mean Sigma Unknown Two Tailed Test

```
                                                    # Page no. :
                                                       372 - 373

# Population Mean Sigma Unknown Two Tailed Test

mu0 <- 40
alpha <- 0.05
xbar <- 37.4
s <- 11.79
n <- 25

t_value <- (xbar - mu0) / (s / sqrt(n))

df <- n - 1  # Degree of Freedom

# Two Tail Test
# P - value Approach

pval <- 2 *(1 - pt(t_value,df = df,lower.tail = F))
     # Book answer is  0.2822

```

```
if(pval > alpha)
{
  cat("Since p-value ",pval ,"is greater than 0.05,
     therefore we will accept null hypothesis")
} else {
  cat("Since p-value ",pval ,"is less than 0.05,
     therefore we will reject null hypothesis and
     accept
      alternative hypothesis.")
}

# Critical Value Approach

t_half_alpha <- qt(1 - (alpha/2),df)
critical_value_1 <- -t_half_alpha
critical_value_2 <- t_half_alpha

if(t_value >= critical_value_2 || t_value <=
   critical_value_1)
{
  cat("Since t-value", t_value ,"does not lie in the
      range", critical_value_1 ,"and", critical_
     value_2
      ,"therefore we reject the null hypothesis and
         accept the alternative hypothesis.")
} else{
  cat("Since t-value", t_value ,"lies in the range",
      critical_value_1 ,"and", critical_value_2
      ,"therefore we accept the null hypothesis.")
}
```

R code Exa 9.3a Population Proportion

```
                                        # Page no. :
                                          377 - 378
```

```
# Population Proportion

p0 <- 0.20
alpha <- 0.05
n <- 400
x <- 100

p_bar <- x / n

z_value <- (p_bar - p0) / sqrt((p0*(1 - p0)) / n)

# Upper Tail Test

# P-value Approach

pval <- pnorm(z_value,lower.tail = F)
if(pval > alpha)
{
  cat("Since p-value ",pval ,"is greater than 0.05,
     therefore we will accept null hypothesis")
} else {
  cat("Since p-value ",pval ,"is less than 0.05,
     therefore we will reject null hypothesis and
     accept
      alternative hypothesis.")
}

# Critical Value Approach

z_alpha <- qnorm(1 - alpha)
critical_value <- z_alpha

if(z_value >= critical_value)
{
  cat("Since z-value", z_value ,"is greater then or
     equal to", critical_value , "therefore we
     reject
```

```
      the null hypothesis and accept the alternative
          hypothesis.")
} else{
  cat("Since z-value", z_value ,"is less than",
     critical_value,"therefore we accept the null
     hypothesis.")
}
```

R code Exa 9.4a Calculating the Probability of Type Second Errors

```
                                        # Page no. :
                                          382 - 383

# Calculating the Probability of Type Second Errors

mu0 <- 120
alpha <- 0.05
z_value <- 1.645
n <- 36
sigma <- 12

xbar <- mu0 - z_value * (sigma / sqrt(n))

mu <- 112
z <- (xbar - mu) / (sigma / sqrt(n))

# Upper Tail Test

beta_value <- pnorm(z, lower.tail = F)

cat("The type 2nd error is",beta_value)

                                    # Page no. : 384

xbar <- 116.71
```

```
mu <- 115
sigma <- 12
n <- 36

z <- (xbar - mu) / (sigma / sqrt(n))

# Upper Tail Test

beta_value <- pnorm(z, lower.tail = F)

cat("The type 2nd error is",beta_value)
```

R code Exa 9.5a Determining the Sample Size

```
                                        # Page no. :
                                           389

# Determining the Sample Size

alpha <- 0.05
beta <- 0.10
z_alpha <- 1.645
z_beta <- 1.28
mu0 <- 120
mua <- 115
sigma <- 12

n <- ((z_alpha + z_beta)**2) * (sigma)**2 / (mu0 -
   mua)**2    # Sample Size
n <- ceiling(n)
cat("Sampling Size is",n)
```

Chapter 10

Inference About Means and Proportions With Two Populations

R code Exa 10.1a Inference about the Difference between the two Population Means Sigma 1 and Sigma 2 known

```
                                        # Page no. : 410

# Inference about the Difference between the two Population Means Sigma 1 and Sigma 2 known

sigma1 <- 9
sigma2 <- 10
sample_size1 <- 36
sample_size2 <- 49
sample_mean1 <- 40
sample_mean2 <- 35

point_estimate <- sample_mean1 - sample_mean2

z_value <- qnorm(0.975)   # alpha/2 = 0.05/2 = 0.025
```

```
= 1− 0.025 = 0.975

standard_error <- sqrt((((sigma1)^2)/(sample_size1)) + (((sigma2)^2)/(sample_size2)))

IE1 <- point_estimate + z_value*standard_error
IE2 <- point_estimate - z_value*standard_error

cat("The interval estimation for the given information at 95% confidence level is ",IE2 ,"to "
    , IE1)
```

R code Exa 10.1b Hypothesis Tests About Difference between two Means

```
                                    # Page no. : 410 − 412

# Hypothesis Tests About Difference between two Means

sigma1 <- 10
sigma2 <- 10
alpha <- 0.05
n1 <- 30
n2 <- 40
xbar1 <- 82
xbar2 <- 78
D0 <- 0

z_value <- ((xbar1 - xbar2) - D0) / sqrt(((sigma1)**2/n1) + ((sigma2)**2/n2))

# P−value Approach

```

```
# Two Tail Test

pval <- 2 * pnorm(z_value, lower.tail = F)

if(pval <= alpha)
{
  cat("Since P-Value",pval,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis")
} else {
  cat("Since P-Value",pval,"is more than 0.05
     therefore we cannot reject Null Hypothesis")
}

# Critical Value Approach

z_half_alpha <- qnorm(0.975)   # alpha/2 = 0.05/2 =
   0.025 = 1- 0.025 = 0.975
critical_value_1 <- -z_half_alpha
critical_value_2 <- z_half_alpha

if(z_value >= critical_value_2 || z_value <=
   critical_value_1)
{
  cat("Since Z-value",z_value,"does not lie in the
     range",critical_value_1,"to",critical_value_2,
      "therefore we can reject Null Hypothesis")
} else {
  cat("Since Z-value",z_value,"lie in the range",
     critical_value_1,"to",critical_value_2,
      "therefore we cannot reject Null Hypothesis")
}
```

R code Exa 10.2a Inference about the Difference between the two Population Means Sigma 1 and Sigma 2 Unknown

```
                                                # Page no. :
                                                  415 - 417

# Inference about the Difference between the two
   Population Means Sigma 1 and Sigma 2 Unknown

s1 <- 150
s2 <- 125
n1 <- 28
n2 <- 22
xbar1 <- 1025
xbar2 <- 910

point_estimate <- xbar1 - xbar2

numerator <- ((((s1)**2 /n1) + ((s2)**2 /n2))**2)
denomenator <- ((1 /(n1 -1)) * (((s1)**2 / n1)**2))
   + ((1 /(n2 -1)) * (((s2)**2 / n2)**2))

df <- numerator / denomenator  # Degree of Freedom

t_value <- qt(0.975,df)   # alpha/2 = 0.05/2 = 0.025
    = 1- 0.025 = 0.975

standard_error <- sqrt((((s1)^2)/(n1)) + (((s2)^2)/(
   n2)))

IE1 <- point_estimate + t_value*standard_error
IE2 <- point_estimate - t_value*standard_error

cat("The interval estimation for the given
   information at 95% confidence level is ",IE2 ,
    "to", IE1)
```

R code Exa 10.2b Hypothesis Tests About Difference between two Means

```
# Page no. : 418 - 419

# Hypothesis Tests About Difference between two Means

x <- c(300, 280, 344, 385, 372, 360, 288, 321, 376, 290, 301, 283)
y <- c(274, 220, 308, 336, 198, 300, 315, 258, 318, 310, 332, 263)

DF <- data.frame(x,y)

test <- t.test(DF$x, DF$y, paired = F, alternative = "greater")
test

# Upper Tail Test

if(test$p.value <= 0.05)    # 95% Confidence Level
{
  cat("Since P-Value",test$p.value,"is less than or equal to 0.05 therefore we can reject
    Null Hypothesis")
} else {
  cat("Since P-Value",test$p.value,"is more than 0.05 therefore we cannot reject Null Hypothesis")
}
```

R code Exa 10.3a Inference About the Difference Between Two Population Means Matched samples

```r
# Page no. : 424 - 425

# Inference About the Difference Between Two Population Means Matched samples

workers <- c(1,2,3,4,5,6)
method_1 <- c(6.0,5.0,7.0,6.2,6.0,6.4)
method_2 <- c(5.4,5.2,6.5,5.9,6.0,5.8)
diff <- method_1 - method_2

dataFrame <- data.frame(workers, method_1, method_2, diff)

test <- t.test(dataFrame$method_1, dataFrame$method_2, paired = T)
test

# Two Tail Test

if(test$p.value <= 0.05)
{
  cat("Since P-Value",test$p.value,"is less than or equal to 0.05 therefore we can
      reject Null Hypothesis")
} else {
  cat("Since P-Value",test$p.value,"is more than 0.05 therefore we cannot reject Null Hypothesis")
}

# Interval Estimate

IE1 <- test$conf.int[1]
IE2 <- test$conf.int[2]

cat("The interval estimation for the given information at 95% confidence level is ",IE1 ,"to
```

```
    ",
    IE2)
```

R code Exa 10.4a Inference About the Difference Between Two Population Proportions

```
                                        # Page no. :
                                            431

# Inference About the Difference Between Two
    Population Proportions

n1 <- 250
n2 <- 300
x1 <- 35
x2 <- 27
alpha <- 0.1

pbar1 <- x1 / n1
pbar2 <- x2 / n2

qbar1 <- 1 - pbar1
qbar2 <- 1 - pbar2

# Interval Estimation

diff_prop <- pbar1 - pbar2

z_half_alpha <- qnorm(0.95)   # alpha / 2 = 0.1 /
    2 = 0.05 = 0.95 (1 - 0.05)

margin_of_error <- z_half_alpha * sqrt((pbar1 *
    qbar1)/n1 + (pbar2 * qbar2)/n2 )

IE1 <- diff_prop + margin_of_error
```

```
IE2 <- diff_prop - margin_of_error

cat("The interval estimation for the given
   information at 90% confidence level is ",IE2 ,"to
   ", IE1)
```

R code Exa 10.4b Hypothesis Tests About Difference between two Proportions

```
                                              # Page no. :
                                                 432 - 433

# Hypothesis Tests About Difference between two
   Proportions

pbar1 <- 0.14
pbar2 <- 0.09
n1 <- 250
n2 <- 300
alpha <- 0.10   # Significance Level

pbar <- ((n1 * pbar1) + (n2 * pbar2)) / (n1 + n2)
   # Pooled Estrimator

z_value <- (pbar1 - pbar2) / sqrt((pbar*(1 - pbar))*
   ((1/n1)+(1/n2)))

# Two Tail Test

pval <- 2 * pnorm(z_value, lower.tail = F)

if(pval <= alpha)
{
  cat("Since P-Value",pval,"is less than or equal to
     0.10 therefore we can reject Null Hypothesis")
```

```
} else {
  cat("Since P-Value",pval,"is more than 0.10
     therefore we cannot reject Null Hypothesis")
}
```

Chapter 11

Inferences About Population Variances

R code Exa 11.1a Inferences About Population Variance

```
                                            # Page no. :
                                                453

# Inferences About the Population Variance

n <- 20
variance <- 0.0025
df <- 19   # Degrees of Freedom

chisq_1 <- qchisq(0.975,df)
chisq_2 <- qchisq(0.025,df)

# Interval Estimation

IE1 <- sqrt((df * variance) / chisq_1)
IE2 <- sqrt((df * variance) / chisq_2)

cat("Interval Estimation at 95% confidence interval
   for population standard deviation is",IE1,
```

```
"to",IE2)
```

R code Exa 11.1b Hypothesis Testing

```
                                        # Page no. : 455 - 456

# Hypothesis Testing

n <- 24
alpha <- 0.05   # Significance Level
sigma0_sq <- 4
variance <- 4.9

chisq_value <- ((n - 1) * variance) / sigma0_sq

df <- n - 1   # Degree of Freedom

# Upper Tail Test

# P-value Approach

pval <- pchisq(chisq_value, df, lower.tail = F)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to 0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05 therefore we cannot reject Null Hypothesis")
}

# Critical Value Approach

```

```
chisqvalue <- qchisq(0.95,df)   # Chi-square value
   for 0.95 (1 - 0.05)

if(chisq_value >= chisqvalue)
{
  cat("Since Chi-square value",chisq_value,"is
     greater than or equal to Chi-square value",
     chisqvalue,
      "therefore we will reject Null Hypothesis")
} else {
  cat("Since Chi-square value",chisq_value,"is less
     than Chi-square value",chisqvalue,
      "therefore we cannot reject Null Hypothesis")
}


                                        # Page no. :
                                           456 -
                                          457

n <- 30
alpha <- 0.05   # Significance Level
sigma0_sq <- 100
variance <- 162

chisq_value <- ((n - 1) * variance) / sigma0_sq

df <- n - 1   # Degree of Freedom

# Two Tail Test

# P-value Approach

pval <- 2 * pchisq(chisq_value, df, lower.tail = F)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to
```

```
    0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05
    therefore we cannot reject Null Hypothesis")
}
```

R code Exa 11.2a Inferences About Two Population Variances

```
                                          # Page no. :
                                            462 - 464

# Inferences About Two Population Variances

alpha <- 0.10   # Significance Level
n1 <- 26
n2 <- 16
sv1 <- 48
sv2 <- 20

f_value <- (sv1) / (sv2)

df1 <- n1 - 1   # Degrees of Freedom 1
df2 <- n2 - 1   # Degrees of Freedom 2

# Two Tail Test

# P-value Approach

pval <- 2 * pf(f_value,df1,df2,lower.tail = F)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to
    0.10 therefore we can reject Null Hypothesis")
} else{
```

```
  cat("Since pval",pval,"is greater than 0.10
     therefore we cannot reject Null Hypothesis")
}

# Critical Value Approach

half_alpha <- alpha / 2
fval <- qf(0.95,df1,df2)   # half_alpha = 0.05  = 1
    - 0.05 = 0.95

if(f_value >= fval)
{
  cat("Since F value",f_value,"is greater than or
     equal to F value",fval,
      "therefore we will reject Null Hypothesis")
} else {
  cat("Since F value",f_value,"is less than F value"
     ,fval,
      "therefore we cannot reject Null Hypothesis")
}

                                        # Page no. :
                                          464

alpha <- 0.05   # Significance Level
n1 <- 41
n2 <- 31
sv1 <- 120
sv2 <- 80

f_value <- (sv1) / (sv2)

df1 <- n1 - 1   # Degrees of Freedom 1
df2 <- n2 - 1   # Degrees of Freedom 2

# Upper Tail Test

# P-value Approach
```

```r

pval <- pf(f_value,df1,df2,lower.tail = F)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to 0.10 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.10 therefore we cannot reject Null Hypothesis")
}

# Critical Value Approach

fval <- qf(0.90,df1,df2)   # alpha = 0.05  =  1 - 2 * 0.05 = 0.90

if(f_value >= fval)
{
  cat("Since F value",f_value,"is greater than or equal to F value",fval,
      "therefore we will reject Null Hypothesis")
} else {
  cat("Since F value",f_value,"is less than F value",fval,
      "therefore we cannot reject Null Hypothesis")
}
```

Chapter 12

Tests of Goodness of Fit and Independence

R code Exa 12.1a Goodness of Fit Test A Multinomial Population

```
                                        # Page no. :
                                          474 - 476

# Goodness of Fit Test A Multinomial Population

category <- c("Company A","Company B", "Company C")
prop <- c(0.30, 0.50, 0.20)
freq1 <- c(48, 98, 54)
freq2 <- c(60, 100, 40)
diff <- freq1 - freq2
sq_diff <- (diff)**2
answer <- sq_diff / freq2
DF <- data.frame(category, prop, freq1, freq2, diff,
    sq_diff, answer)

total_observe_freq <- sum(DF$freq1)
chisq_value <- sum(DF$answer)

alpha <- 0.05
```

```
df <- nrow(DF) - 1   # Degrees of Freedom

# Upper Tail Test

# P-value Approach

pval <- pchisq(chisq_value,df,lower.tail = F)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05
     therefore we cannot reject Null Hypothesis")
}

# Critical Value Approach

chisqValue <- qchisq(0.95,df)   # 1 - alpha = 1 -
   0.05 = 0.95

if(chisq_value >= chisqValue)
{
  cat("Since Chi-square value",chisq_value,"is
     greater than or equal to Chi-square value",
     chisqValue,
      "therefore we will reject Null Hypothesis")
} else {
  cat("Since Chi-square value",chisq_value,"is less
     than Chi-square value",chisqValue,
      "therefore we cannot reject Null Hypothesis")
}
```

R code Exa 12.2a Tests of Independence

```r
                                        # Page no. :
                                          480 - 482

# Tests of Independence

gender <- c("Male", "Female")
light <- c(20,30)
regular <- c(40,30)
dark <- c(20,10)
total <- c(80,70)

DF <- data.frame(gender, light, regular, dark, total
   )

test <- chisq.test(DF[,2:4])
test

# Upper Tail Test

if(test$p.value <= 0.05)
{
  cat("Since pval",test$p.value,"is less than or
     equal to 0.05 therefore we can reject
       Null Hypothesis")
} else{
  cat("Since pval",test$p.value,"is greater than
     0.05 therefore we cannot reject Null
       Hypothesis")
}
```

R code Exa 12.3a Goodness of Fit Test Poisson Distribution

```r
                                        # Page no. :
                                          488 - 490

```

```r
# Goodness of Fit Test Poisson Distribution

x <- c(0,1,2,3,4,5,6,7,8,9)
observed_freq <- c(2,8,10,12,18,22,22,16,12,6)
y <- x * observed_freq
e <- 2.72

mu <- sum(y) / sum(observed_freq)
fun_of_x <- round(((mu)**x) * (e)**(-mu) / factorial(x), 4)   # Function of x

expected_freq <- round(sum(observed_freq)*fun_of_x, 2)

diff <- observed_freq - expected_freq

sq_diff <- round((diff)**2,2)

answer <- round(sq_diff / expected_freq,2)

DF <- data.frame(x,observed_freq,expected_freq,diff,sq_diff,answer)

total_observe_freq <- sum(DF$observed_freq)
total_expexted_freq <- sum(DF$expected_freq)

chisq_value <- sum(DF$answer)

df <- nrow(DF) - 2   # Degrees of Freedom (k - p - 1 where k = 10, p = 1)

alpha <- 0.05

# Upper Tail Test

# P-value Approach

pval <- pchisq(chisq_value,df,lower.tail = F)
```

```
if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05
     therefore we cannot reject Null Hypothesis")
}

# In Book it is taken 9 categories but we consider
   10 so p-value may vary but answeris correct
```

R code Exa 12.3b Goodness of Fit Test Normal Distribution

```
                                    # Page no. :
                                       491 - 494

# Goodness of Fit Test Normal Distribution

data <- c
   (71,66,61,65,54,93,60,86,70,70,73,73,55,63,56,62,76,54,82,79,76,68

          61,61,64,65,62,90,69,76,79,77,54,64,74,65,65,61,56,63,80,5

xbar <- mean(data)
s <- sd(data)

cat("Value of mean is",xbar)
cat("Value of standard variance is",s)

percentage <- c(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9)
z <- c()
test_score <- c()
```

```r

for (i in 1:length(percentage)) {
 z[i] <- round(qnorm(percentage[i]),2)
 test_score[i] <- round(xbar + (z[i] * round(s,2)), 2)
}

DF <- data.frame(percentage, z, test_score)

interval <- c("Less than 55.10", "55.10 to 59.68", "59.68 to 63.01", "63.01 to 65.82",
              "65.82 to 68.42", "68.42 to 71.02", "71.02 to 73.83", "73.83 to 77.16",
              "77.16 to 81.74", "81.74 and over")
observed_freq <- c(5,5,9,6,2,5,2,5,5,6)
expected_freq <- c(5,5,5,5,5,5,5,5,5,5)

diff <- observed_freq - expected_freq
sq_diff <- (diff)**2

answer <- (sq_diff) / expected_freq

dataset <- data.frame(interval, observed_freq, expected_freq, diff, sq_diff, answer)

chisq <- sum(dataset$answer)

alpha <- 0.10   # Significance Level

df <- nrow(dataset) - 3   # Degrees of Freedom (k - p - 1 where k = 10, p = 2)

# Upper Tail Test

# P-value Approach

pval <- pchisq(chisq,df,lower.tail = F)

```

```
if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05
     therefore we cannot reject Null Hypothesis")
}
```

Chapter 13

Experimental Design and Analysis of Variance

R code Exa 13.1a Analysis of Variance and the Completely Randomized Design

```
                                        # Page no. : 510 - 512

# Analysis of Variance and the Completely Randomized Design

method_A <- c(58,64,55,66,67)
method_B <- c(58,69,71,64,68)
method_C <- c(48,57,59,47,49)

DF <- data.frame(method_A,method_B,method_C)

mean1 <- sum(DF$method_A) / nrow(DF)   # Sample Mean 1
mean2 <- sum(DF$method_B) / nrow(DF)   # Sample Mean 2
mean3 <- sum(DF$method_C) / nrow(DF)   # Sample Mean 3
```

```
variance1 <- sum((DF$method_A - mean1)**2) / (nrow(DF)-1)   # Sample Variance 1
variance2 <- sum((DF$method_B - mean2)**2) / (nrow(DF)-1)   # Sample Variance 2
variance3 <- sum((DF$method_C - mean3)**2) / (nrow(DF)-1)   # Sample Variance 3

sd1 <- sqrt(variance1)   # Sample Standard Variance 1
sd2 <- sqrt(variance2)   # Sample Standard Variance 2
sd3 <- sqrt(variance3)   # Sample Standard Variance 3

sample_mean <- (mean1 + mean2 + mean3) / 3   # Overall Sample Mean

variance <- ((mean1 - sample_mean)**2 + (mean2 - sample_mean)**2 + (mean3 - sample_mean)**2)  / (3 - 1)
# Sample Varince for Overall Sample Mean (3 --> Methods)

sigma_sq <- nrow(DF) * variance   # Between-treatment Estimate of Sigma Square

estimate_sigma_sq <- (variance1 + variance2 + variance3) / 3   # Within-treatment Estimate of Sigma Square

ratio <- sigma_sq / estimate_sigma_sq

cat("Ratio of Between-treatment Estimate of Sigma Square by Within-treatment Estimate
  of Sigma Square is",ratio)
```

R code Exa 13.2a Analysis of Variance and the Completely Randomized Design

```r
                                        # Page no. : 518

# Analysis of Variance and the Completely Randomized Design

method_A <- c(58,64,55,66,67)
method_B <- c(58,69,71,64,68)
method_C <- c(48,57,59,47,49)

DF <- data.frame(method_A,method_B,method_C)

k <- ncol(DF)   # Number of Treatments
n <- nrow(DF)   # Number of Observations for each Treatment
N <- n * k   # Total Observations

df_numerator <- k - 1   # Degrees of Freedom for Numerator
df_denomenator <- N - k   # Degrees of Freedom for Denomenator

alpha <- 0.05

x <- c(t(as.matrix(DF)))
f <- c("method_A", "method_B", "method_C")
tm <- gl(k, 1, n*k, factor(f))
result <- anova(lm(x ~ tm))   # Similar to aov(x ~ tm)

result

```

```
# Upper Tail Test

# Critical Value Approach

fval <- qf(0.95,df_numerator, df_denomenator)   # 1 - alpha = 1 - 0.05 = 0.95
fval <- round(fval, 2)

fvalue <- result$`F value`[1]

if(fvalue >= fval)
{
  cat("Since F value",fvalue,"is greater than or equal to F value",fval,
      "therefore we will reject Null Hypothesis")
} else {
  cat("Since fvalue value",fvalue,"is less than fvalue value",fval,
      "therefore we cannot reject Null Hypothesis")
}

# P-value Approach

pval <- pf(fvalue,df_numerator,df_denomenator,lower.tail = F)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to 0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05 therefore we cannot reject Null Hypothesis")
}
```

R code Exa 13.3a Multiple Comparison Procedures Fishers LSD

```r
                                              # Page no. : 525 - 526

# Multiple Comparison Procedures : Fisher's LSD

method_A <- c(58,64,55,66,67)
method_B <- c(58,69,71,64,68)
method_C <- c(48,57,59,47,49)

DF <- data.frame(method_A,method_B,method_C)

k <- ncol(DF)   # Number of Treatments
n <- nrow(DF)   # Number of Observations for each Treatment
N <- n * k   # Total Observations

x <- c(t(as.matrix(DF)))
f <- c("method_A", "method_B", "method_C")
tm <- gl(k, 1, n*k, factor(f))
result <- anova(lm(x ~ tm))   # Similar to aov(x ~ tm)

# Approach - 1

# Fisher's LSD Procedure for Method A and Method B

t_value1 <- (mean(DF$method_A) - mean(DF$method_B)) / sqrt(result$`Mean Sq`[2] * ((1 / n)

```

```
t_value1 <- round(t_value1, 2)

df <- N - k   # Degrees of Freedom

# Two Tail Test

# P-value Approach

alpha <- 0.05

pval <- 2 * pt(t_value1, df, lower.tail = T)

if(pval <= alpha)
{
  cat("Since pval",pval,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05
     therefore we cannot reject Null Hypothesis")
}

tval <- qt(0.975,df)    # alpha/2 = 0.05 / 2 =
   0.025 = (1 - 0.025) = 0.975
tval <- round(tval, 3)

LSD <- tval * sqrt(result$`Mean Sq`[2] * ((1 / n) +
   (1 / n)))
LSD <- round(LSD, 2)

# Approach - 2

# Fisher's LSD Procedure for Method A and Method C

diff_A_C <- mean(DF$method_A)- mean(DF$method_C)
```

```
if(diff_A_C > LSD)
{
  cat("Since the value of difference", diff_A_C,"is
     greater than LSD",LSD,"therefore we will reject
       Null Hypothesis")
} else{
  cat("Since the value of difference", diff_A_C,"is
     less than LSD",LSD,"therefore we cannot reject
       Null Hypothesis")
}

# Fisher's LSD Procedure for Method B and Method C

diff_B_C <- mean(DF$method_B)- mean(DF$method_C)

if(diff_B_C > LSD)
{
  cat("Since the value of difference", diff_B_C,"is
     greater than LSD",LSD,"therefore we will reject
       Null Hypothesis")
} else{
  cat("Since the value of difference", diff_B_C,"is
     less than LSD",LSD,"therefore we cannot reject
       Null Hypothesis")
}
```

R code Exa 13.4a Randomized Block Design

```
                                              # Page no. :
                                                  534

# Randomized Block Design

blocks <- c("Controller 1", "Controller 2", "
```

```
Controller 3", "Controller 4", "Controller 5", "Controller 6")
system_A <- c(15, 14, 10, 13, 16, 13)
system_B <- c(15, 14, 11, 12, 13, 13)
system_C <- c(18, 14, 15, 17, 16, 13)

DF <- data.frame(blocks, system_A, system_B, system_C)

k <- ncol(DF) - 1  # Number of Treatments (blocks is not the treatement)
b <- nrow(DF)  # Number of blocks
N <- k * b  # Total sample Size

x <- c(t(as.matrix(DF[,-1])))
f <- c("system_A", "system_B", "system_C")
tm <- gl(k, 1, N, factor(f))
blk <- gl(b, k, N)
result <- anova(lm(x ~ tm + blk))  # Similar to aov (x ~ tm)

result

f_value <- result$`F value`[1]

# Upper Tail Test

# P-value Approach

pval <- result$`Pr(>F)`[1]

if(pval <= 0.05)
{
  cat("Since pval",pval,"is less than or equal to 0.05 therefore we can reject Null Hypothesis")
} else{
  cat("Since pval",pval,"is greater than 0.05 therefore we cannot reject Null Hypothesis")
```

```
}
```

R code Exa 13.5a Factorial Design

```
                                   # Page no. : 542

# Factorial Experiment

program <- c("Three-hour review", "Three-hour review", "One-day program", "One-day program", "10-week course",
             "10-week course")
business <- c(500, 580, 460, 540, 560, 600)
engineering <- c(540, 460, 560, 620, 600, 580)
arts_and_science <- c(480, 400, 420, 480, 480, 410)

DF <- data.frame(program, business, engineering, arts_and_science)

a <- 3   # Number of levels in Factor A (Unique Programs)
b <- 3   # Number of levels in Factor B (Columns Except program)
r <- 2   # Number of Replications (Each Program has 2 Replications)
N <- a * b * r   # Total Observations

x <- c(t(as.matrix(DF[,-1])))
f1 <- c("Three-hour review", "One-day program", "10-week course")
f2 <- c("business", "engineering", "arts and science")
tm2 <- gl(a, 1, N, factor(f1))
tm1 <- gl(b, r * a, N, factor(f2))
result <- anova(lm(x ~ tm1 * tm2))   # Similar to
```

```
   aov(x ~ tm)

result

alpha <- 0.05

# Upper Tail Test

pval1 <- result$`Pr(>F)`[1]
pval2 <- result$`Pr(>F)`[2]
pval3 <- result$`Pr(>F)`[3]

if(pval1 <= alpha)
{
  cat("Since pval",pval1,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis
      for Undergraduation.")
} else{
  cat("Since pval",pval1,"is greater than 0.05
     therefore we cannot reject Null Hypothesis
      for Undergraduation.")
}

if(pval2 <= alpha)
{
  cat("Since pval",pval2,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis
      for programs.")
} else{
  cat("Since pval",pval2,"is greater than 0.05
     therefore we cannot reject Null Hypothesis
      for programs.")
}

if(pval3 <= alpha)
{
  cat("Since pval",pval3,"is less than or equal to
     0.05 therefore we can reject Null Hypothesis
```

```
      for interaction.")
} else{
  cat("Since pval",pval3,"is greater than 0.05
     therefore we cannot reject Null Hypothesis
      for interaction.")
}
```

Chapter 14

Simple Linear Regression

R code Exa 14.1a Least Squares Method

```
                                        # Page no. : 565
                                            - 566

# Least Squares Method

restaurant <- c(1,2,3,4,5,6,7,8,9,10)
student_population <- c(2,6,8,8,12,16,20,20,22,26)
quartely_sales <- c
   (58,105,88,118,117,137,157,169,149,202)

DF <- data.frame(restaurant, student_population,
   quartely_sales)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

```

```
ggplot(DF,aes(student_population, quartely_sales)) +
    geom_point() +
  labs(title = "Scatter Plot between Student
     Population and Quartely Sales",x = "Student
       Population (1000s)", y = "Quartely Sales ($
          1000s)")

                                    # Page no. : 567 -
                                        569

regressor <- lm(quartely_sales ~ student_population,
    data = DF)
res <- summary(regressor)

res

b1 <- res$coefficients[[2]]

b0 <- res$coefficients[[1]]

cat("Estimated Regression Equation is y_cap =",b0,"+
   ",b1,"x")

ggplot(DF,aes(student_population, quartely_sales)) +
    geom_point() +
  geom_smooth(method='lm', se = F) + labs(title = "
     Scatter Plot between Student Population
and Quartely Sales",x = "Student Population (1000s)"
   , y = "Quartely Sales ($1000s)")
```

R code Exa 14.2a Coefficient of Determination

```
                                          # Page no. :
                                             576 -
                                            580
```

```r

# Coefficient of Determination

restaurant <- c(1,2,3,4,5,6,7,8,9,10)
student_population <- c(2,6,8,8,12,16,20,20,22,26)
quartely_sales <- c(58,105,88,118,117,137,157,169,149,202)

DF <- data.frame(restaurant, student_population, quartely_sales)

regressor <- lm(quartely_sales ~ student_population, data = DF)
res <- summary(regressor)

table <- anova(regressor)

SSE <- table$`Sum Sq`[2]   # Sum of Squares due to Error

cat("Value of SSE is",SSE)

SSR <- table$`Sum Sq`[1]   # Sum of Squares due to Regression

cat("Value of SSR is",SSR)

SST <- SSE + SSR   # Total Sum of Squares

cat("Value of SST is",SST)

r_sq <- res$r.squared   # Coefficient of Determination

corrcoeff <- sqrt(r_sq)   # Correlation Coefficient

cat("Value of Coefficient of Determination is",r_sq)
cat("Value of correlation Coefficient is",corrcoeff)
```

R code Exa 14.3a Test of Significance

```
                                              # Page no. : 485 - 489

# Test of Significance

restaurant <- c(1,2,3,4,5,6,7,8,9,10)
student_population <- c(2,6,8,8,12,16,20,20,22,26)
quartely_sales <- c(58,105,88,118,117,137,157,169,149,202)

DF <- data.frame(restaurant, student_population, quartely_sales)

regressor <- lm(quartely_sales ~ student_population, data = DF)
res <- summary(regressor)

standard_error_MSE <- res$sigma

cat("Value of square root of MSE is",standard_error_MSE)

b1 <- res$coefficients[2]

tval <- res$coefficients[6]


# T Test

# Two Tail Test

# P-value Approach
```

```r

pval <- round(res$coefficients[8],3)

if(pval >= 0.01)
{
  cat("Since pval",pval,"is greater than or equal to
      0.01 therefore we cannot reject the Null
      Hypothesis")
} else{
  cat("Since pval",pval,"is less than 0.01 therefore
      we can reject the Null Hypothesis")

}

# F Test

test <- anova(regressor)

fval <- test$`F value`

pval <- round(test$`Pr(>F)`[1],3)   # P value is
    extremely small ie negligible to 0

if(pval >= 0.01)
{
  cat("Since pval",pval,"is greater than or equal to
      0.01 therefore we cannot reject the Null
      Hypothesis")
} else{
  cat("Since pval",pval,"is less than 0.01 therefore
      we can reject the Null Hypothesis")

}

# Confidence Interval

confidence <- confint(regressor, "student_population
    ", level = 0.99)
```

```
IE1 <- confidence[1]
IE2 <- confidence[2]

cat("The 99 %  confidence interval is ",IE1, "to",
   IE2)
```

R code Exa 14.4a Using the Estimated Regression Equation for Estimation and Prediction

```
                                        # Page no. : 595 -
                                           597

# Using the Estimated Regression Equation for
   Estimation and Prediction

restaurant <- c(1,2,3,4,5,6,7,8,9,10)
student_population <- c(2,6,8,8,12,16,20,20,22,26)
quartely_sales <- c
   (58,105,88,118,117,137,157,169,149,202)

DF <- data.frame(restaurant, student_population,
   quartely_sales)

regressor <- lm(quartely_sales ~ student_population,
    data = DF)
res <- summary(regressor)

pred <- predict(regressor, data.frame(student_
   population=10), interval="confidence")

PE <- pred[1]
IE1 <- pred[2]
IE2 <- pred[3]

cat("Point estimate is ", PE)
```

```
cat("Confidence Interval is ",IE1, "to", IE2)

pred2 <- predict(regressor, data.frame(student_
   population=10), interval="predict")

IE1 <- pred2[2]
IE2 <- pred2[3]

cat("Prediction Confidence Interval is ",IE1, "to",
   IE2)
```

R code Exa 14.5a Residual Analysis Validating Model Assumptions

```
                                    # Page no. : 605
                                       - 609

# Residual Analysis : Validating Model Assumptions

x <- c(2,6,8,8,12,16,20,20,22,26)   # Student
   Population
y <- c(58,105,88,118,117,137,157,169,149,202)   #
   Sales

estimated_sales <- 60 + (5 * x)   # Regression
   Equation  = 60 + 5 x

residuals <- y - estimated_sales

DF <- data.frame(x, y, estimated_sales, residuals)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library
```

```

library(ggplot2)

ggplot(DF,aes(x, residuals)) + geom_point() + geom_hline(yintercept = 0, linetype=2) +
labs(title = "Residual Plot", x = "X", y = "Residual")

ggplot(DF,aes(estimated_sales, residuals)) + geom_point() + geom_hline(yintercept = 0,
        linetype=2) + labs(title = "Plot between Estimated Sales and Residuals", x =
                                "Estimated Sales", y = "Residual")
```

R code Exa 14.5b Standardized Residuals

```
                                   # Page no. : 610 - 612

# Standardized Residuals

i <- c(1,2,3,4,5,6,7,8,9,10)
x <- c(2,6,8,8,12,16,20,20,22,26)   # Student Population
n <- 10
s <- 13.829   # Standard error
y <- c(58,105,88,118,117,137,157,169,149,202)   # Sales

estimated_sales <- 60 + (5 * x)   # Regression Equation = 60 + 5 x

x_deviation <- x - mean(x)

```

```
x_deviation_sq <- (x_deviation)**2

z <- round(x_deviation_sq / sum(x_deviation_sq),4)

h <- round((1 / n) + z, 4)

s_i <- round(s * sqrt(1 -h), 4)   # Standard Deviation for Residual i

residuals <- y - estimated_sales

standard_residuals <- round((residuals) / (s_i), 4)

DF <- data.frame(x, y, estimated_sales, x_deviation, x_deviation_sq, z, h, s_i, residuals, standard_residuals)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(x, standard_residuals)) + geom_point() + geom_hline(yintercept = 0, linetype=2) +
   labs(title = "Scatter Plot between  x and Standard Residuals", x = "X",
        y = "Standard Residuals")

normal_scores <- round(qqnorm(1:10)$x, 2)
standard_residuals <- sort(standard_residuals, decreasing = F)

table <- data.frame(normal_scores, standard_residuals)

```

```
ggplot(table,aes(normal_scores, standard_residuals))
    + geom_point() +
  geom_smooth(method = "lm", se = F) +
labs(title = "Scatter Plot between Normal Scores and
    Standard Residuals", x = "Normal Scores",
     y = "Standard Residuals")
```

R code Exa 14.6a Detecting Outliers

```
                                  # Page no. : 615

# Detecting Outliers

x <- c(1,1,2,3,3,3,4,4,5,6)
y <- c(45,55,50,75,40,45,30,35,25,15)

DF <- data.frame(x,y)

ggplot(DF,aes(x, y)) + geom_point()+ geom_smooth(
   method = "lm", se = F) +
labs(title = "Scatter Plot between  x and Y Showing
   Outlier", x = "X", y = "Y")
```

R code Exa 14.6b Detecting Influential Observations

```
                                     # Page no. :
                                        617 - 618

# Detecting Influential Observations

x <- c(10, 10, 15, 20, 20, 25, 70)
y <- c(135, 130, 120, 115, 120, 110, 100)

```

```r
DF <- data.frame(x, y)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF, aes(x, y)) + geom_point() + labs(title =
    "Scatter Plot between x and y", x = "X",
                                              y = "Y")

point <- x[7]  # From Scatter Plot

h <- (1 / nrow(DF)) + (((point - mean(DF$x))**2) / (
    sum((DF$x - mean(DF$x))**2)))

cat("Leverage at point 7 is", h)
```

Chapter 15

Multiple Regression

R code Exa 15.1a Least Squares Method

```
                                        # Page no. :
                                           646 - 647

# Least Squares Method

driving_assignment <- c(1,2,3,4,5,6,7,8,9,10)
x <- c(100,50,100,100,50,80,75,65,90,90)   # Miles
   Traveled
y <- c(9.3,4.8,8.9,6.5,4.2,6.2,7.4,6.0,7.6,6.1)   #
   Travel Time (hours)

DF <- data.frame(driving_assignment, x ,y)

# Install Library if not installed

# install.packages("ggpolt2")

# Import Library

library(ggplot2)

```

```
ggplot(DF,aes(x, y)) + geom_point() +labs(title = "Scatter Plot between Miles Traveled
  and Travel Time", x = "Miles Traveled", y = "Travel Time in Hours")

                                        # Page no. : 647 - 648

model <- lm(y ~ x, data = DF)
summ <- summary(model)
summ

b0 <- model$coefficients[1]
b1 <- model$coefficients[2]

cat("Linear Regression Equation is y_cap =",b0,"+",b1,"x1")

# F Test

fval <- summ$fstatistic[1]

# Upper Tail Test

# P-value Approach

alpha <- 0.05

pval <- summ$coefficients[8]  # P-value

if(pval >= alpha)
{
  cat("Since pval",pval,"is greater than or equal to 0.01 therefore we cannot reject the Null Hypothesis")
} else{
  cat("Since pval",pval,"is less than 0.05 therefore we can reject the Null Hypothesis")
```

```
}
```

R code Exa 15.2a Two Independent Variables

```
# Page no. : 648 - 649

# Two Independent Variables

driving_assignment <- c(1,2,3,4,5,6,7,8,9,10)
x1 <- c(100,50,100,100,50,80,75,65,90,90)   # Miles Traveled
x2 <- c(4,3,4,2,2,2,3,4,3,2)   # Number of Deliveries
y <- c(9.3,4.8,8.9,6.5,4.2,6.2,7.4,6.0,7.6,6.1)   # Travel Time (hours)

DF <- data.frame(driving_assignment, x1, x2, y)

model <- lm(y ~ x1 + x2, data = DF)
summary(model)

b0 <- model$coefficients[1]
b1 <- model$coefficients[2]
b2 <- model$coefficients[3]

cat("Multiple Regression Equation is y_cap =",b0,"+",b1,"x1 +",b2,"x2")
```

R code Exa 15.3a Multiple Coefficient of Determination

```
                                              # Page no. : 654 - 655

# Multiple Coefficient of Determination

driving_assignment <- c(1,2,3,4,5,6,7,8,9,10)
x1 <- c(100,50,100,100,50,80,75,65,90,90)   # Miles Traveled
x2 <- c(4,3,4,2,2,2,3,4,3,2)   # Number of Deliveries
y <- c(9.3,4.8,8.9,6.5,4.2,6.2,7.4,6.0,7.6,6.1)   # Travel Time (hours)

DF <- data.frame(driving_assignment, x1, x2, y)

model <- lm(y ~ x1 + x2, data = DF)
summ <- summary(model)

mean_y <- mean(DF$y)   # Mean of Travel time

predicted_travel_time <- round(predict(model), 2)

DF <- cbind(DF, predicted_travel_time)

SSR <- sum((DF$predicted_travel_time - mean_y)**2)    # Sum of Squares due to Regression

cat("Value of SSR is",SSR)

SSE <- sum((DF$y - DF$predicted_travel_time)**2)   # Sum of Squares due to Error

cat("Value of SSE", SSE)

SST <- SSR + SSE   # Total Sum of Squares

cat("Value of SST", SST)

```

```
r_sq <- summ$r.squared  # Multiple Coefficient of Determination

cat("Value of Multiple Coefficient of Determination is", r_sq)

adj_r_sq <- summ$adj.r.squared  # Adjusted Multiple Coefficient of Determination

cat("Value of Adjusted Multiple Coefficient of Determination is", adj_r_sq)
```

R code Exa 15.4a Testing of significance

```
                                        # Page no. : 660 - 662

# Testing for Significance

driving_assignment <- c(1,2,3,4,5,6,7,8,9,10)
x1 <- c(100,50,100,100,50,80,75,65,90,90)  # Miles Traveled
x2 <- c(4,3,4,2,2,2,3,4,3,2)  # Number of Deliveries
y <- c(9.3,4.8,8.9,6.5,4.2,6.2,7.4,6.0,7.6,6.1)  # Travel Time (hours)

DF <- data.frame(driving_assignment, x1, x2, y)

model <- lm(y ~ x1 + x2, data = DF)
summ <- summary(model)

SSR <- 21.6252  # Sum of Squares due to Regression

SSE <- 2.2952  # Sum of Squares due to Error
```

```
n <- nrow(DF)   # Total Observations

p <- 2   # Number of Independent Variables

MSR <- SSR / p   # Mean Square due to Regression

cat("Value of MSR is",MSR)

MSE <- SSE / (n - p - 1)   # Mean Square due to
   Error

# F Test

fval <- summ$fstatistic[1]

alpha <- 0.01

# Upper Tail Test

# P-value Approach

pval <- summ$coefficients[11]

if(pval >= alpha)
{
  cat("Since pval",pval,"is greater than or equal to
     0.01 therefore we cannot reject the Null
    Hypothesis")
} else{
  cat("Since pval",pval,"is less than 0.01 therefore
     we can reject the Null Hypothesis")

}

s <- summ$sigma   # Standard Error of the Estimate

cat("Value of Standard Error of the Estimate is",s)
```

```
coeff <- as.data.frame(summ$coefficients)

b1 <- coeff$Estimate[2]
b2 <- coeff$Estimate[3]

s1 <- coeff$`Std. Error`[2]   # Standard Error of
   the x1 (Miles Traveled)
s2 <- coeff$`Std. Error`[3]   # Standard Error of
   the x2 (Number of Deliveries)

df <- n - p - 1   # Degrees of Freedom

# T Test

tval1 <- round(b1 / s1, 3)
tval2 <- round(b2 / s2, 3)

# Upper Tail Test

# P-value Approach

pval1 <- round(pt(tval1, df, lower.tail = F),5)
pval2 <- round(pt(tval2, df, lower.tail = F),5)


if(pval1 >= alpha && pval2 >= alpha)
{
  cat("Since pval1",pval1,"and pval2",pval2,"is
     greater than or equal to 0.01 therefore we
     cannot reject the
      Null Hypothesis")
} else{
  cat("Since pval1",pval1,"and pval2",pval2,"is less
      than 0.01 therefore we can reject the Null
     Hypothesis")

}
```

R code Exa 15.5a Categorical Independent Variables

```r
# Page no. : 668 - 671

# Categorical Independent Variables

service <- c(1:10)
month <- c(2,6,8,3,2,7,9,8,4,6)
repair <- c("electrical","mechanical","electrical","mechanical","electrical","electrical",
            "mechanical","mechanical","electrical","electrical")
time <- c(2.9, 3.0, 4.8, 1.8, 2.9, 4.9, 4.2, 4.8, 4.4, 4.5)

DF <- data.frame(service, month, repair, time)

# With one Independent variable

regressor <- lm(time ~ month, data = DF)
res <- summary(regressor)

res

b0 <- res$coefficients[1]
b1 <- res$coefficients[2]

cat("Equation is y = ", b0 ,"+",b1,"x1")

# With Categorical Variable

DF$repair <- factor(DF$repair,
                    levels = c('electrical', '
```

```
                              mechanical'),
                         labels = c(1, 0))

regressor <- lm(time ~ month + repair, data = DF)
res <- summary(regressor)

res

b0 <- res$coefficients[1]
b1 <- res$coefficients[2]
b2 <- res$coefficients[3]

# For Electrical Repair

bterm <- b0 + b2 * 1

cat("Equation for electrical is  y = ", bterm ,"+",
    b1,"x1" )

# For Mechanical Repair

bterm2 <- b0 + b2 * 0

cat("Equation for mechanical is  y = ", bterm2 ,"+",
    b1,"x1" )
```

R code Exa 15.6a Residual Analysis

```
                                              # Page no. :
                                                 676 - 679

# Residual Analysis

miles <- c(100, 50, 100, 100, 50, 80, 75, 65, 90,
    90)
```

```r
deliveries <- c(4, 3, 4, 2, 2, 2, 3, 4, 3, 2)
time <- c(9.3, 4.8, 8.9, 6.5, 4.2, 6.2, 7.4, 6.0, 7.6, 6.1)

DF <- data.frame(miles, deliveries, time)

regressor <- lm(time ~ miles + deliveries, data = DF)
res <- summary(regressor)

predict <- predict(regressor)
residuals <- DF$time - predict
std_residuals <- rstandard(regressor)

DF <- cbind(DF, predict, residuals, std_residuals)

View(DF)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(predict, std_residuals)) + geom_point() + geom_hline(yintercept = 0,
      linetype=2) + labs(title = "Plot between predicted values and standardizes residuals", x =
                                            "Prediction", y = "Standardization Residual")


leverage <- hatvalues(regressor)
```

```
cook_dist <- cooks.distance(regressor)

DF <- cbind(DF, leverage, cook_dist)

View(DF)
```

R code Exa 15.6b Influential Observations

```
                                          # Page no. :
                                             679 - 680

# Influential Observations

x <- c(1, 1, 2, 3, 4, 4, 5, 15)
y <- c(18, 21, 22, 21, 23, 24, 26, 39)

DF <- data.frame(x, y)

regressor <- lm(y ~ x, data = DF)
res <- summary(regressor)

lev <- hatvalues(regressor)
DF <- cbind(DF, lev)

View(DF)


# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

```

```
ggplot(DF,aes(x, y)) + geom_point() + geom_smooth(
    method = "lm", se = F) +
  labs(title = "Plot between x and y", x = "X", y =
    "Y")


cat("Equation is y = ",res$coefficients[1],"+",res$
    coefficients[2],"x1")


# Removing Influential Observation

x <- c(1, 1, 2, 3, 4, 4, 5)
y <- c(18, 21, 22, 21, 23, 24, 26)

DF <- data.frame(x, y)

regressor <- lm(y ~ x, data = DF)
res <- summary(regressor)

cat("Equation is y = ",res$coefficients[1],"+",res$
    coefficients[2],"x1")
```

R code Exa 15.7a Logistic Regression

```
                                          # Page no. :
                                            684 - 686

# Logistic Regression

customer <- c(1,2,3,4,5,6,7,8,9,10)
spending <- c
    (2.291,3.215,2.135,3.924,2.528,2.473,2.384,7.076,1.182,3.345)

card <- c(1,1,1,0,1,0,0,0,1,0)
```

```
coupon <- c(0,0,0,0,0,1,0,0,1,0)

DF <- data.frame(customer, spending, card, coupon)

regressor <- glm(coupon ~ spending + card, data = DF
  )
summary(regressor)

# Book answer is different
```

Chapter 16

Regression Analysis Model Building

R code Exa 16.1a General Linear Model

```
                                        # Page no. : 714
                                            - 717

# General Linear Model

x <- c(41,106,76,10,22,12,85,111,40,51,9,12,6,56,19)
      # Months Employed
y <- c
   (275,296,317,376,162,150,367,308,189,235,83,112,67,325,189)
      # Scales Sold

DF <- data.frame(x,y)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

```

```r
library(ggplot2)

ggplot(DF, aes(x,y)) + geom_point() + labs(title = "Scatter Plot between Months
  Employed and Scales sold", x = "Months Employed", y = "Scales Sold")

model <- lm(y ~ x, data = DF)
summary(model)

b1 <- model$coefficients[1]
b2 <- model$coefficients[2]

cat("Regression Equation is sales =",b1,"+",b2," months")
# Book Answer is sales = 111 + 2.38 months   (Data is wrongly used in book)

predicted_scales_sold <- predict(model)

DF <- cbind(DF, predicted_scales_sold)

# Standard Residuals

standard_residuals <- rstandard(model)

DF <- cbind(DF, standard_residuals)

ggplot(DF,aes(predicted_scales_sold, standard_residuals)) + geom_point() +
  geom_hline(yintercept = 0, linetype=2) +
labs(title = "Scatter Plot between Predicted scales sold and Standard Residuals",
    x = "Predicted scales sold", y = "Standard Residuals")

                                              # Page no. : 716 - 717
```

```
x2 <- x**2

DF2 <- data.frame(x,x2,y)

model2 <- lm(y ~ x + x2, data = DF2)
summary(model2)

b1 <- model2$coefficients[1]
b2 <- model2$coefficients[2]
b3 <- model2$coefficients[3]

cat("Regression Equation is sales =",b1,"+",b2,"months","+",b3,"monthsq")
# Book Answer is sales = 45.3 + 6.34 months - 0.0345 monthsq (Data is wrongly used in book)

predicted_scales_sold2 <- predict(model2)

DF2 <- cbind(DF2, predicted_scales_sold2)

# Standard Residuals

standard_residuals2 <- rstandard(model2)

DF2 <- cbind(DF2, standard_residuals2)

ggplot(DF2,aes(predicted_scales_sold2, standard_residuals2)) + geom_point() +
  geom_hline(yintercept = 0, linetype=2) +
  labs(title = "Scatter Plot between Predicted scales sold and Standard Residuals",
       x = "Predicted scales sold", y = "Standard Residuals")
```

R code Exa 16.1b Interaction

```r
                                                        # Page no. : 718 - 721

# Interaction

price <- c(2.00, 2.50, 3.00, 2.00, 2.50, 3.00, 2.00, 2.50, 3.00, 2.00, 2.50, 3.00,
           2.00, 2.50, 3.00, 2.00, 2.50, 3.00, 2.00, 2.50, 3.00, 2.00, 2.50, 3.00)
advertising <- c(50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 100, 100, 100, 100, 100,
                 100, 100, 100, 100, 100, 100, 100)
sales <- c(478, 373, 335, 473, 358, 329, 456, 360, 322, 437, 365, 342, 810, 653, 345, 832, 641,
           372, 800, 620, 390, 790, 670, 393)

DF <- data.frame(price, advertising, sales)

# Install Library if not install

# install.packages("dplyr")

# Import Library

library(dplyr)

DF %>% group_by(price,advertising) %>% summarize(Average=mean(sales))

model <- lm(sales ~ price + advertising + (price * advertising), data = DF)
res <- summary(model)

b0 <- res$coefficients[1]
b1 <- res$coefficients[2]
b2 <- res$coefficients[3]
```

```
b3 <- res$coefficients[4]

cat("Equation is sales = ", b0, "+", b1, "Price +",
    b2, "AdvExp +", b3, "PriAdv")
```

R code Exa 16.1c Transformations Involving the Dependent Variables

```
                                        # Page no. : 721
                                           - 724

# Transformations Involving the Dependent Variables

x <- c
   (2289,2113,2180,2448,2026,2702,2657,2106,3226,3213,3607,2888)
      # Weight
y <- c
   (28.7,29.2,34.2,27.9,33.3,26.4,23.9,30.5,18.1,19.5,14.3,20.9)
      # Miles Per Gallon

DF <- data.frame(x,y)

# Install Library if not install

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF, aes(x,y)) + geom_point() + labs(title = "
   Scatter Plot between Weight and Miles
                                             Per
                                                Gallon
                                                ", x =
                                                 "
```

```
Weight", y = "Miles Per Gallon")

model <- lm(y ~ x, data = DF)
summary(model)

b1 <- model$coefficients[1]
b2 <- model$coefficients[2]

cat("Regression Equation is sales =",b1,"+",b2," months")

# Prediction

predicted_miles_per_gallon <- predict(model)

DF <- cbind(DF, predicted_miles_per_gallon)

# Standard Residuals

standard_residuals <- rstandard(model)

DF <- cbind(DF, standard_residuals)


ggplot(DF,aes(predicted_miles_per_gallon, standard_residuals)) + geom_point() +
  geom_hline(yintercept = 0, linetype=2) +
  labs(title = "Scatter Plot between Predicted Miles per Gallon and Standard Residuals",
       x = "Predicted Miles per Gallon", y = "Standard Residuals")

```

```
# Log Transformation

#log_x <- log(x)
log_y <- log(y)

DF2 <- data.frame(x, log_y)

options(scipen = 999)   # To display not in
    scientific notation
model <- lm(log_y ~ x, data = DF2)
summary(model)

b1 <- model$coefficients[1]
b2 <- model$coefficients[2]

cat("Regression Equation is sales =",b1,"+",b2,"
    months")

# Prediction

predicted_miles_per_gallon <- predict(model)

DF2 <- cbind(DF2, predicted_miles_per_gallon)

# Standard Residuals

standard_residuals <- rstandard(model)

DF2 <- cbind(DF2, standard_residuals)


ggplot(DF2,aes(predicted_miles_per_gallon, standard_
    residuals)) + geom_point() +
  geom_hline(yintercept = 0, linetype=2) + labs(
      title = "Scatter Plot between Log
Transformation of Predicted Miles per Gallon and
    Standard Residuals",
        x = "Predicted Miles per Gallon", y = "
```

```
Standard Residuals")
```

R code Exa 16.2a Analysis of a Large Problem

```
# Page no. : 736 - 738

# Analysis of a Large Problem

Sales <- c(3669.88, 3473.95, 2295.10, 4675.56, 6125.96, 2134.94, 5031.66, 3367.45, 6519.45,
           4876.37, 2468.27, 2533.31, 2408.11, 2337.38, 4586.95, 2729.24, 3289.40, 2800.78,
           3264.20, 3453.62, 1741.45, 2035.75, 1578.00, 4167.44, 2799.97)
Time <- c(43.10, 108.13, 13.82, 186.18, 161.79, 8.94, 365.04, 220.32, 127.64, 105.69, 57.72, 23.58,
          13.82, 13.82, 86.99, 165.85, 116.26, 42.28, 52.84, 165.04, 10.57, 13.82, 8.13, 58.44,
          21.14)
Poten <- c(74065.1, 58117.3, 21118.5, 68521.3, 57805.1, 37806.9, 50935.3, 35602.1, 46176.8, 42053.2,
           36829.7, 33612.7, 21412.8, 20416.9, 36272.0, 23093.3, 26878.6, 39572.0, 51866.1, 58749.8,
           23990.8, 25694.9, 23736.3, 34314.3, 22809.5)
AdvExp <- c(4582.9, 5539.8, 2950.4, 2243.1, 7747.1, 402.4, 3140.6, 2086.2, 8846.2, 5673.1, 2761.8,
            1991.8, 1971.5, 1737.4, 10694.2, 8618.6, 7747.9, 4565.8, 6022.7, 3721.1,
```

```
                861.0, 3571.5,
              2845.5, 5060.1, 3552.0)
Share <- c(2.51, 5.51, 10.91, 8.27, 9.15, 0.15,
    8.54, 7.07, 12.54, 8.85, 5.38, 5.43, 8.48, 7.80,
             10.34, 5.15, 6.64, 5.45, 6.31, 6.35,
               7.37, 8.39, 5.15, 12.88, 9.14)
Change <- c(0.34, 0.15, -0.72, 0.17, 0.50, 0.15,
    0.55, -0.49, 1.24, 0.31, 0.37, -0.65, 0.64, 1.01,
              0.11, 0.04, 0.68, 0.66, -0.10, -0.03,
                -1.63, -0.43, 0.04, 0.22, -0.74)
Accounts <- c(74.86, 107.32, 96.75, 195.12, 180.44,
    104.88, 256.10, 126.83, 203.25, 119.51, 116.26,
               142.28, 89.43, 84.55, 119.51,80.49,
                 136.58, 78.86, 136.58, 138.21,
                 75.61, 102.44,
               76.42, 136.58, 88.62)
Work <- c(15.05, 19.97, 17.34, 13.40, 17.64, 16.22,
    18.80, 19.86, 17.42, 21.41, 16.32, 14.51,
            19.35, 20.02, 15.26, 15.87, 7.81, 16.00,
              17.44, 17.98, 20.99, 21.66, 21.46,
              24.78,
            24.96)
Rating <- c(4.9, 5.1, 2.9, 3.4, 4.6, 4.5, 4.6, 2.3,
    4.9, 2.8, 3.1, 4.2, 4.3, 4.2, 5.5, 3.6, 3.4,
              4.2, 3.6, 3.1, 1.6, 3.4, 2.7, 2.8, 3.9)

DF <- data.frame(Sales, Time, Poten, AdvExp, Share,
    Change, Accounts, Work, Rating)
View(DF)

library(corrplot)

matrix <- round(cor(DF), 3)
matrix

regressor <- lm(Sales ~ Time + Poten + AdvExp +
    Share + Change + Accounts + Work + Rating,
                  data = DF)
```

```
summary(regressor)

cat("Equation is Sales = ",regressor$coefficients
    [1],"+",regressor$coefficients[2],"Time +",
    regressor$coefficients[3],"Poten +",regressor$
      coefficients[4],"AdvExp +",regressor$
      coefficients[5],
    "Share +",regressor$coefficients[6],"Change +",
      regressor$coefficients[7],"Accounts +",
    regressor$coefficients[8],"Work +",regressor$
      coefficients[8],"Rating")

# Book Answer is different

regressor2 <- lm(Sales ~ Poten + AdvExp + Share,
    data = DF)
summary(regressor2)

cat("Equation is Sales = ",regressor2$coefficients
    [1],"+",regressor2$coefficients[3],"Poten +",
    regressor2$coefficients[4],"AdvExp +",regressor2
      $coefficients[5],
    "Share")

# Book Answer is different
```

R code Exa 16.3a Multiple Regression Approach to Experimental Design

```
                                  # Page no. : 745 -
                                     747

# Multiple Regression Approach to Experimental
  Design

A <- c(58,64,55,66,67)
```

```
B <- c(58,69,71,64,68)
C <- c(48,57,59,47,49)

DF <- data.frame(A,B,C)

newA <- c(1,1,1,1,1,0,0,0,0,0,0,0,0,0,0)
newB <- c(0,0,0,0,0,1,1,1,1,1,0,0,0,0,0)
y <- c(58,64,55,66,67,58,69,71,64,68,48,57,59,47,49)

DF2 <- data.frame(newA, newB, y)

regressor <- lm(y ~ newA + newB, data = DF2)
summary(regressor)

cat("Equation is y = ",regressor$coefficients[1],"+"
    ,regressor$coefficients[2],"A +",
     regressor$coefficients[3],"B")
```

Chapter 17

Index Numbers

R code Exa 17.1a Price Relatives

```
                                        # Page no. : 765

# Price Relatives

year <- c
   (1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002

price_per_gallon <- c(1.30,
   1.10,1.09,1.07,1.08,1.11,1.22,1.20,1.03,1.14,1.48,1.42,1.34,1.56,

DF <- data.frame(year,price_per_gallon)

base_year <- 1990
base_year_price <- DF$price_per_gallon[DF$year ==
   base_year]

price_relative <- round(((DF$price_per_gallon) / (
   base_year_price)) * 100, 2)

```

```
DF <- cbind(DF, price_relative)

View(DF)
```

R code Exa 17.2a Aggregate Price Indexes

```
                                        # Page no. :
                                            766

# Aggregate Price Indexes

item <- c("Gallon of gasoline","Quart of oil","Tire"
    ,"Insurance policy")
year_1990 <- c(1.30,2.10,130.00,820.00)
year_2008 <- c(3.25,8.00,140.00,1030.00)
quantity <- c(1000,15,2,1)

DF <- data.frame(item,year_1990,year_2008,quantity)

base_year <- 1990

sum_of_1990_items <- sum(DF$year_1990)

sum_of_2008_items <- sum(DF$year_2008)

aggregate_index_2008 <- (sum_of_2008_items / sum_of_
    1990_items) * 100

cat("The unweighted aggregate index for year 2008 is
    ",aggregate_index_2008)


sum_of_1990_items <- sum(DF$year_1990 * DF$quantity)
        # Weighted Sum

```

```
sum_of_2008_items <- sum(DF$year_2008 * DF$quantity)
      # Weighted Sum

aggregate_index_2008 <- (sum_of_2008_items / sum_of_
   1990_items) * 100

cat("The weighted aggregate index for year 2008 is",
   aggregate_index_2008)
```

R code Exa 17.3a Computing an Aggregate Price Index from Price Relatives

```
                                             # Page no. : 769
                                                - 770

# Computing an Aggregate Price Index from Price
   Relatives

item <- c("Gallon of gasoline","Quart of oil","Tire"
   ,"Insurance policy")
year_1990 <- c(1.30,2.10,130.00,820.00)
year_2008 <- c(3.25,8.00,140.00,1030.00)
quantity <- c(1000,15,2,1)

DF <- data.frame(item,year_1990,year_2008,quantity)

base_year <- 1990

price_relative <- (DF$year_2008 / DF$year_1990) *
   100

weight <- DF$year_1990 * DF$quantity

weight_price_relative <- price_relative * weight

```

```
DF <- cbind(DF, price_relative, weight, weight_price_relative)

aggregate_2008 <- sum(DF$weight_price_relative) / sum(DF$weight)

cat("Aggregate Price Index for year 2008 is", aggregate_2008)
```

R code Exa 17.4a Deflating a Series by Price Indexes

```
                                        # Page no. : 774

# Deflating a Series by Price Indexes

year <- c(2004,2005,2006,2007,2008)
hourly_wage <- c(15.69,16.12,16.76,17.45,18.07)
CPI <- c(188.9,195.3,201.6,207.3,215.3)

DF <- data.frame(year, hourly_wage, CPI)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF, aes(year, hourly_wage)) + geom_line() + geom_point() +
 labs(title = "Year V/S Hourly Wage Graph", x = "Year", y = "Hourly Wage")

```

```
deflated_hourly_wage <- (DF$hourly_wage / DF$CPI) * 100

DF <- cbind(DF, deflated_hourly_wage)

ggplot(DF, aes(year, deflated_hourly_wage)) + geom_line() + geom_point() +
  ylim(c(7.75,9.0)) +
  labs(title = "Year V/S Real Hourly Wage Graph", x = "Year", y = "Real Hourly Wage")
```

Chapter 18

Time Series Analysis and Forecasting

R code Exa 18.1a Time Series Patterns

```
                                            # Page no. : 786
                                                - 787

# Time Series Patterns

week <- c(1,2,3,4,5,6,7,8,9,10,11,12)
sales <- c(17,21,19,23,18,16,20,18,22,20,15,22)

DF <- data.frame(week,sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(week,sales)) + geom_line() + geom_
```

```
point() + ylim(c(0,25)) + labs(title = "Week V/S Sales Time Series Plot", x = "Weeks", y = "Sales")
```

R code Exa 18.1b Time Series Patterns Eg2

```
# Page no. : 787 - 788

# Time Series Patterns Eg-2

week <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22)
sales <- c(17,21,19,23,18,16,20,18,22,20,15,22,31,34,31,33,28,32,30,29,34,3

DF <- data.frame(week,sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(week,sales)) + geom_line() + geom_point() + ylim(c(0,40)) +
labs(title = "Week V/S Sales Time Series Plot", x = "Weeks", y = "Sales")
```

R code Exa 18.1c Time Series Patterns Eg3

```
                                    # Page no. :
                                      788 - 789

# Time Series Patterns Eg-3

year <- c(1,2,3,4,5,6,7,8,9,10)
sales <- c
   (21.6,22.9,25.5,21.9,23.9,27.5,31.5,29.7,28.6,31.4)

DF <- data.frame(year,sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year,sales)) + geom_line() + geom_
   point() + ylim(c(20,34)) +
 labs(title = "Years V/S Sales Time Series Plot", x
    = "Years", y = "Sales")
```

R code Exa 18.1d Time Series Patterns Eg4

```
                                  # Page no. : 789
                                     - 790

```

```
# Time Series Patterns Eg-4

year <- c(1,2,3,4,5,6,7,8,9,10)
revenue <- c
   (23.1,21.3,27.4,34.6,33.8,43.2,59.5,64.4,74.2,99.3)

DF <- data.frame(year,revenue)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year,revenue)) + geom_line() + geom_
   point() + ylim(c(0,120)) +
 labs(title = "Years V/S Revenue Time Series Plot",
    x = "Years", y = "Revenue")
```

R code Exa 18.1e Time Series Patterns Eg5

```
                                   # Page no. : 790
                                      - 791

# Time Series Patterns Eg-5

year_quart <- c("Y1 Q1", "Y1 Q2", "Y1 Q3", "Y1 Q4",
   "Y2 Q1", "Y2 Q2", "Y2 Q3", "Y2 Q4", "Y3 Q1",
                "Y3 Q2", "Y3 Q3", "Y3 Q4", "Y4 Q1",
                   "Y4 Q2", "Y4 Q3", "Y4 Q4", "Y5 Q1
                   ", "Y5 Q2",
                "Y5 Q3", "Y5 Q4")  # Years and
```

```
Quarters
sales <- c(125, 153, 106, 88, 118, 161, 133, 102,
    138, 144, 113, 80, 109, 137, 125, 109, 130, 165,
            128, 96)

DF <- data.frame(year_quart, sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year_quart,sales, group = 1)) + geom_
    line() + geom_point() + ylim(c(0,180)) +
  labs(title = "Years/quart V/S Sales Time Series
      Plot", x = "Year/Quarter", y = "Sales")
```

R code Exa 18.1f Time Series Patterns Eg6

```
                                                # Page no. :
                                                    791 - 792

# Time Series Patterns Eg-6

year_quart <- c("Y1 Q1", "Y1 Q2", "Y1 Q3", "Y1 Q4",
    "Y2 Q1", "Y2 Q2", "Y2 Q3", "Y2 Q4", "Y3 Q1",
                  "Y3 Q2", "Y3 Q3", "Y3 Q4", "Y4 Q1",
                      "Y4 Q2", "Y4 Q3", "Y4 Q4")
                  # Years and Quarters
sales <- c(4.8, 4.1, 6.0, 6.5, 5.8, 5.2, 6.8, 7.4,
    6.0, 5.6, 7.5, 7.8, 6.3, 5.9, 8.0, 8.4)

```

```
DF <- data.frame(year_quart, sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year_quart,sales, group = 1)) + geom_
    line() + geom_point() + ylim(c(0.0,9.0)) +
  labs(title = "Years/quart V/S Sales Time Series
      Plot", x = "Year/Quarter", y = "Sales")
```

R code Exa 18.2a Forecast Accuracy

```
                                                # Page no. :
                                                   793 - 795

# Forecast Accuracy

week <- c(1,2,3,4,5,6,7,8,9,10,11,12)
sales <- c(17,21,19,23,18,16,20,18,22,20,15,22)
forcast <- c(NA,17,21,19,23,18,16,20,18,22,20,15)

DF <- data.frame(week,sales,forcast)

forcast_error <- DF$sales - DF$forcast

absolute_forecast_error <- abs(forcast_error)

square_absolute_forcast_error <- absolute_forecast_
    error**2

```

```r
percent_error <- round((forcast_error / DF$sales) *
   100, 2)

absolute_percent_error <- abs(percent_error)

DF <- cbind(DF,forcast_error,absolute_forecast_error
   ,square_absolute_forcast_error,
             percent_error,absolute_percent_error)

View(DF)

total_forcast_error <- sum(DF$forcast_error, na.rm =
    T)

total_absolute_forcast_error <- sum(DF$absolute_
   forecast_error, na.rm = T)

total_sq_abs_error <- sum(DF$square_absolute_forcast
   _error, na.rm = T)

total_percent_error <- sum(DF$percent_error, na.rm =
    T)

total_absolute_percent_error <- sum(DF$absolute_
   percent_error, na.rm = T)

# Native Value

MAE <- round(total_absolute_forcast_error / (nrow(DF
   )-1), 2)   # Not including 1st row
# Mean Absolute Error

MSE <- round(total_sq_abs_error / (nrow(DF)-1), 2)
     # Not including 1st row
# Mean Square Error

MAPE <- round(total_absolute_percent_error / (nrow(
   DF)-1), 2)   # Not including 1st row
```

```
# Mean Absolute Percent Error

cat("Value of MAE is",MAE)
cat("Value of MSE is",MSE)
cat("Value of MAPE is",MAPE)

# Average of Past Values

forcast2 <- c(NA,17.00,19.00,19.00,20.00,19.60,19.00,19.14,19.00,19.33,19.40,19.0

DF2 <- data.frame(week,sales,forcast2)

forcast_error2 <- DF2$sales - DF2$forcast2

absolute_forecast_error2 <- abs(forcast_error2)

square_absolute_forcast_error2 <- absolute_forecast_error2**2

percent_error2 <- round((forcast_error2 / DF2$sales) * 100, 2)

absolute_percent_error2 <- abs(percent_error2)

DF2 <- cbind(DF2,forcast_error2,absolute_forecast_error2,square_absolute_forcast_error2,
             percent_error2,absolute_percent_error2)

View(DF2)

total_forcast_error2 <- sum(DF2$forcast_error2, na.rm = T)

total_absolute_forcast_error2 <- sum(DF2$absolute_forecast_error2, na.rm = T)

```

```
total_sq_abs_error2 <- sum(DF2$square_absolute_forcast_error2, na.rm = T)

total_percent_error2 <- sum(DF2$percent_error2, na.rm = T)

total_absolute_percent_error2 <- sum(DF2$absolute_percent_error2, na.rm = T)

MAE2 <- round(total_absolute_forcast_error2 / (nrow(DF2)-1), 2)   # Not including 1st row
# Mean Absolute Error

MSE2 <- round(total_sq_abs_error2 / (nrow(DF2)-1), 2)   # Not including 1st row
# Mean Square Error

MAPE2 <- round(total_absolute_percent_error2 / (nrow(DF2)-1), 2)   # Not including 1st row
# Mean Absolute Percent Error

cat("Value of MAE is",MAE2)
cat("Value of MSE is",MSE2)
cat("Value of MAPE is",MAPE2)
```

R code Exa 18.3a Moving Averages

```
                                    # Page no. : 798 - 799

# Moving Averages

week <- c(1,2,3,4,5,6,7,8,9,10,11,12)
sales <- c(17,21,19,23,18,16,20,18,22,20,15,22)
forcast <- c(NA,NA,NA,19,21,20,19,18,18,20,20,19)
```

```r

DF <- data.frame(week,sales,forcast)

forcast_error <- DF$sales - DF$forcast

absolute_forecast_error <- abs(forcast_error)

square_absolute_forcast_error <- absolute_forecast_error**2

percent_error <- round((forcast_error / DF$sales) * 100, 2)

absolute_percent_error <- abs(percent_error)

DF <- cbind(DF,forcast_error,absolute_forecast_error,square_absolute_forcast_error,
            percent_error,absolute_percent_error)

View(DF)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(week)) + geom_line(aes(y = forcast), color = "red")+
  geom_line(aes(y = sales), color = "blue") + geom_point(aes(y = forcast)) +
  geom_point(aes(y = sales))+
  ylim(c(0,25)) +
  labs(title = "Week V/S Sales Time Series Plot", x = "Week", y = "Sales")
```

R code Exa 18.3b Exponential Smoothing

```
                                        # Page no. : 802 - 803

# Exponential Smoothing

week <- c(1,2,3,4,5,6,7,8,9,10,11,12)
sales <- c(17,21,19,23,18,16,20,18,22,20,15,22)
forcast <- c(NA,17.00,17.80, 18.04, 19.03, 18.83, 18.26, 18.61, 18.49, 19.19, 19.35, 18.48)
forcast_error <- sales - forcast
forcast_error_sq <- (forcast_error)**2

DF <- data.frame(week, sales, forcast, forcast_error, forcast_error_sq)

View(DF)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

 ggplot(DF,aes(week)) + geom_line(aes(y = forcast), color = "red")+
   geom_line(aes(y = sales), color = "blue") + geom_point(aes(y = forcast)) +
   geom_point(aes(y = sales))+
    ylim(c(0,25)) +
  labs(title = "Week V/S Sales Time Series Plot", x
```

```
= "Week", y = "Sales")
```

R code Exa 18.4a Trend Projection

```
                                        # Page no. :
                                          807 - 808

# Trend Projection

year <- c(1,2,3,4,5,6,7,8,9,10)
sales <- c
   (21.6,22.9,25.5,21.9,23.9,27.5,31.5,29.7,28.6,31.4)

DF <- data.frame(year,sales)

# Install Library if not installed

# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year,sales)) + geom_line() + geom_
   point() + ylim(c(20,34)) +
 labs(title = "Years V/S Sales Time Series Plot", x
    = "Years", y = "Sales")

ggplot(DF,aes(year,sales)) + geom_line() + geom_
   point() + ylim(c(20,34)) +
  geom_smooth(method = "lm", se = F) +
 labs(title = "Years V/S Sales Time Series and
    Linear Function Plot", x = "Years", y = "Sales")
```

R code Exa 18.4b Trend Projection Part2

```r
                                    # Page no. : 809 - 812

# Trend Projection Part-2

year <- c(1,2,3,4,5,6,7,8,9,10)
sales <- c(21.6,22.9,25.5,21.9,23.9,27.5,31.5,29.7,28.6,31.4)
forcast <- c(21.5,22.6,23.7,24.8,25.9,27.0,28.1,29.2,30.3,31.4)

DF <- data.frame(year,sales,forcast)

regressor <- lm(sales ~ year, data = DF)

cat("Linear Trend Equation is T =",regressor$coefficients[1],"+",regressor$coefficients[2],"t")

res <- anova(regressor)

cat("MSE is ",res$`Sum Sq`[2]/10)

cat("MSE is ",res$`Mean Sq`[2])

regressor
res

# Install Library if not installed

```

```
# install.packages("ggplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year,sales)) + geom_line() + geom_
   point() + ylim(c(20,34)) +
  geom_smooth(method = "lm", se = F) +
  labs(title = "Years V/S Sales Time Series Plot", x
     = "Years", y = "Sales")
```

R code Exa 18.4c Nonlinear Trend Regression

```
                                    # Page no. : 814 −
                                        816

# Nonlinear Trend Regression

year <- c(1,2,3,4,5,6,7,8,9,10)
revenue <- c
   (23.1,21.3,27.4,34.6,33.8,43.2,59.5,64.4,74.2,99.3)

year_sq <- year**2

DF <- data.frame(year,revenue,year_sq)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

```

```
ggplot(DF,aes(year,revenue)) + geom_line() + geom_point() + ylim(c(0,120)) +
  labs(title = "Years V/S Revenue Time Series Plot", x = "Year", y = "Revenue")

regressor <- lm(revenue ~ year + year_sq, data = DF)
summary(regressor)
anova(regressor)

ggplot(DF,aes(year,revenue)) + geom_line() + geom_point() + ylim(c(0,120)) + geom_smooth(method =
"lm", se = F) +
  labs(title = "Years V/S Revenue Time Series Plot", x = "Year", y = "Revenue")
```

R code Exa 18.5a Seasonality Without Trend

```
# Page no. : 821 - 822

# Seasonality Without Trend

```

```r
year_quart <- c("Y1 Q1", "Y1 Q2", "Y1 Q3", "Y1 Q4", "Y2 Q1", "Y2 Q2", "Y2 Q3", "Y2 Q4", "Y3 Q1",
                "Y3 Q2", "Y3 Q3", "Y3 Q4", "Y4 Q1", "Y4 Q2", "Y4 Q3", "Y4 Q4", "Y5 Q1", "Y5 Q2",
                "Y5 Q3", "Y5 Q4")  # Years and Quarters
sales <- c(125, 153, 106, 88, 118, 161, 133, 102, 138, 144, 113, 80, 109, 137, 125, 109, 130, 165,
           128, 96)

DF <- data.frame(year_quart, sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year_quart,sales, group = 1)) + geom_line() + geom_point() + ylim(c(0,180)) +
  labs(title = "Years/quart V/S Sales Time Series Plot", x = "Year/Quarter", y = "Sales")


q1 <- c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0)
q2 <- c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0)
q3 <- c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0)

DF <- cbind(DF,q1,q2,q3)

regressor <- lm(sales ~ q1 + q2 + q3, data = DF)
summary(regressor)

b0 <- regressor$coefficients[1]
b1 <- regressor$coefficients[2]
```

```r
b2 <- regressor$coefficients[3]
b3 <- regressor$coefficients[4]

cat("Equation is sales = ",b0,"+",b1,"Qtr1 +",b2,"Qtr2 +",b3,"Qtr3")
```

R code Exa 18.5b Seasonality andTrend

```r
                                            # Page no. : 823 - 824

# Seasonality and Trend

year_quart <- c("Y1 Q1", "Y1 Q2", "Y1 Q3", "Y1 Q4", "Y2 Q1", "Y2 Q2", "Y2 Q3", "Y2 Q4", "Y3 Q1",
                  "Y3 Q2", "Y3 Q3", "Y3 Q4", "Y4 Q1", "Y4 Q2", "Y4 Q3", "Y4 Q4")
# Years and Quarters
sales <- c(4.8, 4.1, 6.0, 6.5, 5.8, 5.2, 6.8, 7.4, 6.0, 5.6, 7.5, 7.8, 6.3, 5.9, 8.0, 8.4)

DF <- data.frame(year_quart, sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year_quart,sales, group = 1)) + geom_line() + geom_point() + ylim(c(0.0,9.0)) +
  labs(title = "Years/quart V/S Sales Time Series Plot", x = "Year/Quarter", y = "Sales")
```

```
q1 <- c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0)
q2 <- c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0)
q3 <- c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0)
period <- c(1:16)

DF <- cbind(DF,q1,q2,q3,period)

regressor <- lm(sales ~ q1 + q2 + q3 + period, data
    = DF)
summary(regressor)

b0 <- regressor$coefficients[1]
b1 <- regressor$coefficients[2]
b2 <- regressor$coefficients[3]
b3 <- regressor$coefficients[4]
b4 <- regressor$coefficients[5]

cat("Equation is sales = ",b0,"+",b1,"Qtr1 +",b2,"
    Qtr2 +",b3,"Qtr3 +",b4,"t")
```

R code Exa 18.6a Time Series Decomposition

```
                                        # Page no. :
                                          832 - 833

# Time Series Decomposition

year_quart <- c("Y1 Q1", "Y1 Q2", "Y1 Q3", "Y1 Q4",
    "Y2 Q1", "Y2 Q2", "Y2 Q3", "Y2 Q4", "Y3 Q1",
                 "Y3 Q2", "Y3 Q3", "Y3 Q4", "Y4 Q1",
                    "Y4 Q2", "Y4 Q3", "Y4 Q4")
# Years and Quarters
sales <- c(4.8, 4.1, 6.0, 6.5, 5.8, 5.2, 6.8, 7.4,
    6.0, 5.6, 7.5, 7.8, 6.3, 5.9, 8.0, 8.4)
```

```
moving_avg <- c(NA
    ,5.350,5.600,5.875,6.075,6.300,6.350,6.450,6.625,6.725,6.800,6.875
                 7.15,NA,NA)
centered_avg <- c(NA,NA
    ,5.475,5.738,5.975,6.188,6.325,6.400,6.538,6.675,6.763,6.838,6.938
                   NA,NA)

DF <- data.frame(year_quart, sales,moving_avg,
    centered_avg)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year_quart, group = 1)) + geom_line(
    aes(y = sales),color = "red") +
  geom_point(aes(y = sales)) + geom_line(aes(y =
      centered_avg),color = "blue") +
  geom_point(aes(y = centered_avg))+
  ylim(c(0.0,9.0)) +
  labs(title = "Years/quart V/S Sales Time Series
      Plot", x = "Year/Quarter", y = "Sales")
```

R code Exa 18.6b Deseasonalizing the Time Series

```
                                    # Page no. : 835 −
                                        836

# Deseasonalizing the Time Series
```

```r

year_quart <- c("Y1 Q1", "Y1 Q2", "Y1 Q3", "Y1 Q4", "Y2 Q1", "Y2 Q2", "Y2 Q3", "Y2 Q4", "Y3 Q1",
                 "Y3 Q2", "Y3 Q3", "Y3 Q4", "Y4 Q1", "Y4 Q2", "Y4 Q3", "Y4 Q4")
# Years and Quarters
period <- c(1:16)
sales <- c(4.8, 4.1, 6.0, 6.5, 5.8, 5.2, 6.8, 7.4, 6.0, 5.6, 7.5, 7.8, 6.3, 5.9, 8.0, 8.4)
index <- c(0.93,0.84,1.09,1.14,0.93,0.84,1.09,1.14,0.93,0.84,1.09,1.14,0.93
deseasonalized_sales <- c(5.16,4.88,5.50,5.70,6.24,6.19,6.24,6.49,6.45,6.67,6.88,6.84,6.77
                            7.02,7.34,7.37)

DF <- data.frame(year_quart,period,sales,index, deseasonalized_sales)

# Install Library if not installed

# install.packages("gpplot2")

# Import Library

library(ggplot2)

ggplot(DF,aes(year_quart,deseasonalized_sales, group = 1)) + geom_line() + geom_point() +
  ylim(c(0.0,9.0)) +
  labs(title = "Years/quart V/S Deseasonalized Sales Time Series Plot", x = "Year/Quarter",
       y = "Deseasonalized Sales")

regressor <- lm(deseasonalized_sales ~ period, data = DF)
summary(regressor)
```

```
anova(regressor)

b0 <- regressor$coefficients[1]
b1 <- regressor$coefficients[2]

cat("Equation is deseasonalized sales = ",b0,"+",b1,
   "Period")
```

Chapter 19

Non Parametric Tests

R code Exa 19.1a Rank Correlation

```
                                                # Page no. : 887 - 889

# Rank Correlation

sales_person <- c('A','B','C','D','E','F','G','H','I','J')
x <- c(2,4,7,1,6,3,10,9,8,5)   # Ranking of Potential
y <- c(400,360,300,295,280,350,200,260,220,385)   # Two Years Sales
z <- c(1,3,5,6,7,4,10,8,9,2)   # Ranking According to y

DF <- data.frame(sales_person,x,y,z)

d <- DF$x - DF$z
d_sq <- d**2

DF <- cbind(DF,d,d_sq)

```

```
total_d_sq <- sum(DF$d_sq)

n <- nrow(DF)

r_s <- 1 - ((6 * total_d_sq)/(n * (n**2 + 1)))   # Spearman Rank−Correlation Coefficient

mean_rs <- 0

sigma_rs <- sqrt(1 / (n - 1))

z_val <- (r_s - mean_rs) / sigma_rs

alpha <- 0.05

# Two Tail Test

# P−value Approach

pval <- 2 * pnorm(z_val,lower.tail =  F)

if(pval >= alpha)
{
  cat("Since pval",pval,"is greater than or equal to 0.05 therefore we cannot reject the Null Hypothesis")
} else{
  cat("Since pval",pval,"is less than 0.05 therefore we can reject the Null Hypothesis")
}
```

R code Exa 19.2a Sign Test

```
# Page no. : 858 − 861
```

```
# Sign Test

store <- c(56, 19, 36, 128, 12, 63, 39, 84, 102, 44)
sales <- c(485, 562, 415, 860, 426, 474, 662, 380,
   515, 721)

DF <- data.frame(store, sales)

median <- 450
sign <- c()

for (i in 1:nrow(DF)) {
  if(DF$sales[i] >= median)
  {
    sign[i] = "+"
  }else
  {
    sign[i] <- "-"
  }
}

DF <- cbind(DF,sign)
View(DF)

positive <- table(DF$sign)[[2]]
n <- nrow(DF)

test <- binom.test(positive, n)
test

if(test$p.value >= 0.05)
{
  cat("We cannot reject null hypothesis")
} else
{
  cat("We can reject null hypothesis")
}
```

```
N <- 60
n1 <- 22   # "+ sign
n2 <- 38   # "- sign
med <- 236000

mean <- 0.50 * N
sd <- sqrt(0.25 * N)

CF <- 22.5   # Correction Factor

p <- pnorm(CF, mean = mean, sd = sd)

if(p >= 0.05)
{
  cat("We cannot reject null hypothesis")
} else
{
  cat("We can reject null hypothesis")
}
```

R code Exa 19.3a Wilcoxon Signed Rank Test

```
                                        # Page no. : 866 -
                                            868

# Wilcoxon Signed - Rank Test

worker <- c(1:11)
A <- c(10.2, 9.6, 9.2, 10.6, 9.9, 10.2, 10.6, 10.0,
    11.2, 10.7, 10.6)
B <- c(9.5, 9.8, 8.8, 10.1, 10.3, 9.3, 10.5, 10.0,
    10.6, 10.2, 9.8)

```

```
DF <- data.frame(worker, A, B)

options(warn = -1)

test <- wilcox.test(DF$A, DF$B, paired = T)
test

if(test$p.value >= 0.05)
{
  cat("We cannot reject null hypothesis")
} else
{
  cat("We can reject null hypothesis")
}
```

R code Exa 19.4a Mann Whitney Wilcoxon Test

```
                                    # Page no. : 873 -
                                        875

# Mann - Whitney - Wilcoxon - Test

college <- c(1:4)
m1 <- c(15,3,23,8)

high <- c(1:5)
m2 <- c(18,20,32,9,25)

test <- wilcox.test(m1, m2, correct = F)
test

if(test$p.value >= 0.05)
{
  cat("We cannot reject null hypothesis")
} else
```

```
{
  cat("We can reject null hypothesis")
}
```

R code Exa 19.4b Mann Whitney Wilcoxon Test Eg2

```
                                        # Page no. :
                                          876-878

# Mann - Whitney - Wilcoxon - Test Eg - 2

account1 <- c(1:12)
account2 <- c(1:10)

balance1 <- c(1095, 955, 1200, 1195, 925, 950, 805,
    945, 875, 1055, 1025, 975)
balance2 <- c(885, 850, 915, 950, 800, 750, 865,
    1000, 1050, 935)

test <- wilcox.test(balance1, balance2, correct = F)
test

if(test$p.value >= 0.05)
{
  cat("We cannot reject null hypothesis")
} else
{
  cat("We can reject null hypothesis")
}
```

R code Exa 19.5a Kruskal Wallis Test

```
                                    # Page no. : 883 -
                                        884

# Kruskal - Wallis Test

A <- c(25, 70, 60, 85, 95, 90, 80)
B <- c(60, 20, 30, 15, 40, 35)
C <- c(50, 70, 60, 80, 90, 70, 75)

x <- list(A,B,C)

test <- kruskal.test(x)
test

if(test$p.value >= 0.05)
{
  cat("We cannot reject null hypothesis")
} else
{
  cat("We can reject null hypothesis")
}
```

Chapter 20

Statistical Process Control

R code Exa 20.1a Sample Mean Chart Process Mean and SD Unknown

```r
# Page no. : 913 - 919

# Sample Mean Chart : Process Mean and SD Unknown

o1 <- c(3.5056, 3.4882, 3.4897, 3.5153, 3.5059, 3.4977, 3.4910, 3.4991, 3.5099, 3.4880, 3.4881,
        3.5043, 3.5043, 3.5004, 3.4846, 3.5145, 3.5004, 3.4959, 3.4878, 3.4969)
o2 <- c(3.5086, 3.5085, 3.4898, 3.5120, 3.5113, 3.4961, 3.4913, 3.4853, 3.5162, 3.5015, 3.4887,
        3.4867, 3.4769, 3.5030, 3.4938, 3.4832, 3.5042, 3.4823, 3.4864, 3.5144)
o3 <- c(3.5144, 3.4884, 3.4995, 3.4989, 3.5011, 3.5050, 3.4976, 3.4830, 3.5228, 3.5094, 3.5141,
        3.4946, 3.4944, 3.5082, 3.5065, 3.5188, 3.4954, 3.4964, 3.4960, 3.5053)
o4 <- c(3.5009, 3.5250, 3.5130, 3.4900, 3.4773, 3.5014, 3.4831, 3.5083, 3.4958, 3.5102, 3.5175,
        3.5018, 3.5014, 3.5045, 3.5089, 3.4935, 3.5020, 3.5082, 3.5070, 3.4985)
```

```
o5 <- c(3.5030, 3.5031, 3.4969, 3.4837, 3.4801, 3.5060, 3.5044, 3.5094, 3.5004, 3.5146, 3.4863,
        3.4784, 3.4904, 3.5234, 3.5011, 3.4989, 3.4889, 3.4871, 3.4984, 3.4885)

DF <- data.frame(o1, o2, o3, o4, o5)

sample_mean <- rowMeans(DF)
sample_range <- c(0.0135, 0.0368, 0.0233, 0.0316, 0.0340, 0.0099, 0.0213, 0.0264, 0.0270, 0.0266,
                  0.0312, 0.0259, 0.0274, 0.0230, 0.0243, 0.0356, 0.0153, 0.0259, 0.0206, 0.0259)

DF <- cbind(DF, sample_mean, sample_range)
View(DF)

AR <- mean(DF$sample_range)   # Average Range
OM <- mean(DF$sample_mean)  # Overall Mean
n <- 5   # Sample Observations
d2 <- 2.362
A2 <- 3 / (d2 * sqrt(n))

UCL <- OM + (A2 * AR)   # Upper Control Limit
LCL <- OM - (A2 * AR)   # Lower Control Limit

cat("UCL is ",UCL)
cat("LCL is",LCL)

library(qicharts2)

qic(DF$sample_mean, xlab = "Sample Number", ylab = "Sample Mean", title = "Sample Mean Chart")

d3 <- 0.864

D1 <- 1 + (3 * (d3 / d2))
D2 <- 1 - (3 * (d3 / d2))
```

```
UCL2 <- AR * D1
LCL2 <- AR * D2

cat("UCL is ",UCL2)
cat("LCL is",LCL2)    # Book answer is different

qic(DF$sample_range, xlab = "Sample Number", ylab =
    "Sample Range", title = "R Chart")

p <- 0.03
n <- 200

sigma <- sqrt((p * (1 - p)) / n)

UCL3 <- p + 3 * sigma
LCL3 <- p - 3 * sigma

cat("UCL is ",UCL3)
cat("LCL is",LCL3)

# Data for P chart is not available in Book

UCL4 <- n * p + 3 * sqrt(n * p * (1 - p))
LCL4 <- n * p - 3 * sqrt(n * p * (1 - p))

cat("UCL is ",UCL4)
cat("LCL is",LCL4)
```

Chapter 21

Decision Analysis

R code Exa 21.1a Problem Formulation

```
# Page no. : 940 - 941

# Problem Formulation

decision <- factor(c("Small Complex, d1", "Medium Complex, d2", "Large Complex, d3"))
demand1 <- c(8, 14, 20)
demand2 <- c(7, 5, -9)

DF <- data.frame(decision, demand1, demand2)

# Install Library if not install

install.packages("rpart")
install.packages("rpart.plot")

# Import Library

library(rpart)
library(rpart.plot)
```

```
ans <- rpart(decision ~ ., data = DF, method = "class" )
rpart.plot(ans)

# Decision Tree is Different from Book
```