R Textbook Companion for Statistics in Education and Psychology by P. C. Dash and Bhabagrahi Biswal[1]

Created by
Nivedha R
M.Sc.
Information Technology
Tamil Nadu Agricultural University
Cross-Checked by
R TBC Team

August 28, 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 in Education and Psychology

Author: P. C. Dash and Bhabagrahi Biswal

Publisher: Dominant Publishers & Distributors Pvt Ltd

Edition: 2

Year: 2009

ISBN: 978-93-82007-64-7

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

Exa Example (Solved example)

Eqn Equation (Particular equation of the above book)

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

Contents

List of R Codes

Chapter 2

Some Fundamentals of Statistics in Education and Psychology

R code Exa 2.6.1 Class Interval

```
# Page.No 2.5

#(a) class limit
lower_limit<-10
upper_limit<-20

lower_value<-lower_limit-0.5
upper_value<-upper_limit+0.5

cat("The lower and upper values of the class
    interval 10-20 is",lower_value,upper_value)

#(b) class midpoint

midpoint<-lower_limit+(upper_limit-lower_limit)/2

cat("mid point of a class interval 10-20 is",
```

```
  midpoint)
```

Chapter 3

Frequency Distribution

R code Exa 3.3.1 Procedure

```
# Page.No 3.4

#(i) Determining Range
scores<-c
   (10,15,16,17,18,20,22,25,30,35,40,42,45,46,47,48,49,47,48,49,50,1

HS<-max(scores)
LS<-min(scores)
range<-HS-LS

cat("The range of the frequency distribution is ",
   range)

#(ii) Determining size of class interval

no.of.classes<-8
size<-range/no.of.classes

cat("The size of the class intervals is ",size)

#(vii) frequency table
```

```
low_value<-9.5
high_value<-54.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
c<-cut(scores,breaks = X_breaks)
table<-table(c)
data<-data.frame(table)
print(data)

N<-sum(data$Freq)
cat("Total no of frequencies is",N)
```

Chapter 4

Graphic Presentation

R code Exa 4.7.1 Histogram

```
# Page.No 4.8

# Given data
low_value<-9.5
high_value<-34.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
f<-c(5,4,6,4,2)
data<-rep(X_mid,f)

# Histogram
hist(data,main="Histogram",
     xlab="class_limits",
     border="black",
     col="Steelblue3",
     xlim=c(0,40),
     las=1,
     breaks=5)
```

R code Exa 4.8.1 Frequency Polygon

```
# Page.No 4.9

# Given data table
low_value<-4.5
high_value<-39.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_breaks
X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
X_mid
f<-c(0,2,3,6,4,5,0)
data<-rep(X_mid,f)

# Plottinf the frequency polugon
plot(X_mid,f,type = "line",main = "title",xlab = "
    Mid Points",ylab = "Frequencies",col = "red")

## Another method
hist<-(hist(data,main ="Histogram",xlab="class_
    limits",border="black",col="Steelblue3",xlim=c
    (0,50),las=1,breaks=5))
library("agricolae")
polygon.freq(hist,f=1,col="red")
```

R code Exa 4.9.1 Ogive

```
# Page.No 4.11

x<-seq(9.5,34.5,5) # class interval
```

```
f<-c(0,5,4,6,3,2)  # frequency
cf<-cumsum(f)   # cumulative frequency
percent<-cf/sum(f)*100
plot(x,percent,type = "o",main="OGIVE",xlim = c
   (9.5,50),ylim=c(0,100),xlab = "Upper limits",ylab
    = "Cumulative Percentage Frequency")
```

R code Exa 4.10.1 Pie Diagram

```
# Page.No 4.12

numeric<-c(50,25,20,5)
degree<-numeric*360/100

pie(degree,main = "pie-diagram",col = c("gray30","
   gray80","black","white"),labels = c(degree))
```

Chapter 5

Measures of Central Tendency

R code Exa 5.3.1 Mean

```
#Page.No 5.5

marks<-c(70,30,20,90,40)  # Marks in history

avg<-mean(marks) # Mean of history marks

cat("The average marks of the students is",avg)
```

R code Exa 5.3.2 Mean

```
#Page.No 5.5

marks<-c(10,20,30,40,50,60,70,80,90)  # marks in
    English
avg<-mean(marks)                       # average of marks
cat("The average marks of the students is",avg)
```

R code Exa 5.3.3 Mean

```r
#Page.No 5.6

scores<-c(80,90,95,96,85,84,82)  # marks in
   mathematics
avg<-mean(scores)    # average of marks
cat("The average marks of the students is",avg)
```

R code Exa 5.3.4 Calculation of Mean from Grouped Data

```r
#Page.No 5.7

# Given data table
low_value<-9.5
high_value<-34.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
f<-c(5,4,6,3,2)  # frequency of the distribution
data<-rep(X_mid,f)
mean(data)

# to create a table from raw data
x<-cut(data, breaks=X_breaks)
datatable<-table(x)
df<-data.frame(datatable)
df$midpnt<-X_mid
df
```

R code Exa 5.3.5 Calculation of Mean from Grouped Data

```
#Page.No 5.9

Lower_limits<-c(5,11,17,23,29)
Upper_limits<-c(10,15,22,28,35)
frequency<-c(1,4,3,7,5)
library("gds")  # function for Descriptive
    statistics of grouped data
results<-gds(Lower_limits,Upper_limits,frequency)
cat("the mean of the distribution is",results$mean)
```

R code Exa 5.3.6 Calculation of Mean from Grouped Data

```
#Page.No 5.12 -5.14

Lower_limits<-c(10,15,20,25,30)
Upper_limits<-c(14,19,24,29,35)
frequency<-c(5,10,7,2,1)
library("gds")  # Function for Descriptive
    statistics of grouped data
results<-gds(Lower_limits,Upper_limits,frequency)
cat("the mean of the distribution is",results$mean)
```

R code Exa 5.3.7 Combined Mean

```
#Page.No 5.14

mean<-c(50,45,40,45) # mean scores in history
no.of.students<-c(20,20,15,30) # no.of students in
    section
wm<-weighted.mean(mean,no.of.students)
cat("Combined mean is",wm)
```

R code Exa 5.3.8 Combined Mean

```
#Page.No 5.15

No.of.students<-c(12,8,10,15,7)
Mean<-c(50,70,30,45,48)   # Mean in English test
weighted.mean(Mean,No.of.students)
```

R code Exa 5.3.9 Combined Mean

```
#Page.No 5.15

no.of.students<-c(8,12,10,14,9)
Mean_of_achievement_test<-c(55,78,80,50,87)
M_comb<-weighted.mean(Mean_of_achievement_test,no.of
    .students)
cat("combined mean of the achievement test of
    students of 5 classes is",M_comb)
```

R code Exa 5.3.10 Properties of Mean

```
#Page.No 5.16

X<-c(8,5,6,4,7)
d<-X-mean(X)
cat("The sum of deviations is",sum(d))
```

R code Exa 5.3.11 Properties of Mean

```
#Page.No 5.17

X<-c(10,12,18,16,14)
d<-X-mean(X)
sum_of_deviations<-sum(d)
sum_of_dsquares<-sum(d^2)
cat("The Sum of squares of deviations is",sum_of_dsquares)# which is greater than sum of squares
mean(X)
```

R code Exa 5.3.12 Properties of Mean

```
#Page.No 5.17

#If each score is added or subtracted by a constant quantity
#The mean will increased by same quantity
X<-c(1,2,3,4,5,6,7)
mean(X)

# Addition of constant 5
X_add<-X+5
mean(X_add)
mean(X)+5 # the same constant addition

# subtrction of constant 5
X_sub<-X-5
mean(X_sub)
mean(X)-5 # the same constant subtraction
```

R code Exa 5.4.1 Calculation of Median

```
# Page.No 5.24

scores<-c(10,15,16,14,17,20,21)
md<-median(scores)
cat("The median is",md)
```

R code Exa 5.4.2 Calculation of Median

```
#Page.No 5.25

scores<-c(15,10,14,16)
Md<-median(scores)
cat("The median of the scores is",Md)
```

R code Exa 5.4.3 Calculation of Median

```
#Page.No 5.25

# (i)
scores<-c(10,8,25,24,30,45)
median(scores)

#(ii)
Scores<-c(40,48,43,42,41)
median(Scores)

#(iii)
score<-c(11,13,15,15,15,18,21)
median(score)

lower_value<-median(score)-0.5 # from the lower
   limit
corrected_median<-lower_value+2/3
```

```
cat("median is ",corrected_median)
```

R code Exa 5.4.7 Calculation of Median from Grouped Data

```
#Page.No 5.30

Lower_limits<-c(10,15,20,25,30)
Upper_limits<-c(14,19,24,29,34)
frequency<-c(5,4,6,3,2)
library("gds")
results<-gds(Lower_limits,Upper_limits,frequency)
cat("the median of the distribution is",results$
    median)
```

R code Exa 5.4.8 Calculation of Median from Grouped Data

```
#Page.No 5.32 - 5.33

library("gds") #function for Descriptive statistics
    of grouped data
lower_limits<-c(10,15,20,25,30)
upper_limits<-c(14,19,24,29,34)
frequency<-c(5,4,6,3,2)
md<-gds(lower_limits,upper_limits,frequency)
cat("The median is",md$median)
```

R code Exa 5.4.9 Calculation of Median from Grouped Data

```
#Page.No 5.33

```

```
#(i) and (ii)
Lower_limits<-c(40,50,60,70,80,90)
Upper_limits<-c(49,59,69,79,89,99)
frequency<-c(14,12,15,10,15,14)

library("gds")  # Function to find Descriptive statistics of grouped data
results<-gds(Lower_limits,Upper_limits,frequency)
cat("the median of the distribution is",results$median)
#The  answer may slightly vary due to  rounding off values
```

R code Exa 5.5.1 Mode

```
#Page.No 5.42

# Create the function.
mode <- function(x) {
  unique_val <- unique(x)
  counts <- vector()
  for (i in 1:length(unique_val)) {
    counts[i] <- length(which(x==unique_val[i]))
  }
  position <- c(which(counts==max(counts)))
  if (mean(counts)==max(counts))
    mode_x <- 'Mode does not exist'
  else
    mode_x <- unique_val[position]
  return(mode_x)
}

# Create the vector with numbers.
v <- c(10,20,10,25,10,20,10,22,28)

```

```
# Calculate the mode using the user function.
result <- mode(v)
cat("The mode is",result)
```

R code Exa 5.5.2 Mode

```
#Page.No 5.43

x<-c(30,31,32,33,34,35)#size of the garments
f<-c(20,30,40,60,25,10) # no of persons wearing
data<-rep(x,f)

# Creating function to find mode
## Or u can use the previous functions too
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

md<-Mode(data)

cat ("size of the dress most frequently used in
    season is",md)
```

R code Exa 5.5.3 Mode

```
#Page.No 5.43

# Create the function.
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

```
# Create the vector with numbers.
Scores <- c(20,20,30,20,25,20,38,37,38)

# Calculate the mode using the user function.
result <- getmode(Scores)
cat("The mode is",result)
```

R code Exa 5.5.4 Mode

```
#Page.No 5.43

# Create the function.
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

# Create the vector with numbers.
scores <- c(20,25,20,15,19,22,20,22)

# Calculate the mode using the user function.
result <- getmode(scores)
cat("The mode is",result)
```

R code Exa 5.5.6 Bi modal

```
#Page.No 5.43

Scores <- c(20,25,20,28,29,28,30,32)

#Creating a function
mode <- function(x) {
```

```
  unique_val <- unique(x)
  counts <- vector()
  for (i in 1:length(unique_val)) {
    counts[i] <- length(which(x==unique_val[i]))
  }
  position <- c(which(counts==max(counts)))
  if (mean(counts)==max(counts))
    mode_x <- 'Mode does not exist'
  else
    mode_x <- unique_val[position]
  return(mode_x)
}

mode(Scores)
# the series is Bi-modal
```

R code Exa 5.5.7 Calculation of Mode from Grouped Data

```
# Page.no 5.44

##Creating a function
mode <- function(x) {
  unique_val <- unique(x)
  counts <- vector()
  for (i in 1:length(unique_val)) {
    counts[i] <- length(which(x==unique_val[i]))
  }
  position <- c(which(counts==max(counts)))
  if (mean(counts)==max(counts))
    mode_x <- 'Mode does not exist'
  else
    mode_x <- unique_val[position]
  return(mode_x)
}

```

```
y<-c
   (11,11,12,12,12,13,13,13,13,14,14,14,14,15,15,16,17,18)

Md<-mode(y)

## Since 13 and 14 are adjacent numbers their mean
   can be considered as the mode
Mode<-mean(Md)
cat("The mode will be", Mode)
```

R code Exa 5.5.8 Bi modal

```
# Page.no 5.44

##Creating a function
mode <- function(x) {
  unique_val <- unique(x)
  counts <- vector()
  for (i in 1:length(unique_val)) {
    counts[i] <- length(which(x==unique_val[i]))
  }
  position <- c(which(counts==max(counts)))
  if (mean(counts)==max(counts))
    mode_x <- 'Mode does not exist'
  else
    mode_x <- unique_val[position]
  return(mode_x)
}

y<-c
   (11,11,12,12,12,13,13,13,13,13,14,14,14,14,15,15,15,15,16,16,16,1

Md<-mode(y)

# value 13 occurs 5 times which is greater than the
```

```
    frequency of the adjacent values
# Also 15 occurs 4 times which is also greater than
    the frequency of the adjacent values_

# hence this is Bi-modal
cat("The mode values are",13 ,15)
```

R code Exa 5.5.9 Calculation of Mode from Grouped Data

```
#Page.No 5.45

# Given frequency distribution data
low_value<-9.5
high_value<-39.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_breaks
X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
X_mid
f<-c(2,3,2,8,3,2)
data<-rep(X_mid,f)

# creating function to find mode
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
Mode(data)
```

R code Exa 5.5.10 Calculation of Mode from Grouped Data

```
#Page.No 5.45
```

```r

# A function created to find the median of a grouped data
GroupedMedian <- function(frequencies, intervals, sep = NULL, trim = NULL) {
  # If "sep" is specified, the function will try to create the
  #   required "intervals" matrix. "trim" removes any unwanted
  #   characters before attempting to convert the ranges to numeric.
  if (!is.null(sep)) {
    if (is.null(trim)) pattern <- ""
    else if (trim == "cut") pattern <- "\\[|\\]|\\(|\\)"
    else pattern <- trim
    intervals <- sapply(strsplit(gsub(pattern, "", intervals), sep), as.numeric)
  }

  Midpoints <- rowMeans(intervals)
  cf <- cumsum(frequencies)
  Midrow <- findInterval(max(cf)/2, cf) + 1
  L <- intervals[1, Midrow]      # lower class boundary of median class
  h <- diff(intervals[, Midrow]) # size of median class
  f <- frequencies[Midrow]       # frequency of median class
  cf2 <- cf[Midrow - 1]          # cumulative frequency class before median class
  n_2 <- max(cf)/2               # total observations divided by 2

  unname(L + (n_2 - cf2)/f * h)
}

# The Problem
```

```
frequency<-c(5,4,6,3,2)
colnames<-c("numbers")
rownames<-c("[9.5,14.5]"," (14.5,19.5]","(19.5,24.5]","(24.5,29.5]"," (29.5,34.5]")
#The class intevals are changed to exclusive one by adding 0.5 to the upper limits and deducing 0.5 from the lower limits
y<-matrix(frequency,nrow=length(frequency),dimnames=list(rownames,colnames))
md<-GroupedMedian(y[, "numbers"], rownames(y), sep=",", trim="cut")


### find the mean for grouped data
low_value<-9.5
high_value<-34.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
f<-c(5,4,6,3,2)
data<-rep(X_mid,f)
Mn<-mean(data)

## to find mode from mean and median
Mode<-3*md-2*Mn
cat ("The mode is",Mode)

# Page.No 5.47
###(ii)Another method
Lower_limits<-c(10,15,20,25,30)
Upper_limits<-c(14,19,24,29,34)
frequency<-c(5,4,6,3,2)
library("gds") # Package for grouped statistics
results<-gds(Lower_limits,Upper_limits,frequency)
cat("the mode of the distribution is",results$mode)
```

R code Exa 5.5.11 Calculation of Mode from Grouped Data

```r
# Page.No 5.48

# A function created to find the median of a grouped data
GroupedMedian <- function(frequencies, intervals, sep = NULL, trim = NULL) {
  # If "sep" is specified, the function will try to create the
  #  required "intervals" matrix. "trim" removes any unwanted
  #  characters before attempting to convert the ranges to numeric.
  if (!is.null(sep)) {
    if (is.null(trim)) pattern <- ""
    else if (trim == "cut") pattern <- "\\[|\\]|\\(|\\)"
    else pattern <- trim
    intervals <- sapply(strsplit(gsub(pattern, "", intervals), sep), as.numeric)
  }

  Midpoints <- rowMeans(intervals)
  cf <- cumsum(frequencies)
  Midrow <- findInterval(max(cf)/2, cf) + 1
  L <- intervals[1, Midrow]      # lower class boundary of median class
  h <- diff(intervals[, Midrow]) # size of median class
  f <- frequencies[Midrow]       # frequency of median class
  cf2 <- cf[Midrow - 1]          # cumulative frequency class before median class
```

```r
  n_2 <- max(cf)/2                    # total observations divided by 2

  unname(L + (n_2 - cf2)/f * h)
}

# The Problem
frequency<-c(2,3,1,4,2,1,2,4,3,2,7,4)
colnames<-c("numbers")
rownames<-c("[1.5,8.5]"," (8.5,15.5]"," (15.5,22.5]","(22.5,29.5]"," (29.5,36.5]","(36.5,43.5]",
            "(43.5,50.5]","(50.5,57.5]","(57.5,64.5]","(64.5,71.5]","(71.5,78.5]"," (78.5,85.5]")
#The class intevals are changed to exclusive one by adding 0.5 to the upper limits and deducing 0.5 from the lower limits
y<-matrix(frequency,nrow=length(frequency),dimnames=list(rownames,colnames))
md<-GroupedMedian(y[, "numbers"], rownames(y), sep=",", trim="cut")


### find the mean for grouped data
low_value<-1.5
high_value<-85.5
step_value<-7
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
f<-c(2,3,1,4,2,1,2,4,3,2,7,4)
data<-rep(X_mid,f)
Mn<-mean(data)

## to find mode from mean and median
Mode<-3*md-2*Mn
cat ("The mode is",Mode)
```

R code Exa 5.7.1 Calculation of Mean Median Mode from Grouped Data

```r
#Page.No 5.51 - 5.53

Lower_limits<-c(10,15,20,25,30,35,40,45)
Upper_limits<-c(14,19,24,29,34,39,44,49)
frequency<-c(3,6,5,9,8,8,2,8)
library("gds") # Package for grouped statistics
results<-gds(Lower_limits,Upper_limits,frequency)
cat("The mean is ",results$mean)
cat("The median is ",results$median)
cat("The mode is ",results$mode)
# The answers may slightly vary due to roundingoff values



#### To get the Precise answers you can try the following method
# A function created to find the median of a grouped data
GroupedMedian <- function(frequencies, intervals, sep = NULL, trim = NULL) {
  # If "sep" is specified, the function will try to create the
  #   required "intervals" matrix. "trim" removes any unwanted
  #   characters before attempting to convert the ranges to numeric.
  if (!is.null(sep)) {
    if (is.null(trim)) pattern <- ""
    else if (trim == "cut") pattern <- "\\[|\\]|\\(|\\)"
    else pattern <- trim
```

```
    intervals <- sapply(strsplit(gsub(pattern, "", intervals), sep), as.numeric)
  }

  Midpoints <- rowMeans(intervals)
  cf <- cumsum(frequencies)
  Midrow <- findInterval(max(cf)/2, cf) + 1
  L <- intervals[1, Midrow]      # lower class boundary of median class
  h <- diff(intervals[, Midrow]) # size of median class
  f <- frequencies[Midrow]       # frequency of median class
  cf2 <- cf[Midrow - 1]          # cumulative frequency class before median class
  n_2 <- max(cf)/2               # total observations divided by 2

  unname(L + (n_2 - cf2)/f * h)
}

# The Problem
frequency<-c(3,6,5,9,8,8,2,8)
colnames<-c("numbers")
rownames<-c("[9.5,14.5]","(14.5,19.5]","(19.5,24.5]","(24.5,29.5]",
            "(29.5,34.5]","(34.5,39.5]","(39.5,44.5]","(44.5,49.5]")
#The class intevals are changed to exclusive one by adding 0.5 to the upper limits and deducing 0.5 from the lower limits
y<-matrix(frequency,nrow=length(frequency),dimnames=list(rownames,colnames))
md<-GroupedMedian(y[, "numbers"], rownames(y), sep=",", trim="cut")


### find the mean for grouped data
```

```
low_value<-9.5
high_value<-49.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
f<-c(3,6,5,9,8,8,2,8)
data<-rep(X_mid,f)
Mn<-mean(data)
Mn

## to find mode from mean and median
Mode<-3*md-2*Mn
cat ("The mode is",Mode)
cat("The median is",md)
cat("The mean is",Mn)
```

R code Exa 5.7.2 Calculation of Mean Median Mode from Grouped Data

```
# Page.No 5.53

Lower_limits<-c(100,110,120,130,140,150,160)
Upper_limits<-c(109,119,129,139,149,159,169)
frequency<-c(10,12,13,18,6,7,4)
library("gds") # Package for grouped statistics
results<-gds(Lower_limits,Upper_limits,frequency)
cat("The mean is ",results$mean)
cat("The median is ",results$median)

GroupedMedian <- function(frequencies, intervals, sep = NULL, trim = NULL) {
  # If "sep" is specified, the function will try to create the
  #  required "intervals" matrix. "trim" removes any unwanted
```

```
  #   characters before attempting to convert the ranges to numeric.
  if (!is.null(sep)) {
    if (is.null(trim)) pattern <- ""
    else if (trim == "cut") pattern <- "\\[|\\]|\\(|\\)"
    else pattern <- trim
    intervals <- sapply(strsplit(gsub(pattern, "", intervals), sep), as.numeric)
  }

  Midpoints <- rowMeans(intervals)
  cf <- cumsum(frequencies)
  Midrow <- findInterval(max(cf)/2, cf) + 1
  L <- intervals[1, Midrow]      # lower class boundary of median class
  h <- diff(intervals[, Midrow]) # size of median class
  f <- frequencies[Midrow]       # frequency of median class
  cf2 <- cf[Midrow - 1]          # cumulative frequency class before median class
  n_2 <- max(cf)/2               # total observations divided by 2

  unname(L + (n_2 - cf2)/f * h)
}

# The Problem
frequency<-c(10,12,13,18,06,07,04)
colnames<-c("numbers")
rownames<-c("[99.5,109.5]"," (109.5,119.5]","(119.5,129.5]","(129.5,139.5]",
            "(139.5,149.5]","(149.5,159.5]","(159.5,169.5]")
#The class intevals are changed to exclusive one by adding 0.5 to the upper limits and deducing 0.5 from the lower limits
```

```
y<-matrix(frequency,nrow=length(frequency),dimnames=list(rownames,colnames))
md<-GroupedMedian(y[, "numbers"], rownames(y), sep=",", trim="cut")


Mode <- 3*md-2*results$mean
cat("The mode is",Mode)
```

Chapter 6

Percentile and Percentile Rank

R code Exa 6.2.1 Percentile

```
# Page.No 6.5 - 6.8

## Function to find percentile
percentile <- function(frequencies, intervals, sep = NULL, trim = NULL,p) {
  # If "sep" is specified, the function will try to create the
  #   required "intervals" matrix. "trim" removes any unwanted
  #   characters before attempting to convert the ranges to numeric.
  if (!is.null(sep)) {
    if (is.null(trim)) pattern <- ""
    else if (trim == "cut") pattern <- "\\[|\\]|\\(|\\)"
    else pattern <- trim
    intervals <- sapply(strsplit(gsub(pattern, "", intervals), sep), as.numeric)
  }
  Midpoints <- rowMeans(intervals)
  cf <- cumsum(frequencies)
```

```
  row <- findInterval(max(cf)*p, cf) + 1
  L <- intervals[1, row]       # lower class boundary of percentile class
  h <- diff(intervals[, row]) # size of percentile class
  f <- frequencies[row]       # frequency of percentile class
  cf2 <- cf[row - 1]          # cumulative frequency class before percentile class
  n_2 <- max(cf)*p      # total observations multiplied with percentile

  unname(L + (n_2 - cf2)/f * h)
}

# The given Problem
frequency<-c(2,3,4,5,3,2,2,3)
colnames<-c("numbers")
rownames<-c("[9.5,14.5]"," (14.5,19.5]","(19.5,24.5]","(24.5,29.5]"," (29.5,34.5]","(34.5,39.5]","(39.5,44.5]","(44.5,49.5]")
#The class intevals are changed to exclusive one by adding 0.5 to the upper limits and deducing 0.5 from the lower limits

y<-matrix(frequency,nrow=length(frequency),dimnames=list(rownames,colnames))

p=c(0.15,0.25,0.50,0.65,0.75,0.80,0.85)

Percentile_value <-percentile(y[, "numbers"], rownames(y), sep=",", trim="cut",p)

Result<- round(rbind(p*100,Percentile_value),digits = 3)

print(Result)
```

R code Exa 6.3.1 Percentile rank

```
#Page.No 6.11
#Page.No 6.11

data<-as.vector(c
    (10,8,7,15,18,19,20,28,16,13,17,29,30,6,5,4,26,24,14,21)
    )
N<-length(data)

# ranking of scores
r<-rank(-data) # negative for rank the data in
    descending order
rank_matrix<-cbind(r,data)


# Rank of score 18
rank_matrix[5,]
R<-rank_matrix[5,1]

percentile_rank<-round(100-((100*R-50))/N,digits =
    0)
cat("The percentile rank of 18 is approximately ",
    percentile_rank)
```

R code Exa 6.3.2 Percentile rank

```
# Page No.6.12

data<-as.vector(c(65,59,46,32,25,15,14,10,9,7,5,3))
N<-length(data)

```

```
# ranking of scores
r<-rank(-data) # negative for rank the data in
   descending order
rank_matrix<-cbind(r,data)


# Rank of score 18
rank_matrix[5,]
R<-rank_matrix[5,1]

percentile_rank<-100-((100*R-50))/N
cat("The percentile rank of 18 is approximately ",
   percentile_rank)
```

R code Exa 6.3.3 Percentile rank

```
#Page.No 6.13

# Function to find percentile rank of grouped data
percentilerank<-function(class_interval,lower_limit,
   x){
  L<-lower_limit
  row<-which(data$c==class_interval)
  f<-data$Freq[row]
  cf<-data.frame(cumsum(data$Freq))
  cf1<-cf$cumsum.data.Freq.[row]
  cf2<-cf$cumsum.data.Freq.[row-1]
  F<-if(row - 1<= 0) {
    cf1
    }else { cf2 }
  N<-sum(data$Freq)
  i<-step_value
  unname((100/N)*(F+(x-L)/i*f))
}

```

```
# Given data
low_value<-9.5
high_value<-69.5
step_value<-10
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
frequency<-c(4,8,2,2,3,1)
data<-rep(X_mid,frequency)
c<-cut(data,breaks = X_breaks)
table<-table(c)
data<-data.frame(table)
print(data)

# percentile rank of 24

PR22<-percentilerank("(19.5,29.5]",19.5,22)

cat("The percentile rank of 22 is ",PR22)
```

R code Exa 6.3.4 Percentile rank

```
#Page.No 6.13

# Function to find percentile rank of grouped data
percentilerank<-function(class_interval,lower_limit,
   x){
  L<-lower_limit
  row<-which(data$c==class_interval)
  f<-data$Freq[row]
  cf<-data.frame(cumsum(data$Freq))
  cf1<-cf$cumsum.data.Freq.[row]
  cf2<-cf$cumsum.data.Freq.[row-1]
  F<-if(row - 1<= 0) {
    cf1
```

```
    }else { cf2 }
  N<-sum(data$Freq)
  i<-step_value
  unname((100/N)*(F+(x-L)/i*f))
}

# Given data
low_value<-9.5
high_value<-89.5
step_value<-10
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
frequency<-c(2,3,4,5,2,1,1,2)
data<-rep(X_mid,frequency)
c<-cut(data,breaks = X_breaks)
table<-table(c)
data<-data.frame(table)
print(data)

# percentile rank of 24

PR24<-percentilerank("(19.5,29.5]",19.5,24)
PR36<-percentilerank("(29.5,39.5]",29.5,36)
PR52<-percentilerank("(49.5,59.5]",49.5,52)
PR55<-percentilerank("(49.5,59.5]",49.5,55)
PR67<-percentilerank("(59.5,69.5]",59.5,67)

results<-c(PR24,PR36,PR52,PR55,PR67)
Q<-c("PR24","PR36","PR52","PR55","PR67")

percentile_ranks<-data.frame(Q,results)
print(percentile_ranks)
```

R code Exa 6.4.1 Percentile Problems

```r
# Page.No 6.18 - 6.23

percentile <- function(frequencies, intervals, sep = NULL, trim = NULL,p) {
  # If "sep" is specified, the function will try to create the
  #   required "intervals" matrix. "trim" removes any unwanted
  #   characters before attempting to convert the ranges to numeric.
  if (!is.null(sep)) {
    if (is.null(trim)) pattern <- ""
    else if (trim == "cut") pattern <- "\\[|\\]|\\(|\\)"
    else pattern <- trim
    intervals <- sapply(strsplit(gsub(pattern, "", intervals), sep), as.numeric)
  }
  Midpoints <- rowMeans(intervals)
  cf <- cumsum(frequencies)
  row <- findInterval(max(cf)*p, cf) + 1
  L <- intervals[1, row]       # lower class boundary of percentile class
  h <- diff(intervals[, row]) # size of percentile class
  f <- frequencies[row]        # frequency of percentile class
  c <- cf[row - 1]
  cf2 <- if(row - 1<= 0 ) {0}
  else { cf[row - 1] } # cumulative frequency class before percentile class
  n_2 <- max(cf)*p      # total observations multiplied with percentile

  unname(L + (n_2 - cf2)/f * h)
}


```

```
# The Problem
frequency<-c(9,5,7,8,8,6,2,3,2)
colnames<-c("numbers")
rownames<-c("[9.5,14.5]"," (14.5,19.5]","(19.5,24.5]","(24.5,29.5]"," (29.5,34.5]","(34.5,39.5]","(39.5,44.5]","(44.5,49.5]","(49.5,54.5]")
#The class intevals are changed to exclusive one by adding 0.5 to the upper limits and deducing 0.5 from the lower limits
y<-matrix(frequency,nrow=length(frequency),dimnames=list(rownames,colnames))


p1=c(0.90,0.80,0.75,0.70,0.60,0.50,0.48,0.40,0.30,0.20)

p2=c(0.17,0.10)

Percentile_value1<-percentile(y[, "numbers"], rownames(y), sep=",", trim="cut",p1)

Percentile_value2<-percentile(y[, "numbers"], rownames(y), sep=",", trim="cut",p2)

Result1<- round(rbind(p1*100,Percentile_value1), digits = 3) # percentile values of above 20

Result2<- round(rbind(p2*100,Percentile_value2), digits = 3) # percentile values of below 20

print(Result1)

print(Result2)
```

R code Exa 6.5.1 Graphical Representation of Percentile

```
# Page.No 6.23-6.24

low_value<-19.5
high_value<-64.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
frequency<-c(1,2,2,7,10,8,5,3,2)
scores<-rep(X_mid,frequency)
cf<-cumsum(frequency)
cpf<-cf/sum(frequency)*100
plot(X_mid,cpf,xlim=c(10,70),ylim = c(0,100),type =
   "o",xlab = "scores",ylab = "Cumulative percentage
    frequency",main = "Graphic Representation of
   percentiles")
abline(h=c(17,25,50,63,75),untf = FALSE)

## Or to make the clear visualization
abline(h=c(17,25,50,63,75),v=c(33.3,35.5,41,44,47))

## to give colours
abline(h=c(17,25,50,63,75),v=c(33.3,35.5,41,44,47),
   col=c("red","green","blue","orange","yellow"))
```

Chapter 7

Measures of Dispersion

R code Exa 7.5.1 Range

```
#Page. No 7.5

scores<-c(5,7,9,15,17,19)
range<-max(scores)-min(scores)
cat("the range of the scores is",range)
```

R code Exa 7.5.2 Range

```
# Page.No 7.5

scores<-c(19,22,90,100,150,190)
range<-max(scores)-min(scores)
cat("the range of the scores is",range)
```

R code Exa 7.5.3 Coefficient of Range

```
#Page.No 7.6

scores<-c(10,20,30,40,50,60,65,66,70,80)
hs<-max(scores)
ls<-min(scores)
Coef.R<-(hs-ls)/(hs+ls)
cat("The coefficient of Range is",Coef.R)
```

R code Exa 7.5.4 Coefficient of Range for Grouped Data

```
#Page.No 7.6

class_intervals <-c(10,14,15,19,20,24,25,29,30,34)
frequency<-c(2,8,6,2,2)
HS<-max(class_intervals)   # highest score
LS<-min(class_intervals)   #Lowest score

Coef.R<-(HS-LS)/(HS+LS)

cat("The coefficient of Range is",Coef.R)
```

R code Exa 7.5.5 Coefficient of Range for Grouped Data

```
#Page.No 7.7

class_intervals <-c(5,9,10,14,15,19,20,24,25,29)
frequency<-c(2,4,5,7,2)
HS<-max(class_intervals)   #Highest score
LS<-min(class_intervals)   # Lowest score

Coef.R<-(HS-LS)/(HS+LS)

cat("The coefficient of Range is",Coef.R)
```

R code Exa 7.5.6 Coefficient of Range for Grouped Data

```
#Page.No 7.7

frequency<-c(2,2,3,3)
scores<-c(10,14,15,19,20,24,25,29)
HS<-max(scores)  #Highest score
LS<-min(scores)  #Lowest score

Coef.R<-(HS-LS)/(HS+LS)

cat("The coefficient of Range is",Coef.R)
```

R code Exa 7.6.1 Quartile Deviation

```
#Page.No 7.10 -7.13

library("gds") # function to find descriptive
   statistics of Grouped data

f<-c(2,8,10,6,5,9)  # frequency

Lower_limits<-c(9.5,14.5,19.5,24.5,29.5,34.5)
Upper_limits<-c(14.5,19.5,24.5,29.5,34.5,39.5)

results<-gds(Lower_limits,Upper_limits,f)

cat("The first quartile is",results$quartile1)  # 1
   st Quartile

cat("The third quartile is",results$quartile3)  # 3
   rd Quartile
```

```
#Quartile deviation
Q<-(results$quartile3-results$quartile1)/2

cat("The quartile deviation is ",Q)
```

R code Exa 7.6.2 Quartile Deviation

```
#Page.No 7.13-7.15

library("gds")   # function to find descriptive
    statistics of Grouped data

# Given data
f<-c(2,8,10,12,8,10)
Lower_limits<-c(39.5,49.5,59.5,69.5,79.5,89.5)
Upper_limits<-c(49.5,59.5,69.5,79.5,89.5,99.5)
results<-gds(Lower_limits,Upper_limits,f)

cat("The first quartile is",results$quartile1)

cat("The third quartile is",results$quartile3)

# Quartile deviation
Q<-(results$quartile3-results$quartile1)/2
cat("The quartile deviation is ",Q)
```

R code Exa 7.6.3 Quartile Deviation

```
#Page.No 7.15 -7.17

library("gds")  # function to find descriptive
    statistics of Grouped data
```

```
f<-c(6,4,6,4,8,7,5)
Lower_limits<-c(39.5,44.5,49.5,54.5,59.5,64.5,69.5)
Upper_limits<-c(44.5,49.5,54.5,59.5,64.5,69.5,74.5)
results<-gds(Lower_limits,Upper_limits,f)

cat("The first quartile is",results$quartile1)

cat("The third quartile is",results$quartile3)

# Quartile deviation
Q<-(results$quartile3-results$quartile1)/2
cat("The quartile deviation of the distibution is ",
    Q)
```

R code Exa 7.7.2 Average Deviation

```
#Page.No 7.19 -7.20

scores<-c(20,18,42,46,50,64,75)
A.D<-mean(abs(scores-mean(scores)))

cat("Average Deviation of the scores is ",A.D)

"The answer provided in the textbook is wrong"
### The first value 20 - mean(scores) is 25 not 15
```

R code Exa 7.7.3 Average Deviation

```
#Page.No 7.20

x<-c(25,45,15,16,84) # Scores
A.D<-mean(abs(x-mean(x)))
```

```
cat("Average Deviation of the distribution is ",A.D)
```

R code Exa 7.7.5 Average Deviation from Grouped data

```
#Page.No 7.22

# Given distribution table
low_value<-99.5
high_value<-169.5
step_value<-10
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
f<-c(10,8,7,6,12,9,8)
data<-rep(X_mid,f)

# Average deviation
A.D<-mean(abs(data-mean(data)))

cat("Average Deviation of the distribution is ",A.D)
```

R code Exa 7.7.6 Average Deviation from Grouped data

```
#Page.No 7.23

# Given distribution table
low_value<-69.5
high_value<-81.5
step_value<-2
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
```

```
f<-c(3,3,4,1,4,5)
data<-rep(X_mid,f)

# Average deviation
A.D<-mean(abs(data-mean(data)))

cat("Average Deviation of the distribution is ",A.D)
```

R code Exa 7.7.7 Average Deviation from Grouped data

```
#Page.No 7.24

# Given distribution table
low_value<-103.5
high_value<-128.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
f<-c(4,3,2,6,5)
data<-rep(X_mid,f)

# Average deviation
A.D<-mean(abs(data-mean(data)))

cat("Average Deviation of the distribution is ",A.D)
```

R code Exa 7.8.1 Standard Deviation

```
#Page.No 7.27 - 7.28

scores<-c(8,9,10,11,12,13,14,15)

```

```
n <- length(scores)  # number of values

std <- sd(scores)    # sample standard deviation

S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
    std deveiation

cat("The standard deviation of the scores is",S.D)

"The Answer given in the book is wrong"
```

R code Exa 7.8.2 Standard Deviation

```
#Page.No 7.28 – 7.29

x<-c(10,20,30,40,50,60,70)

n <- length(x)  # number of values

std <- sd(x)    # sample standard deviation

S.D<-sqrt((std ^ 2) * ((n - 1) / n))  # Population
    std deveiation

cat("The standard deviation of the scores is",S.D)
```

R code Exa 7.8.5 Standard Deviation from Grouped data

```
#Page.No 7.31 – 7.32

# Given distribution table
low_value<-9.5
high_value<-34.5
```

```
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
f<-c(2,8,6,2,2)
data<-rep(X_mid,f)

n <- length(data)  # number of values

std <- sd(data)    # sample standard deviation

S.D<-sqrt((std ^ 2) * ((n - 1) / n))  # Population
   std deveiation

cat("The standard deviation of the scores is",S.D)
```

R code Exa 7.8.6 Standard Deviation from Grouped data

```
#Page.No 7.32 - 7.33

# Given distribution data
low_value<-9.5
high_value<-34.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
f<-c(2,8,6,2,2)
data<-rep(X_mid,f)

n <- length(data)  # number of values
std <- sd(data)    # sample standard deviation

S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
   std deveiation
```

```
cat("The standard deviation of the scores is",S.D)
```

R code Exa 7.8.9 Effects upon Std Deviation

```
#Page.No 7.37

scores<-c(2,5,4,6,8)
Mn<-mean(scores)
n <- length(scores)  # number of values
std <- sd(scores)    # sample standard deviation
S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population Standrad deviation


#  (i)Addition of 2 in each scores
scores_new<-scores+2
Mn_new<-mean(scores_new)
n <- length(scores_new)  # number of values
std_new <- sd(scores_new)    # sample standard deviation
S.D_new<-sqrt((std_new ^ 2) * ((n - 1) / n))

cat("Previously the Mean and standard deviation respectively are",Mn ,
   "and", S.D)
cat("After addition of 2 in each score Mean has enhanced ",Mn_new-Mn,"
   and standard deviation remains unchanged i.e ",S.D-S.D_new)


#(ii)subtraction of 2 in each scores
scores_sub<-scores-2
Mn_sub<-mean(scores_sub)
```

```
n <- length(scores_sub)  # number of values
std_sub <- sd(scores_sub)    # sample standard deviation
S.D_sub<-sqrt((std_sub ^ 2) * ((n - 1) / n))

cat("Previously the Mean and standard deviation respectively are",Mn ,
    "and", S.D)
cat("After subtraction of 2 in each score Mean has reduced by ",Mn-Mn_sub,
    "and standard deviation remains unchanged i.e ", S.D-S.D_sub)


# (iii) multiplication of 2 in each scores
scores_mult<-scores*2
Mn_mult<-mean(scores_mult)
n <- length(scores_mult)  # number of values
std_mult <- sd(scores_mult)    # sample standard deviation
S.D_mult<-sqrt((std_mult ^ 2) * ((n - 1) / n))

cat("Previously the Mean and standard deviation respectively are",Mn ,
    "and", S.D)
cat("After multiplication of 2 in each score Mean has multiplied by ",2,
    "and the result is ",Mn_mult,"thats is same as ",Mn*2)
cat("standard deviation also multiplied by ",2,
    "and the result is",S.D_mult,"thats is same as", S.D*2 )

# (iii) division of 2 in each scores
scores_div<-scores/2
Mn_div<-mean(scores_div)
n <- length(scores_div)  # number of values
std_div <- sd(scores_div)    # sample standard
```

```
deviation
S.D_div<-sqrt((std_div ^ 2) * ((n - 1) / n))

cat("Previously the Mean and standard deviation respectively are",Mn ,
    "and", S.D)
cat("After division 2 in each score Mean has divided by ",
    2,"and the result is ",Mn_div,"i.e ", Mn/2)

cat("standard deviation also divided by ",2,"and the result is",S.D_div,"i.e",S.D/2)
```

R code Exa 7.8.10 Effects upon Std Deviation

```
#Page.No 7.40 – 7.41

M<-c(40,30) # Mean values
N<-c(20,30) # No.of students in a class
S.D<-c(8,4) # Std deviation of students

M.comb<-weighted.mean(M,N)

# Difference between the mean of the 1st and 2nd distribution and combined mean
d1sqr<-(40-M.comb)^2
d2sqr<-(30-M.comb)^2

sigma.comb<-sqrt(((20*(64-d1sqr))+(30*(16+d2sqr)))/50)

cat("The combined mean is",M.comb)

cat("The combined standrad deviation is ",sigma.comb)
```

Chapter 8

Testing Relationship and Associations Correlation

R code Exa 8.8.1 Spearman s Rank Difference Method

```
# Page.No 8.8 - 8.9

# Spearman's Rank Difference Method
score_in_maths<-c(77,83,85,65,50,60,70,80,78,82)
score_in_science<-c(85,60,70,50,40,45,46,44,42,47)

# Ranks of the scores
R1<-rank(score_in_maths)
R2<-rank(score_in_science)

# Correlation coefficient
P<-cor(R1,R2)
cat("There is moderate positive correlation i.e ",P)
round(P,digits = 1)
```

R code Exa 8.8.2 Spearman s Rank Difference Method

```
#Page.No 8.9 - 8.10

# Rank of students by teachers
teacher1<-c(1,2,3,4,5)
teacher2<-c(4,3,2,1,5)

r<-cor(teacher1,teacher2)

cat("Their is ",r,"or No correlation")
```

R code Exa 8.8.3 Spearman s Rank Difference Method

```
# Page.No 8.10-8.11

# Rank of the students
Rank_in_theory<-c(1,10,2,9,4,8,7,5,6,3)
Rank_in_practical<-c(10,1,4,2,9,3,7,5,6,8)

p<-cor(Rank_in_theory,Rank_in_practical)

cat("There is high negative correlation i.e ",p)
```

R code Exa 8.8.4 Spearman s Rank Difference Method

```
#Page.No 8.11

#Marks of the five students
marks_in_english<-c(78,80,81,60,90)
marks_is_science<-c(80,83,85,95,50)

# Ranks
R1<-rank(marks_in_english)
R2<-rank(marks_is_science)
```

```

P<-cor(R1,R2)

cat("There is Negative correlation i.e ",P)
```

R code Exa 8.8.5 Spearman s Rank Difference Method

```
#Page.No 8.12

# Scores in test
scores_in_oriya<-c(40,40,50,30,60,80,90)
scores_in_Geography<-c(50,60,40,40,50,80,95)

# Ranks
R1<-rank(scores_in_oriya)
R2<-rank(scores_in_Geography)

P<-cor(R1,R2)

cat("There is High positive correlation i.e ",P)
```

R code Exa 8.8.6 Spearman s Rank Difference Method

```
#Page.no 8.12 - 8.15

# Ranks of five students by examiners
R1<-c(3,2,1,4,5)
R2<-c(4,5,3,2,1)
R3<-c(5,4,2,3,1)

P12<-cor(R1,R2) # Rank correlation between examiner
    1 and 2
```

```
P23<-cor(R2,R3) # Rank correlation between examiner
    2 and 3
P13<-cor(R1,R3) # Rank correlation between examiner
    1 and 3

cat("The rank correlation between Examiners is",P12,
    P13,P23)
# above three correlations P23 have the highest co-
    relations
```

R code Exa 8.8.7 Spearman s Rank Difference Method

```
#Page.No 8.15 - 8.16

# Rank in efficiency
Rx<-c(1,2,3,4,5,6,7,8,9,10)

length_of_services_in_years<-c
    (15,12,10,11,13,10,10,18,11,13)

Ry<-rank(length_of_services_in_years)

P<-print(cor(Rx,Ry),digits=0)

#Hence there is no correlaion between the experience
  # and efficiency of employees
```

R code Exa 8.8.8 Spearman s Rank Difference Method

```
# Page.No 8.16

X<-c(57,58,33,58,60,63,71,63,74,63) # Scores on
    creativity
```

```
Y<-c(92,100,100,98,102,95,100,99,106,104)  # Scores
    of intelligence

# Ranks of scores
R1<-rank(X)
R2<-rank(Y)

P=cor(R1,R2) # Correlation

cat("Hence the correlation coefficient between
    creativity
     and intelligence scores of the student is",P)
```

R code Exa 8.8.9 Spearman s Rank Difference Method

```
#Page.no 8.18

test_scores1<-c(50,56,59,62,65)
test_scores2<-c(22,34,28,30,32)

# Ranks of the given scores
R1<-rank(test_scores1)
R2<-rank(test_scores2)

P<-cor(R1,R2)
cat("The coefficient of correlation is",P,"which is
    very low")
```

R code Exa 8.8.10 Spearman s Rank Difference Method

```
# Page.No 8.19

```

```
#Given Scores
X<-c(45,48,52,47,49,50,46,51)
Y<-c(70,72,80,75,73,76,71,71)

# Ranks of scores
R1<-rank(X)
R2<-rank(Y)

P=cor(R1,R2)

cat("Hence the correlation coefficient is",P , " i.e
    moderate")
```

R code Exa 8.8.11 Spearman s Rank Difference Method

```
#Page.N# Page.No 8.20

marks_in_maths<-c(50,26,46,25,31,22,29,30,28,18)
marks_in_g.Sci<-c(40,35,34,30,32,28,28,32,29,15)

#Ranks of the marks
R1<-rank(marks_in_maths)
R2<-rank(marks_in_g.Sci)

P=cor(R1,R2)

cat("Here the coeffient of co-ordinatore",P ,"Highly
    Positive")
```

R code Exa 8.9.1 Pearson s Product Moment Coefficient of Correlation

```
# Page.No 8.22 - 8.25

```

```
#(i) By rank difference method
marks_in_history<-c(6,8,4,5,2)
marks_in_engilsh<-c(9,6,12,10,13)

#Ranks of marks
R1<-rank(marks_in_history)
R2<-rank(marks_in_engilsh)

P<-cor(R1,R2) # Correlation

cat("The correlation is ",P,"Perfect negative correlation")


##(ii) By product moment method

r<-cor(marks_in_history,marks_in_engilsh,method = "pearson")
cat("The correlation is",r,"It is very high negative correlation")
```

R code Exa 8.9.2 Pearson s Product Moment Coefficient of Correlation

```
# Page.No 8.25 - 8.26

# Scores
X<-c(55,65,25,50,40,35,15,20,50,45)
Y<-c(50,75,45,55,50,25,35,40,70,55)

r<-cor(X,Y,method = "pearson")
cat("coefficient of correlation is ",r,"It is a high positive correlation")
```

R code Exa 8.9.3 Pearson s Product Moment Coefficient of Correlation

```
# Page.No 8.27 - 8.28

# Scores
X<-c(60,51,43,56,39,48,59,44,37,55)
Y<-c(46,35,35,42,31,34,48,31,30,44)

r<-cor(X,Y,method = "pearson")

cat("coefficient of correlation of the given set is
   ",r,
    "which is very high positive correlation")
```

R code Exa 8.9.4 Pearson s Product Moment Coefficient of Correlation

```
# Page.No 8.28 - 8.29

#Scores
X<-c(65,60,50,40,30,20,80)
Y<-c(50,55,80,90,85,79,77)

r<-cor(Y,X,method = "pearson")

cat("The product moment co-efficient of correlation
   between the above sets of scores is ",r)
"The answer provided in the text book is wrong"
```

R code Exa 8.10.1 Product Mement Method from Group Data

```
#Page.No 8.30 - 8.35

```

```
X<-c(50,41,42,45,60,43,46,47,48,50,58,45,40,35,30,20,18,15,14,10) # TEST I
Y<-c(80,90,95,96,75,70,50,60,55,50,45,45,40,30,30,25,19,18,17,10) # TEST II

# X table caclculations
low_value<-9.5
high_value<-64.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
c1<-cut(X,breaks = X_breaks)
table1<-table(c1)
data1<-data.frame(table1)

f1<-data1$Freq   # X values grouped and their frequencies

x_dash<-(X_mid-round(median(X),digits = 0))/step_value # standard values x'

fxdash<-f1*x_dash    #fx'

fxdashsqr<-fxdash*x_dash# fx'^2


# Claculation of Y variables
low_value<-9.5
high_value<-99.5
step_value<-10
Y_breaks<-seq(low_value,high_value,step_value)
Y_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)
c2<-cut(Y,breaks = Y_breaks)
table2<-table(c2)
```

```
data2<-data.frame(table2)

f2<-data2$Freq

y_dash<-round(((Y_mid-(median(Y)))/10),digits = 0) # y'

fydash<-f2*y_dash    #fy'

fydashsqr<-fydash*y_dash   #fy'^2

N<-sum(f1)

fxdashydash<-108 # from scatter diagram

# calculation of correlation coefficient

numerator<-(N*sum(fxdashydash))-(sum(fxdash)*sum(fydash))

denominator<-sqrt(((N*sum(fxdashsqr))-((sum(fxdash))^2))*((N*sum(fydashsqr))-((sum(fxdash))^2)))

r<-numerator/denominator

cat("The correlation coefficient is",r,"positive")

"The answer provided in the book is wrong"
```

R code Exa 8.10.2 Raw Score Method

```
#Page.No 8.36

#Raw score method

```

```
X<-c(10,15,18,22,25)
Y<-c(32,18,25,21,24)

r<-cor(X,Y,method="pearson")
cat("The coefficient of correlation is ",r,"low
   negative")
```

R code Exa 8.10.3 Raw Score Method

```
#Page.No 8.37

x<-c(5,10,15,20,80)
y<-c(30,37,80,90,95)

r<-cor(x,y,method="pearson")

cat("The coefficient of correlation is ",r,"moderate
    positive")
```

R code Exa 8.10.4 Raw Score Method

```
#Page.No 8.39 - 8.40

x<-c(5,4,6,8,7)
y<-c(4,6,7,8,5)

r<-cor(x,y,method="pearson")

cat("The coefficient of correlation is ",r,"moderate
    positive")
```

R code Exa 8.11.1 Biserial Correlation

```
#Page.No 8.41 - 8.43

library("ltm")

#The data
low_value<-9.5
high_value<-34.5
step_value<-5
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)

#Trained data
f1<-c(4,3,6,5,2)
data1<-rep(X_mid,f1)

#Untrained data
f2<-c(8,6,12,10,4)
data2<-rep(X_mid,f2)

# Calculation of correlation

X_mid
f<-f1+f2
d<-rep(X_mid,f)
#d<-c(rep(12,12),rep(17,9),rep(22,18),rep(27,15),rep
   (32,6))     #Mid points and their total frequency

# 0 represents trained and 1 represents untrained
# y gives trainded and untrained groups with their
   respective frequencies

y<-c(rep(0,4),rep(1,8),rep(0,3),rep(1,6),rep(0,6),
   rep(1,12),rep(0,5),rep(1,10),rep(0,2),rep(1,4))


```

```
biserial.cor(d,y)

"The answer given in the book is wrong"

# Book error
# mean of trained data
Mp<-mean(data1)
print(Mp)
"which is correct "

# Mean of Untrained data"
Mq<-mean(data2)
print(Mq)
"They mentioned it as 20.9"

# The difference will be zero hence the whole answer will become zero
Mp-Mq
```

R code Exa 8.11.2 Biserial Correlation

```
#Page.No 8.44-8.46

library("ltm")

#The data
low_value<-9.5
high_value<-79.5
step_value<-10
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)

#Trained data
f1<-c(2,2,4,7,4,3,3)
```

```r
data1<-rep(X_mid,f1)

#Untrained data
f2<-c(8,9,9,9,4,4,7)
data2<-rep(X_mid,f2)


# Calculation of correlation

f<-f1+f2
X_mid
d<-rep(X_mid,f)

# 0 represents trained and 1 represents untrained
# y gives trainded and untrained groups with their respective frequencies
y<-c(rep(0,2),rep(1,8),rep(0,2),rep(1,9),rep(0,4),rep(1,9),rep(0,7),rep(1,9),rep(0,4),rep(1,4),rep(0,3),rep(1,4),rep(0,3),rep(1,7))

rbis<-biserial.cor(d,y)

cat("Bi-serial correlation is ",rbis,"which is low positive")

"The answer may slightly vary due to roundingoff values"
```

R code Exa 8.11.3 Biserial Correlation

```r
#Page.No 8.47- 8.48

library("ltm")

#The data
```

```r
low_value<-4.5
high_value<-19.5
step_value<-3
X_breaks<-seq(low_value,high_value,step_value)
X_mid<-seq(low_value+step_value/2, high_value-step_value/2, step_value)

#Trained data
f1<-c(4,4,5,4,3)
data1<-rep(X_mid,f1)

#Untrained data
f2<-c(7,5,14,7,7)
data2<-rep(X_mid,f2)


# Calculation of correlation

f<-f1+f2 # Total frequency
X_mid  # mid point
d<-rep(X_mid,f)

# 0 represents trained and 1 represents untrained groups
# y gives trainded and untrained groups with their respective frequencies
y<-c(rep(0,4),rep(1,7),rep(0,4),rep(1,5),rep(0,5),rep(1,14),rep(0,4),rep(1,7),rep(0,3),rep(1,7))

rbis<-biserial.cor(d,y)
cat("The bi-serial correlation is",rbis,"Negligible Negative")
"The answer given in the book is wrong "

# Book Error
#negative correlation
# Mean values given differ from the actual mean values
```

```
mean(data1)    #  Mp
mean(data2)    #Mq
```

R code Exa 8.12.1 Point Biserial Correlation

```
#Page.no 8.48 - 8.50

scores<-c(35,40,32,31,20,18,12,32,10,10)

items<-c(1,1,1,1,0,0,0,1,0,0)

library("ltm") # Packge contains point biserial correlation function

r_pbis<-biserial.cor(scores,items,level = 2)

cat("Point-biserial correlation is ",r_pbis)
```

R code Exa 8.12.2 Point Biserial Correlation

```
#Page.No 8.51-8.52

scores<-c(40,80,90,17,25)
items<-c(1,1,1,0,0)

library("ltm") # Packge contains point biserial correlation function

r_pbis<-biserial.cor(scores,items,level = 2)

cat("Point-biserial correlation is ",r_pbis,"high positive")
```

R code Exa 8.13.1 Partial Correlation

```
#Page.No 8.53 - 8.57


r12<-0.9 # correlation between 1st and 2nd variable
r13<-0.6 # correlation between 1st and 3rd variable
r23<-0.2 # correlation between 2nd and 3rd variable

# (a) keeping 3rd  variable constat
r12.3<-(r12-(r13*r23))/(sqrt(1-r13^2)*sqrt(1-r23^2))

# (b) keeping 1st variable constant
r23.1<-(r23-(r12*r13))/(sqrt(1-r12^2)*sqrt(1-r13^2))

#(c)Keeping 2nd variable constant
r13.2<-(r13-(r12*r23))/(sqrt(1-r12^2)*sqrt(1-r23^2))

cat("The partial correlation coefficients keeping the
  1st 2nd 3rd variables as constant respectly are",r23.1,r13.2,r12.3)
```

R code Exa 8.13.2 Partial Correlation

```
#Page.No 8.55

r12<-15.7  # correlation between 1st and 2nd variable
r13<-0.80  # correlation between 1st and 3rd variable
r23<-4    # correlation between 2nd and 3rd variable
```

```
# (a) keeping 3rd  variable constat
r12.3<-(r12-(r13*r23))/(sqrt(1-r13^2)*-sqrt(abs(1-
   r23^2)))

# (b) keeping 1st variable constant
r23.1<-(r23-(r12*r13))/(-sqrt(abs(1-r12^2))*-sqrt(
   abs(1-r13^2)))

#(c)Keeping 2nd variable constant
r13.2<-(r13-(r12*r23))/(-sqrt(abs(1-r12^2))*-sqrt(
   abs(1-r23^2)))

cat("The partial correlation coefficients keeping
   the
     1st 2nd 3rd variables as constant respectly are"
        ,r12.3,r13.2,r23.1)
```

R code Exa 8.14.1 Tetrachoric Correlation

```
#Page.No 8.57 - 8.59

A<-30
B<-40
C<-20
D<-30

ans<-(180*sqrt(B*C)/(sqrt(A*D)+sqrt(B*C)))

cat("converting cos",ans,"to r gives correlation
   value")

rt<-0.087  # From the table value corresponding to
   cosine

```

```
cat("tetrachoric correlation is",rt,"positively
   correlated")
```

R code Exa 8.14.2 Tetrachoric Correlation

```
#Page.No 8.59

A<-50
B<-35
C<-40
D<-25

ans<-(180*sqrt(A*D)/(sqrt(B*C)+sqrt(A*D)))

cat("converting cos",ans,"to r gives correlation
   value") # From the table value

rt<-0.087  # From the table value corresponding to
   cosine
# since AD greater than BC correlation is negative

cat("tetrachoric correlation is",-rt,"Negatively
   correlated")
```

R code Exa 8.15.1 Phi Coefficient

```
#Page.No 8.60 - 8.61

library("psych") # attaching the function

t<-c(50,80,60,40) # vector of frequencies

Phi_coefficient<-phi(t,digits = 4)
```

```
cat("The phi coeffient is ",Phi_coefficient)
```

R code Exa 8.15.2 Phi Coefficient

```
#Page.No 8.61-8.62

library("psych") # attaching the function

t<-c(50,40,30,15) # vector of frequencies

Phi_coefficient<-phi(t,digits = 4)

cat("The phi coeffient is ",Phi_coefficient,"Low
   negative correlation")
```

Chapter 9

Normal Distribution

R code Exa 9.7.1 Uses of Normal Curve

```
#Page.No 9.13 − 9.14

#(a)
ans<-(pnorm(60, mean=50, sd=6)-pnorm(40, mean=50, sd
   =6))*100

cat("Percentage of cases lie between 40 and 60 is",
   ans)

"The answer may slightly vary due to rounding off
   values"
```

R code Exa 9.7.2 Uses of Normal Curve

```
# Page.no 9.14

# To compare scores on tow different tests
z1<-scale(70,center =  75,scale = 10)
```

```
z2<-scale(60,center =  50,scale = 6)

cat("z-score in mathematics is",z2,
    "which is greater than z-score in general
        science",z1)
"so his performance is better in mathematics"
```

R code Exa 9.7.3 Uses of Normal Curve

```
# Page.no. 9.15

# To determine the percentile rank of candidate in
    his group
scale(50,center = 60,scale = 6)

pnorm(50, mean=60, sd=6)*100

ans<-round(pnorm(50, mean=60, sd=6),digits = 2)*100

cat("The percentile rank of the student who secored
    50 marks
    in an achievement test is ",ans,"%")
```

R code Exa 9.7.4 Uses of Normal Curve

```
# Page.no. 9.16

#To determine percentile rank of a candidate
# To find the z- score
scale(60,center = 50,scale = 6)

#to find the percentage
ans<-round(pnorm(60, mean=50, sd=6),digits = 2)*100
```

```r

cat("The percentile rank of the student is ",ans)
```

R code Exa 9.7.5 Uses of Normal Curve

```r
#Page.No 9.16 - 9.17

#To determine the limits of the scores
# To find the z- score
Z1<-scale(60,center = 50,scale = 5)
Z2<-scale(40,center = 50,scale = 5)

#To find percentage of cases
ans<-(pnorm(60, mean=50, sd=5)-pnorm(40, mean=50, sd=5))*100

cat("Percentage of cases lie between",Z2," and",Z1," is",ans,"%")
```

R code Exa 9.7.6 Relative difficulty value

```r
#Page.No 9.17

# To determine the relative difficulty of the test items
A<--qnorm(0.30)
B<--qnorm(0.40)

cat("difficulty values of A and B respectively are ",A,B)

"Question A is relatively difficult"
```

R code Exa 9.7.7 Percentage of Cases

```
#Page.No 9.17-9.18
# To determine the percentage of cases that are
    above and below given scores

# To find the z- score
scale(55,center = 52,scale = 5)

# (a) to find the percentage of cases lie above 55

ans<-round(pnorm(55, mean=52, sd=5,lower.tail =
    FALSE)*100,digits = 2)
cat("The percentage of cases lie above 55 i.e above
    0.6sd is ",ans,"%")

# (b) to find the percentage of cases lie below 55

ans<-round(pnorm(55, mean=52, sd=5,lower.tail = TRUE
    )*100,digits = 2)
cat("The percentage of cases lie above 55 i.e below
    0.6sd is ",ans,"%")

# (C) to find the percentage of cases lie below 50
scale(50,center = 52,scale = 5)

ans<-round(pnorm(50, mean=52, sd=5,lower.tail = TRUE
    )*100,digits = 2)
cat("The percentage of cases lie above 50 i.e above
    -0.4sd is ",ans,"%")
```

R code Exa 9.7.8 Uses of Normal Curve

```
#Page.NO 9.19
# To determine the score of a candidate whose
  percentile rank is known

X<-qnorm(0.84,50,7)

cat("Thus a candidate secured",X,"marks in an
  achievement test")
```

Chapter 10

Interpretation of Scores

R code Exa 10.5.1 Standard Score

```
#Page.No 10.5

x<-50  #Marks
M<-60  # Mean
S.D<-10  # Std deviation

z<-scale(x,center = M,scale = S.D)

cat("Rama's score is ",z[1,],"  i.e 1-sigma
   distance below the mean")
```

R code Exa 10.5.2 Standard Score

```
#Page.No 10.5

x<-50  #Marks
M<-40  # Mean
S.D<-10  # Std deviation
```

```
z<-scale(x,center = M,scale = S.D)

cat("Hari's score is ",z[1,],"i.e 1sigma distance
   above the mean")
```

R code Exa 10.5.3 Standard Score

```
#Page.No 10.6

#Rama In english
z<-scale(60,center = 40,scale = 10)
cat("Rama secures ",z[1,],"in mothertougue")
"Rama's Z-score is +2sigma"

#Hari in english
z<-scale(60,center = 70,scale = 10)
cat("Hari secures ",z[1,],"in mothertougue")
"Hari's Z-score is -1sigma"

"Rama remains +2 sigma distance from mean
 Hari remains -1 sigma distance below from mean
 So Rama's performance is better"
```

R code Exa 10.6.1 T Score

```
#Page.No 10.6 - 10.7

# In english
z<-scale(60,center = 40,scale = 8)
z<-z[1,]
T_score_E<-10*z+50
cat("Rama T_score in english is ",T_score_E)
```

```
#In mother tongue
z<-scale(50,center = 50,scale = 6)
z<-z[1,]
T_score_M<-10*z+50
cat("Rama T_score in mother tongue is ",T_score_M)

"Rama performance better in English than mother
   tongue"
```

R code Exa 10.7.1 H score

```
#Page.no 10.7 - 10.8

x<-50  #Marks
M<-40  # Mean
S.D<-6  # Std deviation

z<-scale(x,center = M,scale = S.D)

z<-z[1,]

H_score<-50+14*z

cat("Rama H_score is ",H_score)
```

Chapter 11

Chi Square

R code Exa 11.4.1 Chi Square Testing

```
# Page no. : 11.5 − 11.6

condition <- c("Favourable","Unfavourable","
    Undecided")
observed <- c(20,40,21)
expected <- c(27,27,27)

DF <- data.frame(condition, observed, expected)
View(DF)

alpha5 <- 0.05   # 5% significance level
alpha1<-0.01     #1% significance level
df <- nrow(DF) - 1   # Degree of freedom

chi_sq_val_5 <- round(qchisq(alpha5,df,lower.tail =
    F), 2)   # Chi−square value
chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F
    ), 2)

diff <- DF$observed - DF$expected   # Difference
diff_sq <- diff ** 2   # Difference squared
```

```r
ans <- round(diff_sq / DF$expected, 2)   # Difference squared weighted by expected frequency

chi_sq <- sum(ans)   # Chi-square

cat("Value of chi-square is", chi_sq)

if(chi_sq > chi_sq_val_5)
  if(chi_sq > chi_sq_val_1)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and no significance")
}
```

R code Exa 11.4.2 Chi Square Testing

```r
# Page.No: 11.6 - 11.7

condition <- c("Heads","Tails")
observed <- c(50,130)
expected <- c(90,90)

DF <- data.frame(condition, observed, expected)
View(DF)

alpha5 <- 0.05   # 5% significance level
alpha1<-0.01     #1% significance level
df <- nrow(DF) - 1   # Degree of freedom

chi_sq_val_5 <- round(qchisq(alpha5,df,lower.tail = F), 2)   # Chi-square value
```

```
chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F), 2)

diff <- DF$observed - DF$expected   # Difference
diff_sq <- diff ** 2   # Difference squared
ans <- round(diff_sq / DF$expected, 2)   # Difference squared weighted by expected frequency

chi_sq <- sum(ans)   # Chi-square

cat("Value of chi-square is", chi_sq)

if(chi_sq > chi_sq_val_5)
  if(chi_sq > chi_sq_val_1)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and no significance")
}

"The answer provided in the text book is wrong"
```

R code Exa 11.4.3 Chi Square Testing

```
# Page no. : 11.7 - 11.8

condition <- c("Good","Average","Poor")
observed <- c(20,24,6)
expected <- c(8,34,8)

DF <- data.frame(condition, observed, expected)
View(DF)
```

```
alpha5 <- 0.05   # 5% significance level
alpha1<-0.01     #1% significance level
df <- nrow(DF) - 1   # Degree of freedom

chi_sq_val_5 <- round(qchisq(alpha5,df,lower.tail =
    F), 2)   # Chi−square value
chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F
    ), 2)

diff <- DF$observed - DF$expected   # Difference
diff_sq <- diff ** 2   # Difference squared
ans <- round(diff_sq / DF$expected, 2)   #
    Difference squared weighted by expected frequency

chi_sq <- sum(ans)   # Chi−square

cat("Value of chi−square is", chi_sq)

if(chi_sq > chi_sq_val_5)
  if(chi_sq > chi_sq_val_1)
{
  cat("Reject null hypothesis and it is significant"
    )
} else
{
  cat("Cannot reject null hypothesis and no
    significance")
}
```

R code Exa 11.4.4 Chi Square Testing Equal Probability Cases

```
# Page no. : 11.7 − 11.8

condition <- c("Good","Average","Bad")
```

```
observed <- c(25,60,35)
expected <- c(40,40,40)

DF <- data.frame(condition, observed, expected)
View(DF)

alpha5 <- 0.05   # 5% significance level
alpha1<-0.01     #1% significance level
df <- nrow(DF) - 1   # Degree of freedom

chi_sq_val_5 <- round(qchisq(alpha5,df,lower.tail = F), 2)   # Chi-square value
chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F), 2)

diff <- DF$observed - DF$expected   # Difference
diff_sq <- diff ** 2   # Difference squared
ans <- round(diff_sq / DF$expected, 2)   # Difference squared weighted by expected frequency

chi_sq <- sum(ans)   # Chi-square

cat("Value of chi-square is", chi_sq)

if(chi_sq > chi_sq_val_5)
  if(chi_sq > chi_sq_val_1)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and no significance")
}
```

R code Exa 11.4.7 Chi Square of Independence

```
# Page no. : 11.10 - 11.11

condition <- c("First","Second","Third")
matrix <- matrix(c(15,25,10,20,25,5),2,3,byrow = TRUE)
expected<-c((sum(matrix[,1])*50)/100,(sum(matrix[,2])*50)/100,
            (sum(matrix[,3])*50)/100,(sum(matrix[,1])*50)/100,
            (sum(matrix[,2])*50)/100,(sum(matrix[,3])*50)/100)
expected
observed<-c(15,25,10,20,25,5)
DF <- data.frame(condition, observed, expected)
View(DF)

alpha <- 0.05   # 5% significance level
df <- nrow(DF) - 1   # Degree of freedom

chi_sq_val <- round(qchisq(alpha,df,lower.tail = F), 2)   # Chi-square value

diff <- DF$observed - DF$expected   # Difference
diff_sq <- diff ** 2   # Difference squared
ans <- round(diff_sq / DF$expected, 2)   # Difference squared weighted by expected frequency

chi_sq <- sum(ans)   # Chi-square

cat("Value of chi-square is", chi_sq)


if(chi_sq > chi_sq_val)
  {
    cat("Reject null hypothesis and it is significant")
```

```
  } else
  {
    cat("Cannot reject null hypothesis and no
       significance")
  }

"Opinion of rural and Urban population differ
   significantly"
```

R code Exa 11.4.8 chi Square of 2 by 2 Table

```
#Page.no 11.11 - 11.12

R1 = c(20,10)
R2 = c(5,15)
rows   = 2

#2*2 contingency table
table = matrix(c(R1, R2),
                nrow=rows,
                byrow=TRUE)

rownames(table) = c("Good Achievement", "Bad
   achievement")# Naming the rows and
colnames(table) = c("High", "Low")   #  columns is
   optional.
table

chi_sq<-chisq.test(table,correct=FALSE)

chi_sq<-chi_sq$statistic

## Chi-squared table value @ 5%
alpha <- 0.05   # 5% significance level
df <- nrow(table) - 1   # Degree of freedom
```

```
chi_sq_val <- round(qchisq(alpha,df,lower.tail = F),
    2)   # Chi-square table value


cat("The calculated chi square value is ",chi_sq)

if(chi_sq > chi_sq_val)
{
  cat("Reject null hypothesis and observed Chi-
     square is significant")
} else
{
  cat("Cannot reject null hypothesis and there is no
      significance")
}

"so Interest has significantly affect the
   performance level of children"
```

Chapter 12

Testing the Difference Between Two Group Means

R code Exa 12.9.3.1 Students t Test Correlated Large Sample

```
#Page.No 12.9− 12.11

# A function created to find t−ratio
t.test <- function(m1,m2,s1,s2,n1,n2,r,equal.
    variance=FALSE)
{
  {
  se1 <-s1/sqrt(n1-1)
  se2<-s2/sqrt(n2-1)
  df <- n1-1
  seD<-sqrt((se1^2)+(se2^2)-2*r*se1*se2)
  }
  t <- (m1-m2)/seD
  dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
  names(dat) <- c("Difference of means", "Std Error"
     , "t", "p−value")
  return(dat)
}

```

```
Mn.P<-70   # Mean of Pre- test
Mn.E<-67   # Mean Of end - test
SD.P<-6    # Sd of pre- test
SD.E<-5.8  # sd of end- test
N.P<-30    # No.of.Observations of Pre- test
N.E<-30    # No.of.Observations of end- test
r<-.82     # correlation coefficient

t.test(Mn.P,Mn.E,SD.E,SD.P,N.P,N.E,r)

critical_t<-abs(qt(0.01/2,29))

cat("The p-value <0.01 and the table value of t @ 0.01 level of significance is ",critical_t,"which is less than the obtained t-ratio")

"Therefore the mean difference is significant @ 0.01 level of significance"
```

R code Exa 12.9.3.2 Students t Test Correlated Large Sample

```
# Page.No 12.11 - 12.13

# A function created to find t-ratio
t.test <- function(m1,m2,s1,s2,n1,n2,r,equal.variance=FALSE)
{
  {
    se1 <-s1/sqrt(n1-1)
    se2<-s2/sqrt(n2-1)
    df <- n1-1
    seD<-sqrt((se1^2)+(se2^2)-2*r*se1*se2)
  }
  t <- (m1-m2)/seD
  dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
```

```
  names(dat) <- c("Difference of means", "Std Error", "t", "p-value")
  return(dat)
}


Mn.P<-88  # Mean of Pre- test
Mn.E<-85  # Mean Of end - test
SD.P<-16  # Sd of pre- test
SD.E<-12  # sd of end- test
N.P<-64   # No.of.Observations of Pre- test
N.E<-64   # No.of.Observations of end- test
r<-.5    # correlation coefficient

t.test(Mn.P,Mn.E,SD.E,SD.P,N.P,N.E,r)

ct<-abs(qt(0.05/2,63))

cat("The p-value > 0.05 and the table value of t @ 0.05 level of significance is ",ct,"which is greater than the obtained t-ratio")
"The mean difference is significant"
```

R code Exa 12.9.3.3 Students t Test Correlated Small Sample

```
#Page.no 12.13

x<-c(40,36,35,34,40,28,31,34,37,25)
y<-c(50,42,51,46,35,52,68,51,84,63)

t<-t.test(x, y, paired = TRUE, alternative = "two.sided",conf.level = 0.99)

degrees_of_freedom<-t$parameter  # Degrees of freedom
```

```
t_ratio<-abs(t$statistic)   # t calculated value

critical_t<-abs(qt(0.01/2,degrees_of_freedom)) # t table value

p<-t$p.value           # P value

cat("The p value is",p,"which is less than 0.01")

if(t_ratio > critical_t)
  {
    cat("Reject null hypothesis and it is significant")
  } else
  {
    cat("Cannot reject null hypothesis and no significance")
  }

"Therefore, the gain from trial 1 to 5 is significance at 0.01 level of significance"
```

R code Exa 12.9.3.4 Students t Test Correlated Small Sample

```
#Page.no 12.15-12.16

x<-c(16,18,20,24,24,22,20,18,10,8,20)
y<-c(24,20,24,28,30,20,24,22,18,18,24)

t<-t.test(x, y, paired = TRUE, alternative = "two.sided",conf.level = 0.99)

degrees_of_freedom<-t$parameter  # Degrees of freedom
```

```

t_ratio<-t$statistic   # t calculated value

critical_t<-abs(qt(0.01/2,degrees_of_freedom)) # t table value

p<-t$p.value           # P value

cat("The calculated t value is",t_ratio)

if(t_ratio > critical_t)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and insignificant")
}

"Therefore, the gain from viewing the film is insignificant"
```

R code Exa 12.9.3.5 Students t Test Uncorrelated Large Sample

```
#Page.No 12.16- 12.18

# A function created to find t-ratio
t.test <- function(m1,m2,s1,s2,n1,n2,equal.variance=FALSE)
{
  {
    se1 <-s1/sqrt(n1-1)
    se2<-s2/sqrt(n2-1)
    df <- n1+n2-1
```

```
    seD<-sqrt((se1^2)+(se2^2))
  }
  t <- (m1-m2)/seD
  dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
  names(dat) <- c("Difference of means", "Std Error", "t", "p-value")
  return(dat)
}

t.test(62,57,9.7,6.8,30,25)

ct<-abs(qt(0.05/2,53))
cat("Since the p-value <0.05 and the table value of t @ 0.05 level of significance is ",ct,"which is less than the obtained t-ratio",2.199)
"Therefore the mean difference is statistically significant @ 0.05 level of significance"

#### Another Method by using the library function

library(BSDA)
result<-tsum.test(mean.x = 62,s.x = 9.7,n.x = 30,
                  mean.y = 57,s.y = 6.8,n.y = 25)

t_ratio<-result$statistic

cat("The calculated t value is",t_ratio)

critical_t<-abs(qt(0.05/2,53))

if(t_ratio > critical_t)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
```

```
}
```

R code Exa 12.9.3.6 Students t Test Uncorrelated Large Sample

```
#Page.No 12.18- 12.20

# A function created to find t-ratio
t.test <- function(m1,m2,s1,s2,n1,n2,equal.variance=
    FALSE)
{
  {
    se1 <-s1/sqrt(n1-1)
    se2<-s2/sqrt(n2-1)
    df <- n1+n2-1
    seD<-sqrt((se1^2)+(se2^2))
  }
  t <- (m1-m2)/seD
  dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
  names(dat) <- c("Difference of means", "Std Error"
      , "t", "p-value")
  return(dat)
}

t.test(100,105,10,10,50,50)
n1<-50
n2<-50
df<-n1+n2-1

ct<-abs(qt(0.05/2,df))

cat("Since the p-value <0.05 and the table value of
    t @ 0.05 level of significance is ",ct,"which is
    less than the obtained t-ratio",2.199)
"Therefore the mean difference is statistically
    significant @ 0.05 level of significance"
```

```
#### Another Method by using the library function

library(BSDA)
result<-tsum.test(mean.x = 100,s.x = 10,n.x = 50,
                  mean.y = 105,s.y = 10,n.y = 50)

t_ratio<-abs(result$statistic)

critical_t<-abs(qt(0.05/2,53))

if(t_ratio > critical_t)
{
  cat("Reject null hypothesis and it is significant"
    )
} else
{
  cat("Cannot reject null hypothesis and their is No
      significance")
}
```

R code Exa 12.9.3.7 Students t Test Uncorrelated Small Sample

```
# Page.No 12.20-12.23


boys<-c(59,43,37,25,31,36,33,41,20,18)
girls<-c(57,43,44,39,21,18,15,55,40,30)

library(BSDA)
result<-tsum.test(mean.x = mean(boys),s.x = sd(boys)
   ,n.x = length(boys),
                  mean.y = mean(girls),s.y = sd(
                     girls),n.y = length(girls))

```

```
t_ratio<-abs(result$statistic)   # T calculated value

cat("The calculated t value is",t_ratio)

df<-length(girls)+length(boys)-2 # degrees of freedom

p_value<-result$p.value     #P value

critical_t<-abs(qt(0.05/2,df)) # t table value

if(t_ratio > critical_t)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
}

"Thus their is no significant difference between the mean intelligence scores of boys and girls"

cat("p value is greater than 0.05 i.e ",p_value," null hypothesis retained")
```

R code Exa 12.9.3.8 Students t Test Uncorrelated Small Sample

```
#Page.No 12.23- 12.26

control_group<-c(11,11,10,12,8,15,10,8,10,8)
experimental_group<-c(4,4,8,9,12,15,3,13,9,9)

```

```r
library(BSDA)
result<-tsum.test(mean.x = mean(control_group),s.x = sd(control_group),n.x = length(control_group),
            mean.y = mean(experimental_group),s.y = sd(experimental_group),n.y = length(experimental_group))

t_ratio<-result$statistic   # T calculated value

cat("The calculated tvalue is ",t_ratio)

df<-length(control_group)-1 # degrees of freedom

p_value<-result$p.value     #P value

critical_t<-abs(qt(0.05/2,df)) # t table value

if(t_ratio > critical_t)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
}

"Thus their is no significant difference between the mean score of control group and experimental group"
cat("p value is greater than 0.05 i.e",p_value,"null hypothesis retained")
```

R code Exa 12.9.3.9 Students t Test Uncorrelated Small Sample

```r
# Page.No 12.26 - 12.29

GroupX<-c(26,24,18,17,18,20,18)
GroupY<-c(38,26,24,24,30,22)

library(BSDA)
result<-tsum.test(mean.x = mean(GroupX),s.x = sd(GroupX),n.x = length(GroupX),
                  mean.y = mean(GroupY),s.y = sd(GroupY),n.y = length(GroupY))

t_ratio<-abs(result$statistic)  # T calculated value

cat("The calculated t value is",t_ratio)

df<-length(GroupX)+length(GroupY)-1 # degrees of freedom

p_value<-result$p.value     #P value

critical_t<-abs(qt(0.05/2,df)) # t table value

if(t_ratio > critical_t)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
}

"Thus their is  significant difference between Group X and Group Y"
cat("p value is less than 0.05 i.e",p_value,"null hypothesis is rejected")
```

Chapter 13

Analysis of Variance

R code Exa 13.8.1 One Way Anova

```
#Page.No 13.7 – 13.9

lecture<-c(10,12,13,13,14)
seminar<-c(13,15,15,17,18)
discussion<-c(7,7,10,11,12)

combined_groups<-data.frame(cbind(lecture,seminar,
   discussion))
stackedgroup<-stack(combined_groups)
anova_results<-aov(values~ind,data = stackedgroup)
summary(anova_results)

"The ANOVA shows that there is significant
   differences among the three methods of
   instructions"
```

R code Exa 13.8.2 Deviation Score Method

```
#Page.No 13.10 - 13.12

lecture<-c(10,12,13,13,14)
seminar<-c(13,15,15,17,18)
discussion<-c(7,7,10,11,12)

combined_groups<-data.frame(cbind(lecture,seminar,
    discussion))
stackedgroup<-stack(combined_groups)
anova_results<-aov(values~ind,data = stackedgroup)
summary(anova_results)

"The ANOVA shows that there is significant
    differences among the three methods of
    instructions"
```

R code Exa 13.8.3 Anova

```
#Page.No 13.7 - 13.9

Group1<-c(6,7,9,10,8)
Group2<-c(11,10,8,12,9)

combined_groups<-data.frame(cbind(Group1,Group2))
stackedgroup<-stack(combined_groups)
anova_results<-aov(values~ind,data = stackedgroup)
summary(anova_results)

"The ANOVA shows that there is significant
    differences among the three methods of
    instructions"

# To find the t-value
library(BSDA)

```

```
result<-tsum.test(mean.x = mean(Group1),s.x = sd(
   Group1),n.x = length(Group1),
                  mean.y = mean(Group2),s.y = sd(
                     Group2),n.y = length(Group2))

t<-abs(result$statistic)

t^2 ->F # From the anova_results F= 4
```

R code Exa 13.8.4 Two Way ANOVA

```
#Page.No 13.14 – 13.17

specialist<-c
   (1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,2,2,2,2,2)

performance<-c
   (5,4,3,2,1,3,4,8,5,6,4,5,5,6,5,4,6,8,5,6,8,9,10,6,7,6,8,5,9,8)

activity<-c
   (1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3)

activity<-as.factor(activity)
specialist<-as.factor(specialist)
combined_groups<-data.frame(cbind(activity,
   specialist,performance))

# Two– way anova
anova_results<-aov(performance~ specialist+activity+
   specialist*activity) # anova model specification
summary(anova_results)

# Interpretation from anova table p_value

" It can attributed that the co–curricular
```

```
    activities have differ significantly in their
    relation with selection of the specialist"

# Using F value
# Interpretation on Specialist

Fcalculated<-1.716

# at 0.05 level of significance

Ftable_0.05<-qf(.95, df1=1, df2=24)

if(Fcalculated > Ftable_0.05)
{
  cat("Reject null hypothesis and it is significant"
    )
} else
{
  cat("Cannot reject null hypothesis and their is No
      significance")
}

# At 0.01 level of significance

Ftable_0.01<-qf(.99, df1=2, df2=24)

if(Fcalculated > Ftable_0.01)
{
  cat("Reject null hypothesis and it is significant"
    )
} else
{
  cat("Cannot reject null hypothesis and their is No
      significance")
}

### interpretation on activities

```

```
Fcalculated<-13.319

Ftable_0.05<-qf(.95, df1=2, df2=24)

if(Fcalculated > Ftable_0.05)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
}

## Interpretation on interaction effect

# at 0.05 level of significance

Fcalculated<-2.397

Ftable_0.05<-qf(.95, df1=2, df2=24)

if(Fcalculated > Ftable_0.05)
{
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
}

# at 0.01 level of significance

Ftable_0.01<-qf(.99, df1=2, df2=24)

if(Fcalculated > Ftable_0.01)
{
```

```
  cat("Reject null hypothesis and it is significant")
} else
{
  cat("Cannot reject null hypothesis and their is No significance")
}
```

Chapter 14

Analysis of Covariance

R code Exa 14.4.1 ANCOVA

```
# Page.No 14.6 − 14.12

X<-c(5,6,3,2,4,4,8,5,6,2,6,5,6,2,1)

Y<-c(6,7,5,4,3,8,7,6,5,4,4,6,5,2,3)

#ANOVA table for X
X1<-c(5,6,3,2,4)
X2<-c(4,8,5,6,2)
X3<-c(6,5,6,2,1)
combined_groups<-data.frame(cbind(X1,X2,X3))
stackedgroup<-stack(combined_groups)
anova_results_X<-aov(values~ind,data = stackedgroup)
summary(anova_results_X)

# ANOVA table for Y
Y1<-c(6,7,5,4,3)
Y2<-c(8,7,6,5,4)
Y3<-c(4,6,5,2,3)
combined_groups<-data.frame(cbind(Y1,Y2,Y3))
stackedgroup<-stack(combined_groups)
```

```
anova_results_Y<-aov(values~ind,data = stackedgroup)
summary(anova_results_Y)

table_value_1<-qf(.99,2,12)
table_value_5<-qf(.95,2,12)

## From the calculated F value and critical value @ 0.05 and 0.01
## There is no difference among covariates of X

##### calculation of anova table by another method


sum(X)
sum(Y)
sum(X^2)
sum(Y^2)
sum(X*Y)

N<-length(X1)
sum_of_N<-length(X)

#Step 1  correction
Cx<-(sum(X))^2/sum_of_N
Cy<-(sum(Y)^2)/sum_of_N
Cxy<-sum(X)*sum(Y)/sum_of_N

#Step 2 TotalSS
TSSx<-sum(X^2)-Cx
TSSy<-sum(Y^2)-Cy
TSSxy<-sum(X*Y)-Cxy  # Cx<-(sum(X))^2/N

#step 3 Between means
Bet_Mn_X<-(sum(X1)^2+sum(X2)^2+sum(X3)^2)/N-Cx
Bet_Mn_Y<-(sum(Y1)^2+sum(Y2)^2+sum(Y3)^2)/N-Cy
Bet_Mn_XY<-(sum(X1)*sum(Y1)+sum(X2)*sum(Y2)+sum(X3)*sum(Y3))/N-Cxy

```

```
# step 4
# Within Group SS
WSS_x<-TSSx-Bet_Mn_X
WSS_Y<-TSSy-Bet_Mn_Y
WSS_XY<-TSSxy-Bet_Mn_XY

# Degrees of freedom
Bdf<-3-1  # 3 groups
Tdf<-length(X)-1
Wdf<-Tdf-Bdf
df<-c(Bdf,Wdf,Tdf)

#sum of squares
SSx<-c(Bet_Mn_X,WSS_x,TSSx)
SSy<-c(Bet_Mn_Y,WSS_Y,TSSy)

# Mean sum of squares
MSx<-SSx/df
MSy<-SSy/df

Anova_table<-matrix(c(df,SSx,SSy,MSx,MSy),3,5)

# step 6 computing adjusted sum of squares for Y

TSSy.x<-TSSy-(TSSxy^2)/TSSx
WSSy.x<-WSS_Y-(WSS_XY^2)/WSS_x
BSSy.x<-TSSy.x-WSSy.x

SSy.x<-c(BSSy.x,WSSy.x,TSSy.x)
SSx.y<-c(BSSy.x,WSS_XY,TSSxy)

df<-c(Bdf,Wdf-1,Tdf-1)
MSy.x<-SSy.x/df

#Step 7 ANCOVA

matrix(c(df,SSx,SSy,SSx.y,SSy.x,MSy.x),3,6) # Ancova table
```

```
Fy.x<-2.94/1.72  # from the ANCOVA table

Ftable<-qf(.95,2,13)

# Since F calculated value less than F table value Accept null hypothesis
# The computation of adjusted SS for y is not significant @ 0.05

SDy.x<-sqrt(WSSy.x/11)
#From comparison both are not significant at 0.05 level

## Step 8 Adjusted  y means

Mx<-c(mean(X1),mean(X2),mean(X3))
My<-c(mean(Y1),mean(Y2),mean(Y3))
GMx<-mean(Mx)
r_total<-cor(X,Y)
r_between<-Bet_Mn_XY/sqrt(Bet_Mn_X*Bet_Mn_Y)
r_within<-WSS_XY/sqrt(WSS_x*WSS_Y)
bx<-WSS_XY/WSS_x

M1<-mean(Y1)-bx*(mean(X1)-GMx)
M2<-mean(Y2)-bx*(mean(X2)-GMx)
M3<-mean(Y3)-bx*(mean(X3)-GMx)

My.x<-c(M1,M2,M3)

Table<- matrix(c(Mx,My,My.x),3,3)
print(Table)

# Step 9 Comparison of Adjusted Means

SDy.x
SEmd<-SDy.x*sqrt(1/length(X1)+1/length(X2))

```

```
# t_value

(M1-M2)/SEmd

(M1-M3)/SEmd

(M2-M3)/SEmd

# All pairs are not significant
cat("There is high correlation",r_between,r_within,r_total)

"hence Group 1 , Group 2 , Group 3 not differ significantly "
```