

R Textbook Companion for
Statistics in Education and Psychology
by P. C. Dash and Bhabagrahi Biswal¹

Created by

Nivedha R

M.Sc.

Information Technology

Tamil Nadu Agricultural University

Cross-Checked by

R TBC Team

August 28, 2020

¹Funded 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	4
2 Some Fundamentals of Statistics in Education and Psychology	5
3 Frequency Distribution	7
4 Graphic Presentation	9
5 Measures of Central Tendency	12
6 Percentile and Percentile Rank	34
7 Measures of Dispersion	43
8 Testing Relationship and Associations Correlation	56
9 Normal Distribution	76
10 Interpretation of Scores	81
11 Chi Square	84
12 Testing the Difference Between Two Group Means	92
13 Analysis of Variance	103
14 Analysis of Covariance	109

List of R Codes

Exa 2.6.1	Class Interval	5
Exa 3.3.1	Procedure	7
Exa 4.7.1	Histogram	9
Exa 4.8.1	Frequency Polygon	10
Exa 4.9.1	Ogive	10
Exa 4.10.1	Pie Diagram	11
Exa 5.3.1	Mean	12
Exa 5.3.2	Mean	12
Exa 5.3.3	Mean	13
Exa 5.3.4	Calculation of Mean from Grouped Data	13
Exa 5.3.5	Calculation of Mean from Grouped Data	13
Exa 5.3.6	Calculation of Mean from Grouped Data	14
Exa 5.3.7	Combined Mean	14
Exa 5.3.8	Combined Mean	15
Exa 5.3.9	Combined Mean	15
Exa 5.3.10	Properties of Mean	15
Exa 5.3.11	Properties of Mean	16
Exa 5.3.12	Properties of Mean	16
Exa 5.4.1	Calculation of Median	16
Exa 5.4.2	Calculation of Median	17
Exa 5.4.3	Calculation of Median	17
Exa 5.4.7	Calculation of Median from Grouped Data	18
Exa 5.4.8	Calculation of Median from Grouped Data	18
Exa 5.4.9	Calculation of Median from Grouped Data	18
Exa 5.5.1	Mode	19
Exa 5.5.2	Mode	20
Exa 5.5.3	Mode	20
Exa 5.5.4	Mode	21

Exa 5.5.6	Bi modal	21
Exa 5.5.7	Calculation of Mode from Grouped Data	22
Exa 5.5.8	Bi modal	23
Exa 5.5.9	Calculation of Mode from Grouped Data	24
Exa 5.5.10	Calculation of Mode from Grouped Data	24
Exa 5.5.11	Calculation of Mode from Grouped Data	27
Exa 5.7.1	Calculation of Mean Median Mode from Grouped Data	29
Exa 5.7.2	Calculation of Mean Median Mode from Grouped Data	31
Exa 6.2.1	Percentile	34
Exa 6.3.1	Percentile rank	36
Exa 6.3.2	Percentile rank	36
Exa 6.3.3	Percentile rank	37
Exa 6.3.4	Percentile rank	38
Exa 6.4.1	Percentile Problems	39
Exa 6.5.1	Graphical Representation of Percentile	42
Exa 7.5.1	Range	43
Exa 7.5.2	Range	43
Exa 7.5.3	Coefficient of Range	43
Exa 7.5.4	Coefficient of Range for Grouped Data	44
Exa 7.5.5	Coefficient of Range for Grouped Data	44
Exa 7.5.6	Coefficient of Range for Grouped Data	45
Exa 7.6.1	Quartile Deviation	45
Exa 7.6.2	Quartile Deviation	46
Exa 7.6.3	Quartile Deviation	46
Exa 7.7.2	Average Deviation	47
Exa 7.7.3	Average Deviation	47
Exa 7.7.5	Average Deviation from Grouped data	48
Exa 7.7.6	Average Deviation from Grouped data	48
Exa 7.7.7	Average Deviation from Grouped data	49
Exa 7.8.1	Standard Deviation	49
Exa 7.8.2	Standard Deviation	50
Exa 7.8.5	Standard Deviation from Grouped data	50
Exa 7.8.6	Standard Deviation from Grouped data	51
Exa 7.8.9	Effects upon Std Deviation	52
Exa 7.8.10	Effects upon Std Deviation	54
Exa 8.8.1	Spearman s Rank Difference Method	56
Exa 8.8.2	Spearman s Rank Difference Method	56
Exa 8.8.3	Spearman s Rank Difference Method	57

Exa 8.8.4	Spearman s Rank Difference Method	57
Exa 8.8.5	Spearman s Rank Difference Method	58
Exa 8.8.6	Spearman s Rank Difference Method	58
Exa 8.8.7	Spearman s Rank Difference Method	59
Exa 8.8.8	Spearman s Rank Difference Method	59
Exa 8.8.9	Spearman s Rank Difference Method	60
Exa 8.8.10	Spearman s Rank Difference Method	60
Exa 8.8.11	Spearman s Rank Difference Method	61
Exa 8.9.1	Pearson s Product Moment Coefficient of Correlation .	61
Exa 8.9.2	Pearson s Product Moment Coefficient of Correlation .	62
Exa 8.9.3	Pearson s Product Moment Coefficient of Correlation .	63
Exa 8.9.4	Pearson s Product Moment Coefficient of Correlation .	63
Exa 8.10.1	Product Mement Method from Group Data	63
Exa 8.10.2	Raw Score Method	65
Exa 8.10.3	Raw Score Method	66
Exa 8.10.4	Raw Score Method	66
Exa 8.11.1	Biserial Correlation	67
Exa 8.11.2	Biserial Correlation	68
Exa 8.11.3	Biserial Correlation	69
Exa 8.12.1	Point Biserial Correlation	71
Exa 8.12.2	Point Biserial Correlation	71
Exa 8.13.1	Partial Correlation	72
Exa 8.13.2	Partial Correlation	72
Exa 8.14.1	Tetrachoric Correlation	73
Exa 8.14.2	Tetrachoric Correlation	74
Exa 8.15.1	Phi Coefficient	74
Exa 8.15.2	Phi Coefficient	75
Exa 9.7.1	Uses of Normal Curve	76
Exa 9.7.2	Uses of Normal Curve	76
Exa 9.7.3	Uses of Normal Curve	77
Exa 9.7.4	Uses of Normal Curve	77
Exa 9.7.5	Uses of Normal Curve	78
Exa 9.7.6	Relative difficulty value	78
Exa 9.7.7	Percentage of Cases	79
Exa 9.7.8	Uses of Normal Curve	79
Exa 10.5.1	Standard Score	81
Exa 10.5.2	Standard Score	81
Exa 10.5.3	Standard Score	82

Exa 10.6.1 T Score	82
Exa 10.7.1 H score	83
Exa 11.4.1 Chi Square Testing	84
Exa 11.4.2 Chi Square Testing	85
Exa 11.4.3 Chi Square Testing	86
Exa 11.4.4 Chi Square Testing Equal Probability Cases	87
Exa 11.4.7 Chi Square of Independence	88
Exa 11.4.8 chi Square of 2 by 2 Table	90
Exa 12.9.3.1 Students t Test Correlated Large Sample	92
Exa 12.9.3.2 Students t Test Correlated Large Sample	93
Exa 12.9.3.3 Students t Test Correlated Small Sample	94
Exa 12.9.3.4 Students t Test Correlated Small Sample	95
Exa 12.9.3.5 Students t Test Uncorrelated Large Sample	96
Exa 12.9.3.6 Students t Test Uncorrelated Large Sample	98
Exa 12.9.3.7 Students t Test Uncorrelated Small Sample	99
Exa 12.9.3.8 Students t Test Uncorrelated Small Sample	100
Exa 12.9.3.9 Students t Test Uncorrelated Small Sample	101
Exa 13.8.1 One Way Anova	103
Exa 13.8.2 Deviation Score Method	103
Exa 13.8.3 Anova	104
Exa 13.8.4 Two Way ANOVA	105
Exa 14.4.1 ANCOVA	109

Chapter 2

Some Fundamentals of Statistics in Education and Psychology

R code Exa 2.6.1 Class Interval

```
1 # Page.No 2.5
2
3 #(a) class limit
4 lower_limit<-10
5 upper_limit<-20
6
7 lower_value<-lower_limit-0.5
8 upper_value<-upper_limit+0.5
9
10 cat("The lower and upper values of the class
      interval 10-20 is",lower_value,upper_value)
11
12 #(b) class midpoint
13
14 midpoint<-lower_limit+(upper_limit-lower_limit)/2
15
16 cat("mid point of a class interval 10-20 is",
```

midpoint)

Chapter 3

Frequency Distribution

R code Exa 3.3.1 Procedure

```
1 # Page.No 3.4
2
3 #(i) Determining Range
4 scores<-c
      (10,15,16,17,18,20,22,25,30,35,40,42,45,46,47,48,49,47,48,49,50,1
5 HS<-max(scores)
6 LS<-min(scores)
7 range<-HS-LS
8
9 cat("The range of the frequency distribution is ",
      range)
10
11 #(ii) Determining size of class interval
12
13 no.of.classes<-8
14 size<-range/no.of.classes
15
16 cat("The size of the class intervals is ",size)
17
18 #(vii) frequency table
```

```
19
20 low_value<-9.5
21 high_value<-54.5
22 step_value<-5
23 X_breaks<-seq(low_value,high_value,step_value)
24 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
25 c<-cut(scores,breaks = X_breaks)
26 table<-table(c)
27 data<-data.frame(table)
28 print(data)
29
30 N<-sum(data$Freq)
31 cat("Total no of frequencies is",N)
```

Chapter 4

Graphic Presentation

R code Exa 4.7.1 Histogram

```
1 # Page.No 4.8
2
3 # Given data
4 low_value<-9.5
5 high_value<-34.5
6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
9 f<-c(5,4,6,4,2)
10 data<-rep(X_mid,f)
11
12 # Histogram
13 hist(data,main="Histogram",
14       xlab="class_limits",
15       border="black",
16       col="Steelblue3",
17       xlim=c(0,40),
18       las=1,
19       breaks=5)
```

R code Exa 4.8.1 Frequency Polygon

```
1 # Page.No 4.9
2
3 # Given data table
4 low_value<-4.5
5 high_value<-39.5
6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_breaks
9 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
10 X_mid
11 f<-c(0,2,3,6,4,5,0)
12 data<-rep(X_mid,f)
13
14 # Plottinf the frequency polugon
15 plot(X_mid,f,type = "line",main = "title",xlab = "
  Mid Points",ylab = "Frequencies",col = "red")
16
17 ## Another method
18 hist<-(hist(data,main ="Histogram",xlab=" class _
  limits",border="black",col="Steelblue3",xlim=c
  (0,50),las=1,breaks=5))
19 library("agricolae")
20 polygon.freq(hist,f=1,col="red")
```

R code Exa 4.9.1 Ogive

```
1 # Page.No 4.11
2
3 x<-seq(9.5,34.5,5) # class interval
```

```
4 f<-c(0,5,4,6,3,2) # frequency
5 cf<-cumsum(f) # cumulative frequency
6 percent<-cf/sum(f)*100
7 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

```
1 # Page.No 4.12
2
3 numeric<-c(50,25,20,5)
4 degree<-numeric*360/100
5
6 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

```
1 #Page.No 5.5
2
3 marks<-c(70,30,20,90,40) # Marks in history
4
5 avg<-mean(marks) # Mean of history marks
6
7 cat("The average marks of the students is",avg)
```

R code Exa 5.3.2 Mean

```
1 #Page.No 5.5
2
3 marks<-c(10,20,30,40,50,60,70,80,90) # marks in
  English
4 avg<-mean(marks) # average of marks
5 cat("The average marks of the students is",avg)
```

R code Exa 5.3.3 Mean

```
1 #Page.No 5.6
2
3 scores<-c(80,90,95,96,85,84,82) # marks in
  mathematics
4 avg<-mean(scores) # average of marks
5 cat("The average marks of the students is",avg)
```

R code Exa 5.3.4 Calculation of Mean from Grouped Data

```
1 #Page.No 5.7
2
3 # Given data table
4 low_value<-9.5
5 high_value<-34.5
6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
9 f<-c(5,4,6,3,2) # frequency of the distribution
10 data<-rep(X_mid,f)
11 mean(data)
12
13 # to create a table from raw data
14 x<-cut(data, breaks=X_breaks)
15 datatable<-table(x)
16 df<-data.frame(datatable)
17 df$midpnt<-X_mid
18 df
```

R code Exa 5.3.5 Calculation of Mean from Grouped Data

```

1 #Page.No 5.9
2
3 Lower_limits<-c(5,11,17,23,29)
4 Upper_limits<-c(10,15,22,28,35)
5 frequency<-c(1,4,3,7,5)
6 library("gds") # function for Descriptive
   statistics of grouped data
7 results<-gds(Lower_limits,Upper_limits,frequency)
8 cat("the mean of the distribution is",results$mean)

```

R code Exa 5.3.6 Calculation of Mean from Grouped Data

```

1 #Page.No 5.12 -5.14
2
3 Lower_limits<-c(10,15,20,25,30)
4 Upper_limits<-c(14,19,24,29,35)
5 frequency<-c(5,10,7,2,1)
6 library("gds") # Function for Descriptive
   statistics of grouped data
7 results<-gds(Lower_limits,Upper_limits,frequency)
8 cat("the mean of the distribution is",results$mean)

```

R code Exa 5.3.7 Combined Mean

```

1 #Page.No 5.14
2
3 mean<-c(50,45,40,45) # mean scores in history
4 no.of.students<-c(20,20,15,30) # no.of students in
   section
5 wm<-weighted.mean(mean,no.of.students)
6 cat("Combined mean is",wm)

```

R code Exa 5.3.8 Combined Mean

```
1 #Page.No 5.15
2
3 No.of.students<-c(12,8,10,15,7)
4 Mean<-c(50,70,30,45,48) # Mean in English test
5 weighted.mean(Mean,No.of.students)
```

R code Exa 5.3.9 Combined Mean

```
1 #Page.No 5.15
2
3 no.of.students<-c(8,12,10,14,9)
4 Mean_of_achievement_test<-c(55,78,80,50,87)
5 M_comb<-weighted.mean(Mean_of_achievement_test,no.of
  .students)
6 cat("combined mean of the achievement test of
  students of 5 classes is",M_comb)
```

R code Exa 5.3.10 Properties of Mean

```
1 #Page.No 5.16
2
3 X<-c(8,5,6,4,7)
4 d<-X-mean(X)
5 cat("The sum of deviations is",sum(d))
```

R code Exa 5.3.11 Properties of Mean

```
1 #Page.No 5.17
2
3 X<-c(10,12,18,16,14)
4 d<-X-mean(X)
5 sum_of_deviations<-sum(d)
6 sum_of_dsquares<-sum(d^2)
7 cat("The Sum of squares of deviations is",sum_of_
      dsquares)# which is greater than sum of squares
8 mean(X)
```

R code Exa 5.3.12 Properties of Mean

```
1 #Page.No 5.17
2
3 #If each score is added or subtracted by a constant
  quantity
4 #The mean will increased by same quantity
5 X<-c(1,2,3,4,5,6,7)
6 mean(X)
7
8 # Addition of constant 5
9 X_add<-X+5
10 mean(X_add)
11 mean(X)+5 # the same constant addition
12
13 # subtraction of constant 5
14 X_sub<-X-5
15 mean(X_sub)
16 mean(X)-5 # the same constant subtraction
```

R code Exa 5.4.1 Calculation of Median

```
1 # Page.No 5.24
2
3 scores<-c(10,15,16,14,17,20,21)
4 md<-median(scores)
5 cat("The median is",md)
```

R code Exa 5.4.2 Calculation of Median

```
1 #Page.No 5.25
2
3 scores<-c(15,10,14,16)
4 Md<-median(scores)
5 cat("The median of the scores is",Md)
```

R code Exa 5.4.3 Calculation of Median

```
1 #Page.No 5.25
2
3 # (i)
4 scores<-c(10,8,25,24,30,45)
5 median(scores)
6
7 #(ii)
8 Scores<-c(40,48,43,42,41)
9 median(Scores)
10
11 #(iii)
12 score<-c(11,13,15,15,15,18,21)
13 median(score)
14
15 lower_value<-median(score)-0.5 # from the lower
    limit
16 corrected_median<-lower_value+2/3
```

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

R code Exa 5.4.7 Calculation of Median from Grouped Data

```
1 #Page.No 5.30
2
3 Lower_limits<-c(10,15,20,25,30)
4 Upper_limits<-c(14,19,24,29,34)
5 frequency<-c(5,4,6,3,2)
6 library("gds")
7 results<-gds(Lower_limits,Upper_limits,frequency)
8 cat("the median of the distribution is",results$
    median)
```

R code Exa 5.4.8 Calculation of Median from Grouped Data

```
1 #Page.No 5.32 – 5.33
2
3 library("gds") #function for Descriptive statistics
  of grouped data
4 lower_limits<-c(10,15,20,25,30)
5 upper_limits<-c(14,19,24,29,34)
6 frequency<-c(5,4,6,3,2)
7 md<-gds(lower_limits,upper_limits,frequency)
8 cat("The median is",md$median)
```

R code Exa 5.4.9 Calculation of Median from Grouped Data

```
1 #Page.No 5.33
2
```

```

3 #(i) and (ii)
4 Lower_limits<-c(40,50,60,70,80,90)
5 Upper_limits<-c(49,59,69,79,89,99)
6 frequency<-c(14,12,15,10,15,14)
7
8 library("gds") # Function to find Descriptive
   statistics of grouped data
9 results<-gds(Lower_limits,Upper_limits,frequency)
10 cat("the median of the distribution is",results$
   median)
11 #The answer may slightly vary due to rounding off
   values

```

R code Exa 5.5.1 Mode

```

1 #Page.No 5.42
2
3 # Create the function.
4 mode <- function(x) {
5   unique_val <- unique(x)
6   counts <- vector()
7   for (i in 1:length(unique_val)) {
8     counts[i] <- length(which(x==unique_val[i]))
9   }
10  position <- c(which(counts==max(counts)))
11  if (mean(counts)==max(counts))
12    mode_x <- 'Mode does not exist'
13  else
14    mode_x <- unique_val[position]
15  return(mode_x)
16 }
17
18 # Create the vector with numbers.
19 v <- c(10,20,10,25,10,20,10,22,28)
20

```

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

R code Exa 5.5.2 Mode

```
1 #Page.No 5.43
2
3 x<-c(30,31,32,33,34,35)#size of the garments
4 f<-c(20,30,40,60,25,10) # no of persons wearing
5 data<-rep(x,f)
6
7 # Creating function to find mode
8 ## Or u can use the previous functions too
9 Mode <- function(x) {
10   ux <- unique(x)
11   ux[which.max(tabulate(match(x, ux)))]
12 }
13
14 md<-Mode(data)
15
16 cat ("size of the dress most frequently used in
      season is",md)
```

R code Exa 5.5.3 Mode

```
1 #Page.No 5.43
2
3 # Create the function.
4 getmode <- function(v) {
5   uniqv <- unique(v)
6   uniqv[which.max(tabulate(match(v, uniqv)))]
7 }
```

```
8
9 # Create the vector with numbers.
10 Scores <- c(20,20,30,20,25,20,38,37,38)
11
12 # Calculate the mode using the user function.
13 result <- getmode(Scores)
14 cat("The mode is",result)
```

R code Exa 5.5.4 Mode

```
1 #Page.No 5.43
2
3 # Create the function.
4 getmode <- function(v) {
5   uniqv <- unique(v)
6   uniqv[which.max(tabulate(match(v, uniqv)))]
7 }
8
9 # Create the vector with numbers.
10 scores <- c(20,25,20,15,19,22,20,22)
11
12 # Calculate the mode using the user function.
13 result <- getmode(scores)
14 cat("The mode is",result)
```

R code Exa 5.5.6 Bi modal

```
1 #Page.No 5.43
2
3 Scores <- c(20,25,20,28,29,28,30,32)
4
5 #Creating a function
6 mode <- function(x) {
```

```

7  unique_val <- unique(x)
8  counts <- vector()
9  for (i in 1:length(unique_val)) {
10   counts[i] <- length(which(x==unique_val[i]))
11 }
12 position <- c(which(counts==max(counts)))
13 if (mean(counts)==max(counts))
14   mode_x <- 'Mode does not exist'
15 else
16   mode_x <- unique_val[position]
17 return(mode_x)
18 }
19
20 mode(Scores)
21 # the series is Bi-modal

```

R code Exa 5.5.7 Calculation of Mode from Grouped Data

```

1 # Page.no 5.44
2
3 ##Creating a function
4 mode <- function(x) {
5   unique_val <- unique(x)
6   counts <- vector()
7   for (i in 1:length(unique_val)) {
8     counts[i] <- length(which(x==unique_val[i]))
9   }
10  position <- c(which(counts==max(counts)))
11  if (mean(counts)==max(counts))
12    mode_x <- 'Mode does not exist'
13  else
14    mode_x <- unique_val[position]
15  return(mode_x)
16 }
17

```

```

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

19 Md<-mode(y)
20
21 ## Since 13 and 14 are adjacent numbers their mean
    can be considered as the mode
22 Mode<-mean(Md)
23 cat("The mode will be", Mode)

```

R code Exa 5.5.8 Bi modal

```

1 # Page.no 5.44
2
3 ##Creating a function
4 mode <- function(x) {
5   unique_val <- unique(x)
6   counts <- vector()
7   for (i in 1:length(unique_val)) {
8     counts[i] <- length(which(x==unique_val[i]))
9   }
10  position <- c(which(counts==max(counts)))
11  if (mean(counts)==max(counts))
12    mode_x <- 'Mode does not exist'
13  else
14    mode_x <- unique_val[position]
15  return(mode_x)
16 }
17
18 y<-c
    (11,11,12,12,12,13,13,13,13,13,14,14,14,14,15,15,15,15,16,16,16,17,17,17,17,18,18,18,18,18)

19 Md<-mode(y)
20
21 # value 13 occurs 5 times which is greater than the

```

```

    frequency of the adjacent values
22 # Also 15 occurs 4 times which is also greater than
    the frequency of the adjacent values
23
24 # hence this is Bi-modal
25 cat("The mode values are",13 ,15)

```

R code Exa 5.5.9 Calculation of Mode from Grouped Data

```

1 #Page.No 5.45
2
3 # Given frequency distribution data
4 low_value<-9.5
5 high_value<-39.5
6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_breaks
9 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
10 X_mid
11 f<-c(2,3,2,8,3,2)
12 data<-rep(X_mid,f)
13
14 # creating function to find mode
15 Mode <- function(x) {
16   ux <- unique(x)
17   ux[which.max(tabulate(match(x, ux)))]
18 }
19 Mode(data)

```

R code Exa 5.5.10 Calculation of Mode from Grouped Data

```

1 #Page.No 5.45

```

```

2
3 # A function created to find the median of a grouped
  data
4 GroupedMedian <- function(frequencies, intervals,
  sep = NULL, trim = NULL) {
5   # If "sep" is specified, the function will try to
  create the
6   #   required "intervals" matrix. "trim" removes
  any unwanted
7   #   characters before attempting to convert the
  ranges to numeric.
8   if (!is.null(sep)) {
9     if (is.null(trim)) pattern <- ""
10    else if (trim == "cut") pattern <- "
  \\[[\\]]|\\(|\\)"
11    else pattern <- trim
12    intervals <- sapply(strsplit(gsub(pattern, "",
  intervals), sep), as.numeric)
13  }
14
15  Midpoints <- rowMeans(intervals)
16  cf <- cumsum(frequencies)
17  Midrow <- findInterval(max(cf)/2, cf) + 1
18  L <- intervals[1, Midrow]      # lower class
  boundary of median class
19  h <- diff(intervals[, Midrow]) # size of median
  class
20  f <- frequencies[Midrow]      # frequency of
  median class
21  cf2 <- cf[Midrow - 1]         # cumulative
  frequency class before median class
22  n_2 <- max(cf)/2              # total
  observations divided by 2
23
24  unname(L + (n_2 - cf2)/f * h)
25 }
26
27 # The Problem

```

```

28 frequency<-c(5,4,6,3,2)
29 colnames<-c("numbers")
30 rownames<-c("[9.5,14.5]", "(14.5,19.5]", "(19.5,24.5]", "(24.5,29.5]", "(29.5,34.5]")
31 #The class intervals are changed to exclusive one by
    adding 0.5 to the upper limits and deducing 0.5
    from the lower limits
32 y<-matrix(frequency,nrow=length(frequency),dimnames=
    list(rownames,colnames))
33 md<-GroupedMedian(y[, "numbers"], rownames(y), sep="
    ", trim="cut")
34
35
36 ### find the mean for grouped data
37 low_value<-9.5
38 high_value<-34.5
39 step_value<-5
40 X_breaks<-seq(low_value,high_value,step_value)
41 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
42 f<-c(5,4,6,3,2)
43 data<-rep(X_mid,f)
44 Mn<-mean(data)
45
46 ## to find mode from mean and median
47 Mode<-3*md-2*Mn
48 cat ("The mode is",Mode)
49
50 # Page.No 5.47
51 ###(ii) Another method
52 Lower_limits<-c(10,15,20,25,30)
53 Upper_limits<-c(14,19,24,29,34)
54 frequency<-c(5,4,6,3,2)
55 library("gds") # Package for grouped statistics
56 results<-gds(Lower_limits,Upper_limits,frequency)
57 cat("the mode of the distribution is",results$mode)

```

R code Exa 5.5.11 Calculation of Mode from Grouped Data

```
1 # Page.No 5.48
2
3 # A function created to find the median of a grouped
  data
4 GroupedMedian <- function(frequencies, intervals,
  sep = NULL, trim = NULL) {
5   # If "sep" is specified, the function will try to
  create the
6   # required "intervals" matrix. "trim" removes
  any unwanted
7   # characters before attempting to convert the
  ranges to numeric.
8   if (!is.null(sep)) {
9     if (is.null(trim)) pattern <- ""
10    else if (trim == "cut") pattern <- "
  \\[[\\]]|\\(|\\)"
11    else pattern <- trim
12    intervals <- sapply(strsplit(gsub(pattern, "",
  intervals), sep), as.numeric)
13  }
14
15  Midpoints <- rowMeans(intervals)
16  cf <- cumsum(frequencies)
17  Midrow <- findInterval(max(cf)/2, cf) + 1
18  L <- intervals[1, Midrow]      # lower class
  boundary of median class
19  h <- diff(intervals[, Midrow]) # size of median
  class
20  f <- frequencies[Midrow]      # frequency of
  median class
21  cf2 <- cf[Midrow - 1]        # cumulative
  frequency class before median class
```

```

22   n_2 <- max(cf)/2           # total
      observations divided by 2
23
24   unname(L + (n_2 - cf2)/f * h)
25 }
26
27 # The Problem
28 frequency<-c(2,3,1,4,2,1,2,4,3,2,7,4)
29 colnames<-c("numbers")
30 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] ",
31           " (43.5 ,50.5] ", " (50.5 ,57.5] ", " (57.5 ,64.5]
      " , " (64.5 ,71.5] ", " (71.5 ,78.5] ", "
      (78.5 ,85.5] ")
32 #The class intervals are changed to exclusive one by
      adding 0.5 to the upper limits and deducing 0.5
      from the lower limits
33 y<-matrix(frequency,nrow=length(frequency),dimnames=
      list(rownames,colnames))
34 md<-GroupedMedian(y[, "numbers"], rownames(y), sep="
      ,", trim="cut")
35
36
37 ### find the mean for grouped data
38 low_value<-1.5
39 high_value<-85.5
40 step_value<-7
41 X_breaks<-seq(low_value,high_value,step_value)
42 X_mid<-seq(low_value+step_value/2, high_value-step_
      value/2, step_value)
43 f<-c(2,3,1,4,2,1,2,4,3,2,7,4)
44 data<-rep(X_mid,f)
45 Mn<-mean(data)
46
47 ## to find mode from mean and median
48 Mode<-3*md-2*Mn
49 cat ("The mode is",Mode)

```

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

```
1 #Page.No 5.51 – 5.53
2
3 Lower_limits<-c(10,15,20,25,30,35,40,45)
4 Upper_limits<-c(14,19,24,29,34,39,44,49)
5 frequency<-c(3,6,5,9,8,8,2,8)
6 library("gds") # Package for grouped statistics
7 results<-gds(Lower_limits,Upper_limits,frequency)
8 cat("The mean is ",results$mean)
9 cat("The median is ",results$median)
10 cat("The mode is ",results$mode)
11 # The answers may slightly vary due to roundingoff
    values
12
13
14
15 ##### To get the Precise answers you can try the
    following method
16 # A function created to find the median of a grouped
    data
17 GroupedMedian <- function(frequencies, intervals,
    sep = NULL, trim = NULL) {
18   # If "sep" is specified, the function will try to
    create the
19   # required "intervals" matrix. "trim" removes
    any unwanted
20   # characters before attempting to convert the
    ranges to numeric.
21   if (!is.null(sep)) {
22     if (is.null(trim)) pattern <- ""
23     else if (trim == "cut") pattern <- "
    \\[[\\]]\\(\\)"
24     else pattern <- trim
```

```

25     intervals <- sapply(strsplit(gsub(pattern, "",
26         intervals), sep), as.numeric)
27 }
28 Midpoints <- rowMeans(intervals)
29 cf <- cumsum(frequencies)
30 Midrow <- findInterval(max(cf)/2, cf) + 1
31 L <- intervals[1, Midrow]      # lower class
32     boundary of median class
33 h <- diff(intervals[, Midrow]) # size of median
34     class
35 f <- frequencies[Midrow]      # frequency of
36     median class
37 cf2 <- cf[Midrow - 1]         # cumulative
38     frequency class before median class
39 n_2 <- max(cf)/2              # total
40     observations divided by 2
41 }
42 }
43 # The Problem
44 frequency<-c(3,6,5,9,8,8,2,8)
45 colnames<-c("numbers")
46 rownames<-c("[9.5,14.5]", "(14.5,19.5]",
47     "(19.5,24.5]", "(24.5,29.5]",
48     "(29.5,34.5]", "(34.5,39.5]", "(39.5,44.5]",
49     "(44.5,49.5]")
50 #The class intervals are changed to exclusive one by
51     adding 0.5 to the upper limits and deducing 0.5
52     from the lower limits
53 y<-matrix(frequency,nrow=length(frequency),dimnames=
54     list(rownames,colnames))
55 md<-GroupedMedian(y[, "numbers"], rownames(y), sep="
56     ", trim="cut")
57 }
58 }
59 }
60 ### find the mean for grouped data

```

```

51 low_value<-9.5
52 high_value<-49.5
53 step_value<-5
54 X_breaks<-seq(low_value,high_value,step_value)
55 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
56 f<-c(3,6,5,9,8,8,2,8)
57 data<-rep(X_mid,f)
58 Mn<-mean(data)
59 Mn
60
61 ## to find mode from mean and median
62 Mode<-3*md-2*Mn
63 cat ("The mode is",Mode)
64 cat("The median is",md)
65 cat("The mean is",Mn)

```

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

```

1 # Page.No 5.53
2
3 Lower_limits<-c(100,110,120,130,140,150,160)
4 Upper_limits<-c(109,119,129,139,149,159,169)
5 frequency<-c(10,12,13,18,6,7,4)
6 library("gds") # Package for grouped statistics
7 results<-gds(Lower_limits,Upper_limits,frequency)
8 cat("The mean is ",results$mean)
9 cat("The median is ",results$median)
10
11 GroupedMedian <- function(frequencies, intervals,
    sep = NULL, trim = NULL) {
12 # If "sep" is specified, the function will try to
    create the
13 # required "intervals" matrix. "trim" removes
    any unwanted

```

```

14 # characters before attempting to convert the
    ranges to numeric.
15 if (!is.null(sep)) {
16   if (is.null(trim)) pattern <- ""
17   else if (trim == "cut") pattern <- "
    \\[[\\]]|\\(|\\)"
18   else pattern <- trim
19   intervals <- sapply(strsplit(gsub(pattern, "",
    intervals), sep), as.numeric)
20 }
21
22 Midpoints <- rowMeans(intervals)
23 cf <- cumsum(frequencies)
24 Midrow <- findInterval(max(cf)/2, cf) + 1
25 L <- intervals[1, Midrow] # lower class
    boundary of median class
26 h <- diff(intervals[, Midrow]) # size of median
    class
27 f <- frequencies[Midrow] # frequency of
    median class
28 cf2 <- cf[Midrow - 1] # cumulative
    frequency class before median class
29 n_2 <- max(cf)/2 # total
    observations divided by 2
30
31 unname(L + (n_2 - cf2)/f * h)
32 }
33
34 # The Problem
35 frequency<-c(10,12,13,18,06,07,04)
36 colnames<-c("numbers")
37 rownames<-c("[99.5,109.5]", "(109.5,119.5]",
    "(119.5,129.5]", "(129.5,139.5]",
38 "(139.5,149.5]", "(149.5,159.5]",
    "(159.5,169.5]")
39 #The class intervals are changed to exclusive one by
    adding 0.5 to the upper limits and deducing 0.5
    from the lower limits

```

```
40 y<-matrix(frequency,nrow=length(frequency),dimnames=  
    list(rownames,colnames))  
41 md<-GroupedMedian(y[, "numbers"], rownames(y), sep="  
    ", trim="cut")  
42  
43  
44 Mode <- 3*md-2*results$mean  
45 cat("The mode is",Mode)
```

Chapter 6

Percentile and Percentile Rank

R code Exa 6.2.1 Percentile

```
1 # Page.No 6.5 – 6.8
2
3 ## Function to find percentile
4 percentile <- function(frequencies, intervals, sep =
  NULL, trim = NULL,p) {
5   # If "sep" is specified, the function will try to
  create the
6   #   required "intervals" matrix. "trim" removes
  any unwanted
7   #   characters before attempting to convert the
  ranges to numeric.
8   if (!is.null(sep)) {
9     if (is.null(trim)) pattern <- ""
10    else if (trim == "cut") pattern <- "
  \\[[\\]]\\(\\)"
11    else pattern <- trim
12    intervals <- sapply(strsplit(gsub(pattern, "",
  intervals), sep), as.numeric)
13  }
14  Midpoints <- rowMeans(intervals)
15  cf <- cumsum(frequencies)
```

```

16  row <- findInterval(max(cf)*p, cf) + 1
17  L <- intervals[1, row]      # lower class boundary
    of percentile class
18  h <- diff(intervals[, row]) # size of percentile
    class
19  f <- frequencies[row]      # frequency of
    percentile class
20  cf2 <- cf[row - 1]        # cumulative frequency
    class before percentile class
21  n_2 <- max(cf)*p          # total observations
    multiplied with percentile
22
23  unname(L + (n_2 - cf2)/f * h)
24 }
25
26 # The given Problem
27 frequency<-c(2,3,4,5,3,2,2,3)
28 colnames<-c("numbers")
29 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]")
30 #The class intervals are changed to exclusive one by
    adding 0.5 to the upper limits and deducing 0.5
    from the lower limits
31
32 y<-matrix(frequency,nrow=length(frequency),dimnames=
    list(rownames,colnames))
33
34 p=c(0.15,0.25,0.50,0.65,0.75,0.80,0.85)
35
36 Percentile_value <-percentile(y[, "numbers"],
    rownames(y), sep=" ", trim="cut",p)
37
38 Result<- round(rbind(p*100,Percentile_value),digits
    = 3)
39
40 print(Result)

```

R code Exa 6.3.1 Percentile rank

```
1 #Page.No 6.11
2 #Page.No 6.11
3
4 data<-as.vector(c
      (10,8,7,15,18,19,20,28,16,13,17,29,30,6,5,4,26,24,14,21)
      )
5 N<-length(data)
6
7 # ranking of scores
8 r<-rank(-data) # negative for rank the data in
      descending order
9 rank_matrix<-cbind(r,data)
10
11
12 # Rank of score 18
13 rank_matrix[5,]
14 R<-rank_matrix[5,1]
15
16 percentile_rank<-round(100-((100*R-50))/N,digits =
      0)
17 cat("The percentile rank of 18 is approximately ",
      percentile_rank)
```

R code Exa 6.3.2 Percentile rank

```
1 # Page No.6.12
2
3 data<-as.vector(c(65,59,46,32,25,15,14,10,9,7,5,3))
4 N<-length(data)
5
```

```

6 # ranking of scores
7 r<-rank(-data) # negative for rank the data in
   descending order
8 rank_matrix<-cbind(r,data)
9
10
11 # Rank of score 18
12 rank_matrix[5,]
13 R<-rank_matrix[5,1]
14
15 percentile_rank<-100-((100*R-50))/N
16 cat("The percentile rank of 18 is approximately ",
   percentile_rank)

```

R code Exa 6.3.3 Percentile rank

```

1 #Page.No 6.13
2
3 # Function to find percentile rank of grouped data
4 percentilerank<-function(class_interval,lower_limit,
   x){
5   L<-lower_limit
6   row<-which(data$c==class_interval)
7   f<-data$Freq[row]
8   cf<-data.frame(cumsum(data$Freq))
9   cf1<-cf$cumsum.data.Freq.[row]
10  cf2<-cf$cumsum.data.Freq.[row-1]
11  F<-if(row - 1<= 0) {
12    cf1
13  }else { cf2 }
14  N<-sum(data$Freq)
15  i<-step_value
16  unname((100/N)*(F+(x-L)/i*f))
17 }
18

```

```

19 # Given data
20 low_value<-9.5
21 high_value<-69.5
22 step_value<-10
23 X_breaks<-seq(low_value,high_value,step_value)
24 X_mid<-seq(low_value+step_value/2, high_value-step_
      value/2, step_value)
25 frequency<-c(4,8,2,2,3,1)
26 data<-rep(X_mid,frequency)
27 c<-cut(data,breaks = X_breaks)
28 table<-table(c)
29 data<-data.frame(table)
30 print(data)
31
32 # percentile rank of 24
33
34 PR22<-percentilerank("(19.5,29.5]",19.5,22)
35
36 cat("The percentile rank of 22 is ",PR22)

```

R code Exa 6.3.4 Percentile rank

```

1 #Page.No 6.13
2
3 # Function to find percentile rank of grouped data
4 percentilerank<-function(class_interval,lower_limit,
      x){
5   L<-lower_limit
6   row<-which(data$c==class_interval)
7   f<-data$Freq[row]
8   cf<-data.frame(cumsum(data$Freq))
9   cf1<-cf$cumsum.data.Freq.[row]
10  cf2<-cf$cumsum.data.Freq.[row-1]
11  F<-if(row - 1<= 0) {
12    cf1

```

```

13     }else { cf2 }
14     N<-sum(data$Freq)
15     i<-step_value
16     unname((100/N)*(F+(x-L)/i*f))
17 }
18
19 # Given data
20 low_value<-9.5
21 high_value<-89.5
22 step_value<-10
23 X_breaks<-seq(low_value,high_value,step_value)
24 X_mid<-seq(low_value+step_value/2, high_value-step_
      value/2, step_value)
25 frequency<-c(2,3,4,5,2,1,1,2)
26 data<-rep(X_mid,frequency)
27 c<-cut(data,breaks = X_breaks)
28 table<-table(c)
29 data<-data.frame(table)
30 print(data)
31
32 # percentile rank of 24
33
34 PR24<-percentilerank(" (19.5,29.5]" ,19.5,24)
35 PR36<-percentilerank(" (29.5,39.5]" ,29.5,36)
36 PR52<-percentilerank(" (49.5,59.5]" ,49.5,52)
37 PR55<-percentilerank(" (49.5,59.5]" ,49.5,55)
38 PR67<-percentilerank(" (59.5,69.5]" ,59.5,67)
39
40 results<-c(PR24,PR36,PR52,PR55,PR67)
41 Q<-c("PR24", "PR36", "PR52", "PR55", "PR67")
42
43 percentile_ranks<-data.frame(Q,results)
44 print(percentile_ranks)

```

R code Exa 6.4.1 Percentile Problems

```

1 # Page.No 6.18 – 6.23
2
3 percentile <- function(frequencies, intervals, sep =
  NULL, trim = NULL,p) {
4   # If "sep" is specified, the function will try to
  create the
5   #   required "intervals" matrix. "trim" removes
  any unwanted
6   #   characters before attempting to convert the
  ranges to numeric.
7   if (!is.null(sep)) {
8     if (is.null(trim)) pattern <- ""
9     else if (trim == "cut") pattern <- "
  \\[[\\]]|\\(|\\)"
10    else pattern <- trim
11    intervals <- sapply(strsplit(gsub(pattern, "",
  intervals), sep), as.numeric)
12  }
13  Midpoints <- rowMeans(intervals)
14  cf <- cumsum(frequencies)
15  row <- findInterval(max(cf)*p, cf) + 1
16  L <- intervals[1, row]      # lower class boundary
  of percentile class
17  h <- diff(intervals[, row]) # size of percentile
  class
18  f <- frequencies[row]      # frequency of
  percentile class
19  c <- cf[row - 1]
20  cf2 <- if(row - 1 <= 0 ) {0}
21  else { cf[row - 1] } # cumulative frequency class
  before percentile class
22  n_2 <- max(cf)*p          # total observations
  multiplied with percentile
23
24  unname(L + (n_2 - cf2)/f * h)
25 }
26
27

```

```

28 # The Problem
29 frequency<-c(9,5,7,8,8,6,2,3,2)
30 colnames<-c("numbers")
31 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]")
32 #The class intervals are changed to exclusive one by
  adding 0.5 to the upper limits and deducing 0.5
  from the lower limits
33 y<-matrix(frequency,nrow=length(frequency),dimnames=
  list(rownames,colnames))
34
35
36 p1=c
  (0.90,0.80,0.75,0.70,0.60,0.50,0.48,0.40,0.30,0.20)

37
38 p2=c(0.17,0.10)
39
40 Percentile_value1<-percentile(y[, "numbers"],
  rownames(y), sep=",", trim="cut",p1)
41
42 Percentile_value2<-percentile(y[, "numbers"],
  rownames(y), sep=",", trim="cut",p2)
43
44 Result1<- round(rbind(p1*100,Percentile_value1),
  digits = 3) # percentile values of above 20
45
46 Result2<- round(rbind(p2*100,Percentile_value2),
  digits = 3) # percentile values of below 20
47
48 print(Result1)
49
50 print(Result2)

```

R code Exa 6.5.1 Graphical Representation of Percentile

```
1 # Page.No 6.23–6.24
2
3 low_value<-19.5
4 high_value<-64.5
5 step_value<-5
6 X_breaks<-seq(low_value,high_value,step_value)
7 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
8 frequency<-c(1,2,2,7,10,8,5,3,2)
9 scores<-rep(X_mid,frequency)
10 cf<-cumsum(frequency)
11 cpf<-cf/sum(frequency)*100
12 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")
13 abline(h=c(17,25,50,63,75),untf = FALSE)
14
15 ## Or to make the clear visualization
16 abline(h=c(17,25,50,63,75),v=c(33.3,35.5,41,44,47))
17
18 ## to give colours
19 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

```
1 #Page. No 7.5
2
3 scores<-c(5,7,9,15,17,19)
4 range<-max(scores)-min(scores)
5 cat("the range of the scores is",range)
```

R code Exa 7.5.2 Range

```
1 # Page.No 7.5
2
3 scores<-c(19,22,90,100,150,190)
4 range<-max(scores)-min(scores)
5 cat("the range of the scores is",range)
```

R code Exa 7.5.3 Coefficient of Range

```

1 #Page.No 7.6
2
3 scores<-c(10,20,30,40,50,60,65,66,70,80)
4 hs<-max(scores)
5 ls<-min(scores)
6 Coef.R<-(hs-ls)/(hs+ls)
7 cat("The coefficient of Range is",Coef.R)

```

R code Exa 7.5.4 Coefficient of Range for Grouped Data

```

1 #Page.No 7.6
2
3 class_intervals <-c(10,14,15,19,20,24,25,29,30,34)
4 frequency<-c(2,8,6,2,2)
5 HS<-max(class_intervals) # highest score
6 LS<-min(class_intervals) #Lowest score
7
8 Coef.R<-(HS-LS)/(HS+LS)
9
10 cat("The coefficient of Range is",Coef.R)

```

R code Exa 7.5.5 Coefficient of Range for Grouped Data

```

1 #Page.No 7.7
2
3 class_intervals <-c(5,9,10,14,15,19,20,24,25,29)
4 frequency<-c(2,4,5,7,2)
5 HS<-max(class_intervals) #Highest score
6 LS<-min(class_intervals) # Lowest score
7
8 Coef.R<-(HS-LS)/(HS+LS)
9
10 cat("The coefficient of Range is",Coef.R)

```

R code Exa 7.5.6 Coefficient of Range for Grouped Data

```
1 #Page.No 7.7
2
3 frequency<-c(2,2,3,3)
4 scores<-c(10,14,15,19,20,24,25,29)
5 HS<-max(scores) #Highest score
6 LS<-min(scores) #Lowest score
7
8 Coef.R<-(HS-LS)/(HS+LS)
9
10 cat("The coefficient of Range is",Coef.R)
```

R code Exa 7.6.1 Quartile Deviation

```
1 #Page.No 7.10 -7.13
2
3 library("gds") # function to find descriptive
  statistics of Grouped data
4
5 f<-c(2,8,10,6,5,9) # frequency
6
7 Lower_limits<-c(9.5,14.5,19.5,24.5,29.5,34.5)
8 Upper_limits<-c(14.5,19.5,24.5,29.5,34.5,39.5)
9
10 results<-gds(Lower_limits,Upper_limits,f)
11
12 cat("The first quartile is",results$quartile1) # 1
  st Quartile
13
14 cat("The third quartile is",results$quartile3) # 3
  rd Quartile
```

```

15
16 #Quartile deviation
17 Q<-(results$quartile3-results$quartile1)/2
18
19 cat("The quartile deviation is ",Q)

```

R code Exa 7.6.2 Quartile Deviation

```

1 #Page.No 7.13–7.15
2
3 library("gds") # function to find descriptive
  statistics of Grouped data
4
5 # Given data
6 f<-c(2,8,10,12,8,10)
7 Lower_limits<-c(39.5,49.5,59.5,69.5,79.5,89.5)
8 Upper_limits<-c(49.5,59.5,69.5,79.5,89.5,99.5)
9 results<-gds(Lower_limits,Upper_limits,f)
10
11 cat("The first quartile is",results$quartile1)
12
13 cat("The third quartile is",results$quartile3)
14
15 # Quartile deviation
16 Q<-(results$quartile3-results$quartile1)/2
17 cat("The quartile deviation is ",Q)

```

R code Exa 7.6.3 Quartile Deviation

```

1 #Page.No 7.15 –7.17
2
3 library("gds") # function to find descriptive
  statistics of Grouped data

```

```

4
5 f<-c(6,4,6,4,8,7,5)
6 Lower_limits<-c(39.5,44.5,49.5,54.5,59.5,64.5,69.5)
7 Upper_limits<-c(44.5,49.5,54.5,59.5,64.5,69.5,74.5)
8 results<-gds(Lower_limits,Upper_limits,f)
9
10 cat("The first quartile is",results$quartile1)
11
12 cat("The third quartile is",results$quartile3)
13
14 # Quartile deviation
15 Q<-(results$quartile3-results$quartile1)/2
16 cat("The quartile deviation of the distribution is ",
      Q)

```

R code Exa 7.7.2 Average Deviation

```

1 #Page.No 7.19 -7.20
2
3 scores<-c(20,18,42,46,50,64,75)
4 A.D<-mean(abs(scores-mean(scores)))
5
6 cat("Average Deviation of the scores is ",A.D)
7
8 "The answer provided in the textbook is wrong"
9 ### The first value 20 - mean(scores) is 25 not 15

```

R code Exa 7.7.3 Average Deviation

```

1 #Page.No 7.20
2
3 x<-c(25,45,15,16,84) # Scores
4 A.D<-mean(abs(x-mean(x)))

```

```
5
6 cat(" Average Deviation of the distribution is ",A.D)
```

R code Exa 7.7.5 Average Deviation from Grouped data

```
1 #Page.No 7.22
2
3 # Given distribution table
4 low_value<-99.5
5 high_value<-169.5
6 step_value<-10
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
9 f<-c(10,8,7,6,12,9,8)
10 data<-rep(X_mid,f)
11
12 # Average deviation
13 A.D<-mean(abs(data-mean(data)))
14
15 cat(" Average Deviation of the distribution is ",A.D)
```

R code Exa 7.7.6 Average Deviation from Grouped data

```
1 #Page.No 7.23
2
3 # Given distribution table
4 low_value<-69.5
5 high_value<-81.5
6 step_value<-2
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
```

```

9 f<-c(3,3,4,1,4,5)
10 data<-rep(X_mid,f)
11
12 # Average deviation
13 A.D<-mean(abs(data-mean(data)))
14
15 cat("Average Deviation of the distribution is ",A.D)

```

R code Exa 7.7.7 Average Deviation from Grouped data

```

1 #Page.No 7.24
2
3 # Given distribution table
4 low_value<-103.5
5 high_value<-128.5
6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
9 f<-c(4,3,2,6,5)
10 data<-rep(X_mid,f)
11
12 # Average deviation
13 A.D<-mean(abs(data-mean(data)))
14
15 cat("Average Deviation of the distribution is ",A.D)

```

R code Exa 7.8.1 Standard Deviation

```

1 #Page.No 7.27 – 7.28
2
3 scores<-c(8,9,10,11,12,13,14,15)
4

```

```

5 n <- length(scores) # number of values
6
7 std <- sd(scores) # sample standard deviation
8
9 S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
  std deveiation
10
11 cat("The standard deviation of the scores is",S.D)
12
13 "The Answer given in the book is wrong"

```

R code Exa 7.8.2 Standard Deviation

```

1 #Page.No 7.28 – 7.29
2
3 x<-c(10,20,30,40,50,60,70)
4
5 n <- length(x) # number of values
6
7 std <- sd(x) # sample standard deviation
8
9 S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
  std deveiation
10
11 cat("The standard deviation of the scores is",S.D)

```

R code Exa 7.8.5 Standard Deviation from Grouped data

```

1 #Page.No 7.31 – 7.32
2
3 # Given distribution table
4 low_value<-9.5
5 high_value<-34.5

```

```

6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
9 f<-c(2,8,6,2,2)
10 data<-rep(X_mid,f)
11
12 n <- length(data) # number of values
13
14 std <- sd(data) # sample standard deviation
15
16 S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
  std deveiation
17
18 cat("The standard deviation of the scores is",S.D)

```

R code Exa 7.8.6 Standard Deviation from Grouped data

```

1 #Page.No 7.32 – 7.33
2
3 # Given distribution data
4 low_value<-9.5
5 high_value<-34.5
6 step_value<-5
7 X_breaks<-seq(low_value,high_value,step_value)
8 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
9 f<-c(2,8,6,2,2)
10 data<-rep(X_mid,f)
11
12 n <- length(data) # number of values
13 std <- sd(data) # sample standard deviation
14
15 S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
  std deveiation

```

```
16
17 cat("The standard deviation of the scores is",S.D)
```

R code Exa 7.8.9 Effects upon Std Deviation

```
1 #Page.No 7.37
2
3 scores<-c(2,5,4,6,8)
4 Mn<-mean(scores)
5 n <- length(scores) # number of values
6 std <- sd(scores) # sample standard deviation
7 S.D<-sqrt((std ^ 2) * ((n - 1) / n)) # Population
   Standrad deviation
8
9
10 # (i) Addition of 2 in each scores
11 scores_new<-scores+2
12 Mn_new<-mean(scores_new)
13 n <- length(scores_new) # number of values
14 std_new <- sd(scores_new) # sample standard
   deviation
15 S.D_new<-sqrt((std_new ^ 2) * ((n - 1) / n))
16
17 cat("Previously the Mean and standard deviation
   respectively are",Mn ,
18     "and", S.D)
19 cat("After addition of 2 in each score Mean has
   enhanced ",Mn_new-Mn,"
20     and standard deviation remains unchanged i.e ",S
   .D-S.D_new)
21
22
23 #(ii) subtraction of 2 in each scores
24 scores_sub<-scores-2
25 Mn_sub<-mean(scores_sub)
```

```

26 n <- length(scores_sub) # number of values
27 std_sub <- sd(scores_sub) # sample standard
   deviation
28 S.D_sub<-sqrt((std_sub ^ 2) * ((n - 1) / n))
29
30 cat("Previously the Mean and standard deviation
   respectively are",Mn ,
31     "and", S.D)
32 cat("After subtraction of 2 in each score Mean has
   reduced by ",Mn-Mn_sub ,
33     "and standard deviation remains unchanged i.e ",
   S.D-S.D_sub)
34
35
36 # (iii) multiplication of 2 in each scores
37 scores_mult<-scores*2
38 Mn_mult<-mean(scores_mult)
39 n <- length(scores_mult) # number of values
40 std_mult <- sd(scores_mult) # sample standard
   deviation
41 S.D_mult<-sqrt((std_mult ^ 2) * ((n - 1) / n))
42
43 cat("Previously the Mean and standard deviation
   respectively are",Mn ,
44     "and", S.D)
45 cat("After multiplication of 2 in each score Mean
   has multiplied by ",2,
46     "and the result is ",Mn_mult,"thats is same as "
   ,Mn*2)
47 cat("standard deviation also multiplied by ",2,
48     "and the result is",S.D_mult,"thats is same as",
   S.D*2 )
49
50 # (iii) division of 2 in each scores
51 scores_div<-scores/2
52 Mn_div<-mean(scores_div)
53 n <- length(scores_div) # number of values
54 std_div <- sd(scores_div) # sample standard

```

```

      deviation
55 S.D_div<-sqrt((std_div ^ 2) * ((n - 1) / n))
56
57 cat(" Previously the Mean and standard deviation
      respectively are",Mn ,
58     "and", S.D)
59 cat(" After division 2 in each score Mean has divided
      by ",
60     2,"and the result is ",Mn_div,"i.e ", Mn/2)
61
62 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

```

1 #Page.No 7.40 – 7.41
2
3 M<-c(40,30) # Mean values
4 N<-c(20,30) # No.of students in a class
5 S.D<-c(8,4) # Std deviation of students
6
7 M.comb<-weighted.mean(M,N)
8
9 # Difference between the mean of the 1st and 2nd
   distribution and combined mean
10 d1sqr<-(40-M.comb)^2
11 d2sqr<-(30-M.comb)^2
12
13 sigma.comb<-sqrt(((20*(64-d1sqr)))+(30*(16+d2sqr)))/
   50)
14
15 cat("The combined mean is",M.comb)
16
17 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

```
1 # Page.No 8.8 – 8.9
2
3 # Spearman's Rank Difference Method
4 score_in_maths<-c(77,83,85,65,50,60,70,80,78,82)
5 score_in_science<-c(85,60,70,50,40,45,46,44,42,47)
6
7 # Ranks of the scores
8 R1<-rank(score_in_maths)
9 R2<-rank(score_in_science)
10
11 # Correlation coefficient
12 P<-cor(R1,R2)
13 cat("There is moderate positive correlation i.e ",P)
14 round(P,digits = 1)
```

R code Exa 8.8.2 Spearman s Rank Difference Method

```

1 #Page.No 8.9 – 8.10
2
3 # Rank of students by teachers
4 teacher1<-c(1,2,3,4,5)
5 teacher2<-c(4,3,2,1,5)
6
7 r<-cor(teacher1,teacher2)
8
9 cat("Their is ",r," or No correlation")

```

R code Exa 8.8.3 Spearman s Rank Difference Method

```

1 # Page.No 8.10 –8.11
2
3 # Rank of the students
4 Rank_in_theory<-c(1,10,2,9,4,8,7,5,6,3)
5 Rank_in_practical<-c(10,1,4,2,9,3,7,5,6,8)
6
7 p<-cor(Rank_in_theory,Rank_in_practical)
8
9 cat("There is high negative correlation i.e ",p)

```

R code Exa 8.8.4 Spearman s Rank Difference Method

```

1 #Page.No 8.11
2
3 #Marks of the five students
4 marks_in_english<-c(78,80,81,60,90)
5 marks_is_science<-c(80,83,85,95,50)
6
7 # Ranks
8 R1<-rank(marks_in_english)
9 R2<-rank(marks_is_science)

```

```

10
11 P<-cor(R1,R2)
12
13 cat("There is Negative correlation i.e ",P)

```

R code Exa 8.8.5 Spearman s Rank Difference Method

```

1 #Page.No 8.12
2
3 # Scores in test
4 scores_in_oriya<-c(40,40,50,30,60,80,90)
5 scores_in_Geography<-c(50,60,40,40,50,80,95)
6
7 # Ranks
8 R1<-rank(scores_in_oriya)
9 R2<-rank(scores_in_Geography)
10
11 P<-cor(R1,R2)
12
13 cat("There is High positive correlation i.e ",P)

```

R code Exa 8.8.6 Spearman s Rank Difference Method

```

1 #Page.no 8.12 – 8.15
2
3 # Ranks of five students by examiners
4 R1<-c(3,2,1,4,5)
5 R2<-c(4,5,3,2,1)
6 R3<-c(5,4,2,3,1)
7
8 P12<-cor(R1,R2) # Rank correlation between examiner
   1 and 2

```

```

9 P23<-cor(R2,R3) # Rank correlation between examiner
  2 and 3
10 P13<-cor(R1,R3) # Rank correlation between examiner
  1 and 3
11
12 cat("The rank correlation between Examiners is",P12,
  P13,P23)
13 # above three correlations P23 have the highest co-
  relations

```

R code Exa 8.8.7 Spearman s Rank Difference Method

```

1 #Page.No 8.15 – 8.16
2
3 # Rank in efficiency
4 Rx<-c(1,2,3,4,5,6,7,8,9,10)
5
6 length_of_services_in_years<-c
  (15,12,10,11,13,10,10,18,11,13)
7
8 Ry<-rank(length_of_services_in_years)
9
10 P<-print(cor(Rx,Ry),digits=0)
11
12 #Hence there is no correlaion between the experience
13 # and efficiency of employees

```

R code Exa 8.8.8 Spearman s Rank Difference Method

```

1 # Page.No 8.16
2
3 X<-c(57,58,33,58,60,63,71,63,74,63) # Scores on
  creativity

```

```

4
5 Y<-c(92,100,100,98,102,95,100,99,106,104) # Scores
   of intelligence
6
7 # Ranks of scores
8 R1<-rank(X)
9 R2<-rank(Y)
10
11 P=cor(R1,R2) # Correlation
12
13 cat("Hence the correlation coefficient between
   creativity
14   and intelligence scores of the student is",P)

```

R code Exa 8.8.9 Spearman s Rank Difference Method

```

1 #Page.no 8.18
2
3 test_scores1<-c(50,56,59,62,65)
4 test_scores2<-c(22,34,28,30,32)
5
6 # Ranks of the given scores
7 R1<-rank(test_scores1)
8 R2<-rank(test_scores2)
9
10 P<-cor(R1,R2)
11 cat("The coefficient of correlation is",P,"which is
   very low")

```

R code Exa 8.8.10 Spearman s Rank Difference Method

```

1 # Page.No 8.19
2

```

```

3 #Given Scores
4 X<-c(45,48,52,47,49,50,46,51)
5 Y<-c(70,72,80,75,73,76,71,71)
6
7 # Ranks of scores
8 R1<-rank(X)
9 R2<-rank(Y)
10
11 P=cor(R1,R2)
12
13 cat("Hence the correlation coefficient is",P , " i.e
      moderate")

```

R code Exa 8.8.11 Spearman s Rank Difference Method

```

1 #Page.No# Page.No 8.20
2
3 marks_in_maths<-c(50,26,46,25,31,22,29,30,28,18)
4 marks_in_g.Sci<-c(40,35,34,30,32,28,28,32,29,15)
5
6 #Ranks of the marks
7 R1<-rank(marks_in_maths)
8 R2<-rank(marks_in_g.Sci)
9
10 P=cor(R1,R2)
11
12 cat("Here the coeffient of co-ordinatore",P ," Highly
      Positive")

```

R code Exa 8.9.1 Pearson s Product Moment Coefficient of Correlation

```

1 # Page.No 8.22 – 8.25
2

```

```

3 #(i) By rank difference method
4 marks_in_history<-c(6,8,4,5,2)
5 marks_in_engilsh<-c(9,6,12,10,13)
6
7 #Ranks of marks
8 R1<-rank(marks_in_history)
9 R2<-rank(marks_in_engilsh)
10
11 P<-cor(R1,R2) # Correlation
12
13 cat("The correlation is ",P," Perfect negative
      correlation")
14
15
16 ##(ii) By product moment method
17
18 r<-cor(marks_in_history,marks_in_engilsh,method = "
      pearson")
19 cat("The correlation is",r," It is very high negative
      correlation")

```

R code Exa 8.9.2 Pearson s Product Moment Coefficient of Correlation

```

1 # Page.No 8.25 – 8.26
2
3 # Scores
4 X<-c(55,65,25,50,40,35,15,20,50,45)
5 Y<-c(50,75,45,55,50,25,35,40,70,55)
6
7 r<-cor(X,Y,method = "pearson")
8 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

```
1 # Page.No 8.27 – 8.28
2
3 # Scores
4 X<-c(60,51,43,56,39,48,59,44,37,55)
5 Y<-c(46,35,35,42,31,34,48,31,30,44)
6
7 r<-cor(X,Y,method = "pearson")
8
9 cat("coefficient of correlation of the given set is
    ",r,
10    "which is very high positive correlation")
```

R code Exa 8.9.4 Pearson s Product Moment Coefficient of Correlation

```
1 # Page.No 8.28 – 8.29
2
3 #Scores
4 X<-c(65,60,50,40,30,20,80)
5 Y<-c(50,55,80,90,85,79,77)
6
7 r<-cor(Y,X,method = "pearson")
8
9 cat("The product moment co-efficient of correlation
    between the above sets of scores is ",r)
10 "The answer provided in the text book is wrong"
```

R code Exa 8.10.1 Product Mement Method from Group Data

```
1 #Page.No 8.30 – 8.35
2
```

```

3 X<-c
  (50,41,42,45,60,43,46,47,48,50,58,45,40,35,30,20,18,15,14,10)
  # TEST I
4 Y<-c
  (80,90,95,96,75,70,50,60,55,50,45,45,40,30,30,25,19,18,17,10)
  # TEST II
5
6 # X table cacclulations
7 low_value<-9.5
8 high_value<-64.5
9 step_value<-5
10 X_breaks<-seq(low_value,high_value,step_value)
11 X_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
12 c1<-cut(X,breaks = X_breaks)
13 table1<-table(c1)
14 data1<-data.frame(table1)
15
16 f1<-data1$Freq # X values grouped and their
  frequencies
17
18 x_dash<-(X_mid-round(median(X),digits = 0))/step_
  value # standard values x'
19
20 fxdash<-f1*x_dash #fx'
21
22 fxdashsqr<-fxdash*x_dash# fx'^2
23
24
25 # Claculation of Y variables
26 low_value<-9.5
27 high_value<-99.5
28 step_value<-10
29 Y_breaks<-seq(low_value,high_value,step_value)
30 Y_mid<-seq(low_value+step_value/2, high_value-step_
  value/2, step_value)
31 c2<-cut(Y,breaks = Y_breaks)
32 table2<-table(c2)

```

```

33 data2<-data.frame(table2)
34
35 f2<-data2$Freq
36
37 y_dash<-round(((Y_mid-(median(Y)))/10),digits = 0) #
    y'
38
39 fydash<-f2*y_dash    #fy '
40
41 fydashsqr<-fydash*y_dash    #fy '^2
42
43 N<-sum(f1)
44
45 fxdashydash<-108 # from scatter diagram
46
47 # calculation of correlation coefficient
48
49 numerator<-(N*sum(fxdashydash))-((sum(fxdash)*sum(
    fydash))
50
51 denominator<-sqrt(((N*sum(fxdashsqr))-((sum(fxdash))
    ^2))*((N*sum(fydashsqr))-((sum(fxdash))^2)))
52
53 r<-numerator/denominator
54
55 cat("The correlation coefficient is",r," positive")
56
57 "The answer provided in the book is wrong"

```

R code Exa 8.10.2 Raw Score Method

```

1 #Page.No 8.36
2
3 #Raw score method
4

```

```
5 X<-c(10,15,18,22,25)
6 Y<-c(32,18,25,21,24)
7
8 r<-cor(X,Y,method="pearson")
9 cat("The coefficient of correlation is ",r,"low
    negative")
```

R code Exa 8.10.3 Raw Score Method

```
1 #Page.No 8.37
2
3 x<-c(5,10,15,20,80)
4 y<-c(30,37,80,90,95)
5
6 r<-cor(x,y,method="pearson")
7
8 cat("The coefficient of correlation is ",r,"moderate
    positive")
```

R code Exa 8.10.4 Raw Score Method

```
1 #Page.No 8.39 – 8.40
2
3 x<-c(5,4,6,8,7)
4 y<-c(4,6,7,8,5)
5
6 r<-cor(x,y,method="pearson")
7
8 cat("The coefficient of correlation is ",r,"moderate
    positive")
```

R code Exa 8.11.1 Biserial Correlation

```
1 #Page.No 8.41 – 8.43
2
3 library("ltm")
4
5 #The data
6 low_value<-9.5
7 high_value<-34.5
8 step_value<-5
9 X_breaks<-seq(low_value,high_value,step_value)
10 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
11
12 #Trained data
13 f1<-c(4,3,6,5,2)
14 data1<-rep(X_mid,f1)
15
16 #Untrained data
17 f2<-c(8,6,12,10,4)
18 data2<-rep(X_mid,f2)
19
20 # Calculation of correlation
21
22 X_mid
23 f<-f1+f2
24 d<-rep(X_mid,f)
25 #d<-c(rep(12,12),rep(17,9),rep(22,18),rep(27,15),rep
    (32,6)) #Mid points and their total frequency
26
27 # 0 represents trained and 1 represents untrained
28 # y gives trained and untrained groups with their
    respective frequencies
29
30 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))
31
32
```

```

33 biserial.cor(d,y)
34
35 "The answer given in the book is wrong"
36
37 # Book error
38 # mean of trained data
39 Mp<-mean(data1)
40 print(Mp)
41 "which is correct "
42
43 # Mean of Untrained data"
44 Mq<-mean(data2)
45 print(Mq)
46 "They mentioned it as 20.9"
47
48 # The difference will be zero hence the whole answer
   will become zero
49 Mp-Mq

```

R code Exa 8.11.2 Biserial Correlation

```

1 #Page.No 8.44–8.46
2
3 library("ltm")
4
5 #The data
6 low_value<-9.5
7 high_value<-79.5
8 step_value<-10
9 X_breaks<-seq(low_value,high_value,step_value)
10 X_mid<-seq(low_value+step_value/2, high_value-step_
   value/2, step_value)
11
12 #Trained data
13 f1<-c(2,2,4,7,4,3,3)

```

```

14 data1<-rep(X_mid,f1)
15
16 #Untrained data
17 f2<-c(8,9,9,9,4,4,7)
18 data2<-rep(X_mid,f2)
19
20
21 # Calculation of correlation
22
23 f<-f1+f2
24 X_mid
25 d<-rep(X_mid,f)
26
27 # 0 represents trained and 1 represents untrained
28 # y gives trained and untrained groups with their
    respective frequencies
29 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))
30
31 rbis<-biserial.cor(d,y)
32
33 cat("Bi-serial correlation is ",rbis," which is low
    positive")
34
35 "The answer may slightly vary due to roundingoff
    values"

```

R code Exa 8.11.3 Biserial Correlation

```

1 #Page.No 8.47- 8.48
2
3 library("ltm")
4
5 #The data

```

```

6 low_value<-4.5
7 high_value<-19.5
8 step_value<-3
9 X_breaks<-seq(low_value,high_value,step_value)
10 X_mid<-seq(low_value+step_value/2, high_value-step_
    value/2, step_value)
11
12 #Trained data
13 f1<-c(4,4,5,4,3)
14 data1<-rep(X_mid,f1)
15
16 #Untrained data
17 f2<-c(7,5,14,7,7)
18 data2<-rep(X_mid,f2)
19
20
21 # Calculation of correlation
22
23 f<-f1+f2 # Total frequency
24 X_mid # mid point
25 d<-rep(X_mid,f)
26
27 # 0 represents trained and 1 represents untrained
    groups
28 # y gives trained and untrained groups with their
    respective frequencies
29 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))
30
31 rbis<-biserial.cor(d,y)
32 cat("The bi-serial correlation is",rbis,"Negligible
    Negative")
33 "The answer given in the book is wrong "
34
35 # Book Error
36 #negative correlation
37 # Mean values given differ from the actual mean
    values

```

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

R code Exa 8.12.1 Point Biserial Correlation

```
1 #Page.no 8.48 – 8.50
2
3 scores<-c(35,40,32,31,20,18,12,32,10,10)
4
5 items<-c(1,1,1,1,0,0,0,1,0,0)
6
7 library("ltm") # Packge contains point biserial
  correlation function
8
9 r_pbis<-biserial.cor(scores,items,level = 2)
10
11 cat("Point-biserial correlation is ",r_pbis)
```

R code Exa 8.12.2 Point Biserial Correlation

```
1 #Page.No 8.51–8.52
2
3 scores<-c(40,80,90,17,25)
4 items<-c(1,1,1,0,0)
5
6 library("ltm") # Packge contains point biserial
  correlation function
7
8 r_pbis<-biserial.cor(scores,items,level = 2)
9
10 cat("Point-biserial correlation is ",r_pbis,"high
  positive")
```

R code Exa 8.13.1 Partial Correlation

```
1 #Page.No 8.53 – 8.57
2
3
4 r12<-0.9 # correlation between 1st and 2nd variable
5 r13<-0.6 # correlation between 1st and 3rd variable
6 r23<-0.2 # correlation between 2nd and 3rd variable
7
8 # (a) keeping 3rd variable constant
9 r12.3<-(r12-(r13*r23))/(sqrt(1-r13^2)*sqrt(1-r23^2))
10
11 # (b) keeping 1st variable constant
12 r23.1<-(r23-(r12*r13))/(sqrt(1-r12^2)*sqrt(1-r13^2))
13
14 #(c) Keeping 2nd variable constant
15 r13.2<-(r13-(r12*r23))/(sqrt(1-r12^2)*sqrt(1-r23^2))
16
17 cat("The partial correlation coefficients keeping
18     the
19     1st 2nd 3rd variables as constant respectly are"
20     ,r23.1,r13.2,r12.3)
```

R code Exa 8.13.2 Partial Correlation

```
1 #Page.No 8.55
2
3 r12<-15.7 # correlation between 1st and 2nd
4           variable
5 r13<-0.80 # correlation between 1st and 3rd
6           variable
7 r23<-4    # correlation between 2nd and 3rd variable
```

```

6
7 # (a) keeping 3rd variable constant
8 r12.3<-(r12-(r13*r23))/(sqrt(1-r13^2)*-sqrt(abs(1-
    r23^2)))
9
10 # (b) keeping 1st variable constant
11 r23.1<-(r23-(r12*r13))/(-sqrt(abs(1-r12^2))*-sqrt(
    abs(1-r13^2)))
12
13 #(c) Keeping 2nd variable constant
14 r13.2<-(r13-(r12*r23))/(-sqrt(abs(1-r12^2))*-sqrt(
    abs(1-r23^2)))
15
16 cat("The partial correlation coefficients keeping
    the
17     1st 2nd 3rd variables as constant respectly are"
    ,r12.3,r13.2,r23.1)

```

R code Exa 8.14.1 Tetrachoric Correlation

```

1 #Page.No 8.57 – 8.59
2
3 A<-30
4 B<-40
5 C<-20
6 D<-30
7
8 ans<-(180*sqrt(B*C)/(sqrt(A*D)+sqrt(B*C)))
9
10 cat("converting cos",ans,"to r gives correlation
    value")
11
12 rt<-0.087 # From the table value corresponding to
    cosine
13

```

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

R code Exa 8.14.2 Tetrachoric Correlation

```
1 #Page.No 8.59
2
3 A<-50
4 B<-35
5 C<-40
6 D<-25
7
8 ans<-(180*sqrt(A*D)/(sqrt(B*C)+sqrt(A*D)))
9
10 cat("converting cos",ans,"to r gives correlation
      value") # From the table value
11
12 rt<-0.087 # From the table value corresponding to
      cosine
13 # since AD greater than BC correlation is negative
14
15 cat("tetrachoric correlation is",-rt," Negatively
      correlated")
```

R code Exa 8.15.1 Phi Coefficient

```
1 #Page.No 8.60 – 8.61
2
3 library("psych") # attaching the function
4
5 t<-c(50,80,60,40) # vector of frequencies
6
7 Phi_coefficient<-phi(t,digits = 4)
```

```
8
9 cat("The phi coefficient is ",Phi_coefficient)
```

R code Exa 8.15.2 Phi Coefficient

```
1 #Page.No 8.61–8.62
2
3 library("psych") # attaching the function
4
5 t<-c(50,40,30,15) # vector of frequencies
6
7 Phi_coefficient<-phi(t,digits = 4)
8
9 cat("The phi coefficient is ",Phi_coefficient,"Low
    negative correlation")
```

Chapter 9

Normal Distribution

R code Exa 9.7.1 Uses of Normal Curve

```
1 #Page.No 9.13 – 9.14
2
3 #(a)
4 ans<-(pnorm(60, mean=50, sd=6)-pnorm(40, mean=50, sd
      =6))*100
5
6 cat("Percentage of cases lie between 40 and 60 is",
      ans)
7
8 "The answer may slightly vary due to rounding off
      values"
```

R code Exa 9.7.2 Uses of Normal Curve

```
1 # Page.no 9.14
2
3 # To compare scores on tow different tests
4 z1<-scale(70,center = 75,scale = 10)
```

```

5 z2<-scale(60,center = 50,scale = 6)
6
7 cat("z-score in mathematics is",z2,
8     "which is greater than z-score in general
9     science",z1)
9 "so his performance is better in mathematics"

```

R code Exa 9.7.3 Uses of Normal Curve

```

1 # Page.no. 9.15
2
3 # To determine the percentile rank of candidate in
4   his group
5
6 scale(50,center = 60,scale = 6)
7
8 pnorm(50, mean=60, sd=6)*100
9
10 ans<-round(pnorm(50, mean=60, sd=6),digits = 2)*100
11
12 cat("The percentile rank of the student who secured
13     50 marks
14     in an achievement test is ",ans,"%")

```

R code Exa 9.7.4 Uses of Normal Curve

```

1 # Page.no. 9.16
2
3 #To determine percentile rank of a candidate
4 # To find the z- score
5 scale(60,center = 50,scale = 6)
6
7 #to find the percentage
8 ans<-round(pnorm(60, mean=50, sd=6),digits = 2)*100

```

```
9
10 cat("The percentile rank of the student is ",ans)
```

R code Exa 9.7.5 Uses of Normal Curve

```
1 #Page.No 9.16 - 9.17
2
3 #To determine the limits of the scores
4 # To find the z- score
5 Z1<-scale(60,center = 50,scale = 5)
6 Z2<-scale(40,center = 50,scale = 5)
7
8 #To find percentage of cases
9 ans<-(pnorm(60, mean=50, sd=5)-pnorm(40, mean=50, sd
    =5))*100
10
11 cat("Percentage of cases lie between",Z2," and",Z1,"
    is",ans,"%")
```

R code Exa 9.7.6 Relative difficulty value

```
1 #Page.No 9.17
2
3 # To determine the relative difficulty of the test
  items
4 A<--qnorm(0.30)
5 B<--qnorm(0.40)
6
7 cat("difficulty values of A and B respectively are "
    ,A,B)
8
9 " Question A is relatively difficult"
```

R code Exa 9.7.7 Percentage of Cases

```
1 #Page.No 9.17–9.18
2 # To determine the percentage of cases that are
   above and below given scores
3
4 # To find the z- score
5 scale(55,center = 52,scale = 5)
6
7 # (a) to find the percentage of cases lie above 55
8
9 ans<-round(pnorm(55, mean=52, sd=5,lower.tail =
   FALSE)*100,digits = 2)
10 cat("The percentage of cases lie above 55 i.e above
   0.6sd is ",ans,"%")
11
12 # (b) to find the percentage of cases lie below 55
13
14 ans<-round(pnorm(55, mean=52, sd=5,lower.tail = TRUE
   )*100,digits = 2)
15 cat("The percentage of cases lie above 55 i.e below
   0.6sd is ",ans,"%")
16
17 # (C) to find the percentage of cases lie below 50
18 scale(50,center = 52,scale = 5)
19
20 ans<-round(pnorm(50, mean=52, sd=5,lower.tail = TRUE
   )*100,digits = 2)
21 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

```
1 #Page.NO 9.19
2 # To determine the score of a candidate whose
   percentile rank is known
3
4 X<-qnorm(0.84,50,7)
5
6 cat("Thus a candidate secured",X,"marks in an
   achievement test")
```

Chapter 10

Interpretation of Scores

R code Exa 10.5.1 Standard Score

```
1 #Page.No 10.5
2
3 x<-50 #Marks
4 M<-60 # Mean
5 S.D<-10 # Std deviation
6
7 z<-scale(x,center = M,scale = S.D)
8
9 cat("Rama's score is ",z[1,]," i.e 1-sigma
    distance below the mean")
```

R code Exa 10.5.2 Standard Score

```
1 #Page.No 10.5
2
3 x<-50 #Marks
4 M<-40 # Mean
5 S.D<-10 # Std deviation
```

```

6
7 z<-scale(x,center = M,scale = S.D)
8
9 cat("Hari's score is ",z[1,],"i.e 1sigma distance
    above the mean")

```

R code Exa 10.5.3 Standard Score

```

1 #Page.No 10.6
2
3 #Rama In english
4 z<-scale(60,center = 40,scale = 10)
5 cat("Rama secures ",z[1,],"in mothertougue")
6 "Rama's Z-score is +2sigma"
7
8 #Hari in english
9 z<-scale(60,center = 70,scale = 10)
10 cat("Hari secures ",z[1,],"in mothertougue")
11 "Hari's Z-score is -1sigma"
12
13 "Rama remains +2 sigma distance from mean
14 Hari remains -1 sigma distance below from mean
15 So Rama's performance is better"

```

R code Exa 10.6.1 T Score

```

1 #Page.No 10.6 - 10.7
2
3 # In english
4 z<-scale(60,center = 40,scale = 8)
5 z<-z[1,]
6 T_score_E<-10*z+50
7 cat("Rama T_score in english is ",T_score_E)

```

```

8
9 #In mother tongue
10 z<-scale(50,center = 50,scale = 6)
11 z<-z[1,]
12 T_score_M<-10*z+50
13 cat("Rama T_score in mother tongue is ",T_score_M)
14
15 "Rama performance better in English than mother
    tongue"

```

R code Exa 10.7.1 H score

```

1 #Page.no 10.7 - 10.8
2
3 x<-50 #Marks
4 M<-40 # Mean
5 S.D<-6 # Std deviation
6
7 z<-scale(x,center = M,scale = S.D)
8
9 z<-z[1,]
10
11 H_score<-50+14*z
12
13 cat("Rama H_score is ",H_score)

```

Chapter 11

Chi Square

R code Exa 11.4.1 Chi Square Testing

```
1 # Page no. : 11.5 - 11.6
2
3 condition <- c("Favourable", "Unfavourable", "
  Undecided")
4 observed <- c(20, 40, 21)
5 expected <- c(27, 27, 27)
6
7 DF <- data.frame(condition, observed, expected)
8 View(DF)
9
10 alpha5 <- 0.05 # 5% significance level
11 alpha1 <- 0.01 # 1% significance level
12 df <- nrow(DF) - 1 # Degree of freedom
13
14 chi_sq_val_5 <- round(qchisq(alpha5, df, lower.tail =
  F), 2) # Chi-square value
15 chi_sq_val_1 <- round(qchisq(alpha1, df, lower.tail = F
  ), 2)
16
17 diff <- DF$observed - DF$expected # Difference
18 diff_sq <- diff ** 2 # Difference squared
```

```

19 ans <- round(diff_sq / DF$expected, 2) #
    Difference squared weighted by expected frequency
20
21 chi_sq <- sum(ans) # Chi-square
22
23 cat("Value of chi-square is", chi_sq)
24
25 if(chi_sq > chi_sq_val_5)
26   if(chi_sq > chi_sq_val_1)
27   {
28     cat("Reject null hypothesis and it is significant"
29         )
29   } else
30   {
31     cat("Cannot reject null hypothesis and no
32         significance")
32   }

```

R code Exa 11.4.2 Chi Square Testing

```

1 # Page.No: 11.6 – 11.7
2
3 condition <- c("Heads", "Tails")
4 observed <- c(50, 130)
5 expected <- c(90, 90)
6
7 DF <- data.frame(condition, observed, expected)
8 View(DF)
9
10 alpha5 <- 0.05 # 5% significance level
11 alpha1 <- 0.01 # 1% significance level
12 df <- nrow(DF) - 1 # Degree of freedom
13
14 chi_sq_val_5 <- round(qchisq(alpha5, df, lower.tail =
    F), 2) # Chi-square value

```

```

15 chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F
    ), 2)
16
17 diff <- DF$observed - DF$expected # Difference
18 diff_sq <- diff ** 2 # Difference squared
19 ans <- round(diff_sq / DF$expected, 2) #
    Difference squared weighted by expected frequency
20
21 chi_sq <- sum(ans) # Chi-square
22
23 cat("Value of chi-square is", chi_sq)
24
25 if(chi_sq > chi_sq_val_5)
26   if(chi_sq > chi_sq_val_1)
27   {
28     cat("Reject null hypothesis and it is significant"
        )
29   } else
30   {
31     cat("Cannot reject null hypothesis and no
        significance")
32   }
33
34 "The answer provided in the text book is wrong"

```

R code Exa 11.4.3 Chi Square Testing

```

1 # Page no. : 11.7 - 11.8
2
3 condition <- c("Good", "Average", "Poor")
4 observed <- c(20,24,6)
5 expected <- c(8,34,8)
6
7 DF <- data.frame(condition, observed, expected)
8 View(DF)

```

```

9
10 alpha5 <- 0.05 # 5% significance level
11 alpha1<-0.01 #1% significance level
12 df <- nrow(DF) - 1 # Degree of freedom
13
14 chi_sq_val_5 <- round(qchisq(alpha5,df,lower.tail =
      F), 2) # Chi-square value
15 chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F
      ), 2)
16
17 diff <- DF$observed - DF$expected # Difference
18 diff_sq <- diff ** 2 # Difference squared
19 ans <- round(diff_sq / DF$expected, 2) #
      Difference squared weighted by expected frequency
20
21 chi_sq <- sum(ans) # Chi-square
22
23 cat("Value of chi-square is", chi_sq)
24
25 if(chi_sq > chi_sq_val_5)
26   if(chi_sq > chi_sq_val_1)
27   {
28     cat("Reject null hypothesis and it is significant"
          )
29   } else
30   {
31     cat("Cannot reject null hypothesis and no
          significance")
32   }

```

R code Exa 11.4.4 Chi Square Testing Equal Probability Cases

```

1 # Page no. : 11.7 - 11.8
2
3 condition <- c("Good", "Average", "Bad")

```

```

4 observed <- c(25,60,35)
5 expected <- c(40,40,40)
6
7 DF <- data.frame(condition, observed, expected)
8 View(DF)
9
10 alpha5 <- 0.05 # 5% significance level
11 alpha1<-0.01 #1% significance level
12 df <- nrow(DF) - 1 # Degree of freedom
13
14 chi_sq_val_5 <- round(qchisq(alpha5,df,lower.tail =
      F), 2) # Chi-square value
15 chi_sq_val_1<- round(qchisq(alpha1,df,lower.tail = F
      ), 2)
16
17 diff <- DF$observed - DF$expected # Difference
18 diff_sq <- diff ** 2 # Difference squared
19 ans <- round(diff_sq / DF$expected, 2) #
      Difference squared weighted by expected frequency
20
21 chi_sq <- sum(ans) # Chi-square
22
23 cat("Value of chi-square is", chi_sq)
24
25 if(chi_sq > chi_sq_val_5)
26   if(chi_sq > chi_sq_val_1)
27   {
28     cat("Reject null hypothesis and it is significant"
29       )
29   } else
30   {
31     cat("Cannot reject null hypothesis and no
32       significance")

```

R code Exa 11.4.7 Chi Square of Independence

```
1 # Page no. : 11.10 - 11.11
2
3 condition <- c("First", "Second", "Third")
4 matrix <- matrix(c(15, 25, 10, 20, 25, 5), 2, 3, byrow =
  TRUE)
5 expected <- c((sum(matrix[, 1]) * 50) / 100, (sum(matrix
  [, 2]) * 50) / 100,
6             (sum(matrix[, 3]) * 50) / 100, (sum(matrix[, 1]) *
  50) / 100,
7             (sum(matrix[, 2]) * 50) / 100, (sum(matrix[, 3]) *
  50) / 100)
8 expected
9 observed <- c(15, 25, 10, 20, 25, 5)
10 DF <- data.frame(condition, observed, expected)
11 View(DF)
12
13 alpha <- 0.05 # 5% significance level
14 df <- nrow(DF) - 1 # Degree of freedom
15
16 chi_sq_val <- round(qchisq(alpha, df, lower.tail = F),
  2) # Chi-square value
17
18 diff <- DF$observed - DF$expected # Difference
19 diff_sq <- diff ** 2 # Difference squared
20 ans <- round(diff_sq / DF$expected, 2) #
  Difference squared weighted by expected frequency
21
22 chi_sq <- sum(ans) # Chi-square
23
24 cat("Value of chi-square is", chi_sq)
25
26
27 if(chi_sq > chi_sq_val)
28 {
29   cat("Reject null hypothesis and it is
  significant")
}
```

```

30 } else
31 {
32     cat("Cannot reject null hypothesis and no
        significance")
33 }
34
35 "Opinion of rural and Urban population differ
    significantly"

```

R code Exa 11.4.8 chi Square of 2 by 2 Table

```

1 #Page.no 11.11 - 11.12
2
3 R1 = c(20,10)
4 R2 = c(5,15)
5 rows = 2
6
7 #2*2 contingency table
8 table = matrix(c(R1, R2),
9                 nrow=rows,
10                byrow=TRUE)
11
12 rownames(table) = c("Good Achievement", "Bad
    achievement")# Naming the rows and
13 colnames(table) = c("High", "Low") # columns is
    optional.
14 table
15
16 chi_sq<-chisq.test(table,correct=FALSE)
17
18 chi_sq<-chi_sq$statistic
19
20 ## Chi-squared table value @ 5%
21 alpha <- 0.05 # 5% significance level
22 df <- nrow(table) - 1 # Degree of freedom

```

```
23
24 chi_sq_val <- round(qchisq(alpha,df,lower.tail = F),
25                       2) # Chi-square table value
26
27 cat("The calculated chi square value is ",chi_sq)
28
29 if(chi_sq > chi_sq_val)
30 {
31   cat("Reject null hypothesis and observed Chi-
32       square is significant")
33 } else
34 {
35   cat("Cannot reject null hypothesis and there is no
36       significance")
37 }
38
39 "so Interest has significantly affect the
40 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

```
1 #Page.No 12.9– 12.11
2
3 # A function created to find t-ratio
4 t.test <- function(m1,m2,s1,s2,n1,n2,r,equal.
   variance=FALSE)
5 {
6   {
7     se1 <-s1/sqrt(n1-1)
8     se2<-s2/sqrt(n2-1)
9     df <- n1-1
10    seD<-sqrt((se1^2)+(se2^2)-2*r*se1*se2)
11   }
12   t <- (m1-m2)/seD
13   dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
14   names(dat) <- c("Difference of means", "Std Error"
15     , "t", "p-value")
16   return(dat)
17 }
```

```

18 Mn.P<-70 # Mean of Pre- test
19 Mn.E<-67 # Mean Of end - test
20 SD.P<-6 # Sd of pre- test
21 SD.E<-5.8 # sd of end- test
22 N.P<-30 # No.of.Observations of Pre- test
23 N.E<-30 # No.of.Observations of end- test
24 r<-.82 # correlation coefficient
25
26 t.test(Mn.P,Mn.E,SD.E,SD.P,N.P,N.E,r)
27
28 critical_t<-abs(qt(0.01/2,29))
29
30 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")
31
32 " Therefore the mean difference is significant @ 0.01
      level of significance"

```

R code Exa 12.9.3.2 Students t Test Correlated Large Sample

```

1 # Page.No 12.11 - 12.13
2
3 # A function created to find t-ratio
4 t.test <- function(m1,m2,s1,s2,n1,n2,r,equal.
      variance=FALSE)
5 {
6   {
7     se1 <-s1/sqrt(n1-1)
8     se2<-s2/sqrt(n2-1)
9     df <- n1-1
10    seD<-sqrt((se1^2)+(se2^2)-2*r*se1*se2)
11  }
12  t <- (m1-m2)/seD
13  dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))

```

```

14  names(dat) <- c("Difference of means", "Std Error"
    , "t", "p-value")
15  return(dat)
16 }
17
18
19 Mn.P<-88 # Mean of Pre- test
20 Mn.E<-85 # Mean Of end - test
21 SD.P<-16 # Sd of pre- test
22 SD.E<-12 # sd of end- test
23 N.P<-64 # No.of.Observations of Pre- test
24 N.E<-64 # No.of.Observations of end- test
25 r<-.5 # correlation coefficient
26
27 t.test(Mn.P,Mn.E,SD.E,SD.P,N.P,N.E,r)
28
29 ct<-abs(qt(0.05/2,63))
30
31 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")
32 "The mean difference is significant"

```

R code Exa 12.9.3.3 Students t Test Correlated Small Sample

```

1 #Page.no 12.13
2
3 x<-c(40,36,35,34,40,28,31,34,37,25)
4 y<-c(50,42,51,46,35,52,68,51,84,63)
5
6 t<-t.test(x, y, paired = TRUE, alternative = "two.
    sided",conf.level = 0.99)
7
8 degrees_of_freedom<-t$parameter # Degrees of
    freedom

```

```

9
10 t_ratio<-abs(t$statistic) # t calculated value
11
12 critical_t<-abs(qt(0.01/2,degrees_of_freedom)) # t
    table value
13
14 p<-t$p.value # P value
15
16 cat("The p value is",p,"which is less than 0.01")
17
18 if(t_ratio > critical_t)
19 {
20     cat("Reject null hypothesis and it is
        significant")
21 } else
22 {
23     cat("Cannot reject null hypothesis and no
        significance")
24 }
25
26 "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

```

1 #Page.no 12.15–12.16
2
3 x<-c(16,18,20,24,24,22,20,18,10,8,20)
4 y<-c(24,20,24,28,30,20,24,22,18,18,24)
5
6 t<-t.test(x, y, paired = TRUE, alternative = "two.
    sided",conf.level = 0.99)
7
8 degrees_of_freedom<-t$parameter # Degrees of
    freedom

```

```

9
10 t_ratio<-t$statistic # t calculated value
11
12 critical_t<-abs(qt(0.01/2,degrees_of_freedom)) # t
    table value
13
14 p<-t$p.value # P value
15
16 cat("The calculated t value is",t_ratio)
17
18 if(t_ratio > critical_t)
19 {
20   cat("Reject null hypothesis and it is significant"
    )
21 } else
22 {
23   cat("Cannot reject null hypothesis and
    insignificant")
24 }
25
26 "Therefore, the gain from viewing the film is
    insignificant"

```

R code Exa 12.9.3.5 Students t Test Uncorrelated Large Sample

```

1 #Page.No 12.16– 12.18
2
3 # A function created to find t-ratio
4 t.test <- function(m1,m2,s1,s2,n1,n2,equal.variance=
    FALSE)
5 {
6   {
7     se1 <-s1/sqrt(n1-1)
8     se2<-s2/sqrt(n2-1)
9     df <- n1+n2-1

```

```

10     seD<-sqrt((se1^2)+(se2^2))
11   }
12   t <- (m1-m2)/seD
13   dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
14   names(dat) <- c("Difference of means", "Std Error"
15     , "t", "p-value")
16   return(dat)
17 }
18 t.test(62,57,9.7,6.8,30,25)
19
20 ct<-abs(qt(0.05/2,53))
21 cat("Since the p-value <0.05 and the table value of
22   t @ 0.05 level of significance is ",ct," which is
23   less than the obtained t-ratio",2.199)
24 "Therefore the mean difference is statistically
25   significant @ 0.05 level of significance"
26
27 ##### Another Method by using the library function
28
29 library(BSDA)
30 result<-tsum.test(mean.x = 62,s.x = 9.7,n.x = 30,
31                   mean.y = 57,s.y = 6.8,n.y = 25)
32
33 t_ratio<-result$statistic
34
35 cat("The calculated t value is",t_ratio)
36
37 critical_t<-abs(qt(0.05/2,53))
38
39 if(t_ratio > critical_t)
40 {
41   cat("Reject null hypothesis and it is significant"
42     )
43 } else
44 {
45   cat("Cannot reject null hypothesis and their is No
46     significance")

```

42 }

R code Exa 12.9.3.6 Students t Test Uncorrelated Large Sample

```
1 #Page.No 12.18– 12.20
2
3 # A function created to find t-ratio
4 t.test <- function(m1,m2,s1,s2,n1,n2,equal.variance=
      FALSE)
5 {
6   {
7     se1 <-s1/sqrt(n1-1)
8     se2<-s2/sqrt(n2-1)
9     df <- n1+n2-1
10    seD<-sqrt((se1^2)+(se2^2))
11   }
12   t <- (m1-m2)/seD
13   dat <- c(m1-m2, seD, t, 2*pt(-abs(t),df))
14   names(dat) <- c("Difference of means", "Std Error"
      , "t", "p-value")
15   return(dat)
16 }
17
18 t.test(100,105,10,10,50,50)
19 n1<-50
20 n2<-50
21 df<-n1+n2-1
22
23 ct<-abs(qt(0.05/2,df))
24
25 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)
26 "Therefore the mean difference is statistically
      significant @ 0.05 level of significance"
```

```

27
28 ##### Another Method by using the library function
29
30 library(BSDA)
31 result<-tsum.test(mean.x = 100,s.x = 10,n.x = 50,
32                   mean.y = 105,s.y = 10,n.y = 50)
33
34 t_ratio<-abs(result$statistic)
35
36 critical_t<-abs(qt(0.05/2,53))
37
38 if(t_ratio > critical_t)
39 {
40   cat("Reject null hypothesis and it is significant"
41       )
42 } else
43 {
44   cat("Cannot reject null hypothesis and their is No
45       significance")
46 }

```

R code Exa 12.9.3.7 Students t Test Uncorrelated Small Sample

```

1 # Page.No 12.20 –12.23
2
3
4 boys<-c(59,43,37,25,31,36,33,41,20,18)
5 girls<-c(57,43,44,39,21,18,15,55,40,30)
6
7 library(BSDA)
8 result<-tsum.test(mean.x = mean(boys),s.x = sd(boys)
9                   ,n.x = length(boys),
10                   mean.y = mean(girls),s.y = sd(
11                       girls),n.y = length(girls))

```

```

11 t_ratio<-abs(result$statistic) # T calculated
    value
12
13 cat("The calculated t value is",t_ratio)
14
15 df<-length(girls)+length(boys)-2 # degrees of
    freedom
16
17 p_value<-result$p.value #P value
18
19 critical_t<-abs(qt(0.05/2,df)) # t table value
20
21 if(t_ratio > critical_t)
22 {
23   cat("Reject null hypothesis and it is significant"
24     )
25 } else
26 {
27   cat("Cannot reject null hypothesis and their is No
28     significance")
29 }
30
31 "Thus their is no significant difference between the
    mean intelligence scores of boys and girls"
32
33 cat("p value is greater than 0.05 i.e ",p_value,"
34   null hypothesis retained")

```

R code Exa 12.9.3.8 Students t Test Uncorrelated Small Sample

```

1 #Page.No 12.23– 12.26
2
3 control_group<-c(11,11,10,12,8,15,10,8,10,8)
4 experimental_group<-c(4,4,8,9,12,15,3,13,9,9)
5

```

```

6 library(BSDA)
7 result<-tsum.test(mean.x = mean(control_group),s.x =
      sd(control_group),n.x = length(control_group),
8      mean.y = mean(experimental_group),s.y = sd
      (experimental_group),n.y = length(
      experimental_group))
9
10 t_ratio<-result$statistic # T calculated value
11
12 cat("The calculated tvalue is ",t_ratio)
13
14 df<-length(control_group)-1 # degrees of freedom
15
16 p_value<-result$p.value #P value
17
18 critical_t<-abs(qt(0.05/2,df)) # t table value
19
20 if(t_ratio > critical_t)
21 {
22   cat("Reject null hypothesis and it is significant"
23     )
24 } else
25 {
26   cat("Cannot reject null hypothesis and their is No
27     significance")
28 }
29 "Thus their is no significant difference between the
30   mean score of control group and experimental
31   group"
32 cat("p value is greater than 0.05 i.e",p_value,"null
33   hypothesis retained")

```

R code Exa 12.9.3.9 Students t Test Uncorrelated Small Sample

```

1 # Page.No 12.26 - 12.29
2
3 GroupX<-c(26,24,18,17,18,20,18)
4 GroupY<-c(38,26,24,24,30,22)
5
6 library(BSDA)
7 result<-tsum.test(mean.x = mean(GroupX),s.x = sd(
      GroupX),n.x = length(GroupX),
8                      mean.y = mean(GroupY),s.y = sd(
      GroupY),n.y = length(GroupY))
9
10 t_ratio<-abs(result$statistic) # T calculated value
11
12 cat("The calculated t value is",t_ratio)
13
14 df<-length(GroupX)+length(GroupY)-1 # degrees of
      freedom
15
16 p_value<-result$p.value #P value
17
18 critical_t<-abs(qt(0.05/2,df)) # t table value
19
20 if(t_ratio > critical_t)
21 {
22   cat("Reject null hypothesis and it is significant"
23     )
24 } else
25 {
26   cat("Cannot reject null hypothesis and their is No
27     significance")
28 }
29
30 "Thus their is significant difference between Group
31   X and Group Y"
32 cat("p value is less than 0.05 i.e",p_value,"null
33   hypothesis is rejected")

```

Chapter 13

Analysis of Variance

R code Exa 13.8.1 One Way Anova

```
1 #Page.No 13.7 – 13.9
2
3 lecture<-c(10,12,13,13,14)
4 seminar<-c(13,15,15,17,18)
5 discussion<-c(7,7,10,11,12)
6
7 combined_groups<-data.frame(cbind(lecture,seminar,
8     discussion))
9 stackedgroup<-stack(combined_groups)
10 anova_results<-aov(values~ind,data = stackedgroup)
11 summary(anova_results)
12 "The ANOVA shows that there is significant
    differences among the three methods of
    instructions"
```

R code Exa 13.8.2 Deviation Score Method

```

1 #Page.No 13.10 – 13.12
2
3 lecture<-c(10,12,13,13,14)
4 seminar<-c(13,15,15,17,18)
5 discussion<-c(7,7,10,11,12)
6
7 combined_groups<-data.frame(cbind(lecture,seminar,
      discussion))
8 stackedgroup<-stack(combined_groups)
9 anova_results<-aov(values~ind,data = stackedgroup)
10 summary(anova_results)
11
12 "The ANOVA shows that there is significant
      differences among the three methods of
      instructions"

```

R code Exa 13.8.3 Anova

```

1 #Page.No 13.7 – 13.9
2
3 Group1<-c(6,7,9,10,8)
4 Group2<-c(11,10,8,12,9)
5
6 combined_groups<-data.frame(cbind(Group1,Group2))
7 stackedgroup<-stack(combined_groups)
8 anova_results<-aov(values~ind,data = stackedgroup)
9 summary(anova_results)
10
11 "The ANOVA shows that there is significant
      differences among the three methods of
      instructions"
12
13 # To find the t-value
14 library(BSDA)
15

```

```

16 result<-tsum.test(mean.x = mean(Group1),s.x = sd(
      Group1),n.x = length(Group1),
17                    mean.y = mean(Group2),s.y = sd(
      Group2),n.y = length(Group2))
18
19 t<-abs(result$statistic)
20
21 t^2 ->F # From the anova_results F= 4

```

R code Exa 13.8.4 Two Way ANOVA

```

1 #Page.No 13.14 – 13.17
2
3 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)
4 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)
5 activity<-c
      (1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3)
6 activity<-as.factor(activity)
7 specialist<-as.factor(specialist)
8 combined_groups<-data.frame(cbind(activity,
      specialist,performance))
9
10 # Two- way anova
11 anova_results<-aov(performance~ specialist+activity+
      specialist*activity) # anova model specification
12 summary(anova_results)
13
14 # Interpretation from anova table p_value
15
16 " It can attributed that the co-curricular

```

```

    activities have differ significantly in their
    relation with selection of the specialist"
17
18 # Using F value
19 # Interpretation on Specialist
20
21 Fcalculated<-1.716
22
23 # at 0.05 level of significance
24
25 Ftable_0.05<-qf(.95, df1=1, df2=24)
26
27 if(Fcalculated > Ftable_0.05)
28 {
29   cat("Reject null hypothesis and it is significant"
30       )
31 } else
32 {
33   cat("Cannot reject null hypothesis and their is No
34       significance")
35 }
36
37 # At 0.01 level of significance
38
39 Ftable_0.01<-qf(.99, df1=2, df2=24)
40
41 if(Fcalculated > Ftable_0.01)
42 {
43   cat("Reject null hypothesis and it is significant"
44       )
45 } else
46 {
47   cat("Cannot reject null hypothesis and their is No
48       significance")
49 }
50
51 ### interpretation on activities

```

```

49 Fcalculated<-13.319
50
51 Ftable_0.05<-qf(.95, df1=2, df2=24)
52
53 if(Fcalculated > Ftable_0.05)
54 {
55   cat("Reject null hypothesis and it is significant"
56       )
57 } else
58 {
59   cat("Cannot reject null hypothesis and their is No
60       significance")
61 }
62
63 ## Interpretation on interaction effect
64 # at 0.05 level of significance
65 Fcalculated<-2.397
66
67 Ftable_0.05<-qf(.95, df1=2, df2=24)
68
69 if(Fcalculated > Ftable_0.05)
70 {
71   cat("Reject null hypothesis and it is significant"
72       )
73 } else
74 {
75   cat("Cannot reject null hypothesis and their is No
76       significance")
77 }
78 # at 0.01 level of significance
79 Ftable_0.01<-qf(.99, df1=2, df2=24)
80
81 if(Fcalculated > Ftable_0.01)
82 {

```

```
83   cat("Reject null hypothesis and it is significant"
      )
84 } else
85 {
86   cat("Cannot reject null hypothesis and their is No
        significance")
87 }
```

Chapter 14

Analysis of Covariance

R code Exa 14.4.1 ANCOVA

```
1 # Page.No 14.6 – 14.12
2
3 X<-c(5,6,3,2,4,4,8,5,6,2,6,5,6,2,1)
4
5 Y<-c(6,7,5,4,3,8,7,6,5,4,4,6,5,2,3)
6
7 #ANOVA table for X
8 X1<-c(5,6,3,2,4)
9 X2<-c(4,8,5,6,2)
10 X3<-c(6,5,6,2,1)
11 combined_groups<-data.frame(cbind(X1,X2,X3))
12 stackedgroup<-stack(combined_groups)
13 anova_results_X<-aov(values~ind,data = stackedgroup)
14 summary(anova_results_X)
15
16 # ANOVA table for Y
17 Y1<-c(6,7,5,4,3)
18 Y2<-c(8,7,6,5,4)
19 Y3<-c(4,6,5,2,3)
20 combined_groups<-data.frame(cbind(Y1,Y2,Y3))
21 stackedgroup<-stack(combined_groups)
```

```

22 anova_results_Y<-aov(values~ind,data = stackedgroup)
23 summary(anova_results_Y)
24
25 table_value_1<-qf(.99,2,12)
26 table_value_5<-qf(.95,2,12)
27
28 ## From the calculated F value and critical value @
    0.05 and 0.01
29 ## There is no difference among covariates of X
30
31 ##### calculation of anova table by another method
32
33
34 sum(X)
35 sum(Y)
36 sum(X^2)
37 sum(Y^2)
38 sum(X*Y)
39
40 N<-length(X1)
41 sum_of_N<-length(X)
42
43 #Step 1 correction
44 Cx<-(sum(X))^2/sum_of_N
45 Cy<-(sum(Y)^2)/sum_of_N
46 Cxy<-sum(X)*sum(Y)/sum_of_N
47
48 #Step 2 TotalSS
49 TSSx<-sum(X^2)-Cx
50 TSSy<-sum(Y^2)-Cy
51 TSSxy<-sum(X*Y)-Cxy # Cx<-(sum(X))^2/N
52
53 #step 3 Between means
54 Bet_Mn_X<-(sum(X1)^2+sum(X2)^2+sum(X3)^2)/N-Cx
55 Bet_Mn_Y<-(sum(Y1)^2+sum(Y2)^2+sum(Y3)^2)/N-Cy
56 Bet_Mn_XY<-(sum(X1)*sum(Y1)+sum(X2)*sum(Y2)+sum(X3)*
    sum(Y3))/N-Cxy
57

```

```

58 # step 4
59 # Within Group SS
60 WSS_x<-TSSx-Bet_Mn_X
61 WSS_Y<-TSSy-Bet_Mn_Y
62 WSS_XY<-TSSxy-Bet_Mn_XY
63
64 # Degrees of freedom
65 Bdf<-3-1 # 3 groups
66 Tdf<-length(X)-1
67 Wdf<-Tdf-Bdf
68 df<-c(Bdf ,Wdf ,Tdf)
69
70 #sum of squares
71 SSx<-c(Bet_Mn_X,WSS_x,TSSx)
72 SSy<-c(Bet_Mn_Y,WSS_Y,TSSy)
73
74 # Mean sum of squares
75 MSx<-SSx/df
76 MSy<-SSy/df
77
78 Anova_table<-matrix(c(df,SSx,SSy,MSx,MSy),3,5)
79
80 # step 6 computing adjusted sum of squares for Y
81
82 TSSy.x<-TSSy-(TSSxy^2)/TSSx
83 WSSy.x<-WSS_Y-(WSS_XY^2)/WSS_x
84 BSSy.x<-TSSy.x-WSSy.x
85
86 SSy.x<-c(BSSy.x,WSSy.x,TSSy.x)
87 SSx.y<-c(BSSy.x,WSS_XY,TSSxy)
88
89 df<-c(Bdf,Wdf-1,Tdf-1)
90 MSy.x<-SSy.x/df
91
92 #Step 7 ANCOVA
93
94 matrix(c(df,SSx,SSy,SSx.y,SSy.x,MSy.x),3,6) # Ancova
      table

```

```

95
96 Fy.x<-2.94/1.72 # from the ANCOVA table
97
98 Ftable<-qf(.95,2,13)
99
100 # Since F calculated value less than F table value
    Accept null hypothesis
101 # The computation of adjusted SS for y is not
    significant @ 0.05
102
103 SDy.x<-sqrt(WSSy.x/11)
104 #From comparison both are not significant at 0.05
    level
105
106 ## Step 8 Adjusted y means
107
108 Mx<-c(mean(X1),mean(X2),mean(X3))
109 My<-c(mean(Y1),mean(Y2),mean(Y3))
110 GMx<-mean(Mx)
111 r_total<-cor(X,Y)
112 r_between<-Bet_Mn_XY/sqrt(Bet_Mn_X*Bet_Mn_Y)
113 r_within<-WSS_XY/sqrt(WSS_x*WSS_Y)
114 bx<-WSS_XY/WSS_x
115
116 M1<-mean(Y1)-bx*(mean(X1)-GMx)
117 M2<-mean(Y2)-bx*(mean(X2)-GMx)
118 M3<-mean(Y3)-bx*(mean(X3)-GMx)
119
120 My.x<-c(M1,M2,M3)
121
122 Table<- matrix(c(Mx,My,My.x),3,3)
123 print(Table)
124
125 # Step 9 Comparison of Adjusted Means
126
127 SDy.x
128 SEmd<-SDy.x*sqrt(1/length(X1)+1/length(X2))
129

```

```
130 # t_value
131
132 (M1-M2)/SEmd
133
134 (M1-M3)/SEmd
135
136 (M2-M3)/SEmd
137
138 # All pairs are not significant
139 cat("There is high correlation",r_between,r_within,r
      _total)
140
141 "hence Group 1 , Group 2 , Group 3 not differ
      significantly "
```
