R Textbook Companion for Biostatistics: Basic Concepts and Methodology for the Health Sciences by Daniel W. Wayne, Chad L. Cross[1]

Created by
Anjana Rajagopal
M.Sc.
Others
Pondicherry University
Cross-Checked by
R TBC Team

June 8, 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: Biostatistics: Basic Concepts and Methodology for the Health Sciences

Author: Daniel W. Wayne, Chad L. Cross

Publisher: John Wiley & Sons, Singapore

Edition: 10

Year: 2014

ISBN: 978-81-265-5189-7

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 1

Getting acquainted with biostatistics

R code Exa 1.4.1 Simple random sampling of size 10 Page 8

```
##Example 1.4.1 Pg.8
##Simple random sampling of size 10

age <- c(48,35,46,44,43,42,39,44,49,49,
         44,39,38,49,49,53,56,57,51,61,
         53,66,71,75,72,65,67,38,37,46,
         44,44,48,49,30,45,47,45,48,47,
         47,44,48,43,45,40,48,49,38,44,
         43,47,46,57,52,54,56,53,64,53,
         58,54,59,56,62,50,64,53,61,53,
         62,57,52,54,61,59,57,52,54,53,
         62,52,62,57,59,59,56,57,53,59,
         61,55,61,56,52,54,51,50,50,55,
         63,50,59,54,60,50,56,68,66,71,
         82,68,78,66,70,66,78,69,71,69,
         78,66,68,71,69,77,76,71,43,47,
         48,37,40,42,38,49,43,46,34,46,
         46,48,47,43,52,53,61,60,53,53,
         50,53,54,61,61,61,64,53,53,54,
```

```
            61,60,51,50,53,64,64,53,60,54,
            55,58,62,62,54,53,61,54,51,62,
            57,50,64,63,65,71,71,73,66)

length(age)
set.seed(12)
srs = sample(age,10)
srs
##Answers change because of random sampling
```

R code Exa 1.4.2 Systematic sample of size 10 Page 11

```
##Example 1.4.2 Pg.11
##Systematic sample of size 10

age <- c(48,35,46,44,43,42,39,44,49,49,
            44,39,38,49,49,53,56,57,51,61,
            53,66,71,75,72,65,67,38,37,46,
            44,44,48,49,30,45,47,45,48,47,
            47,44,48,43,45,40,48,49,38,44,
            43,47,46,57,52,54,56,53,64,53,
            58,54,59,56,62,50,64,53,61,53,
            62,57,52,54,61,59,57,52,54,53,
            62,52,62,57,59,59,56,57,53,59,
            61,55,61,56,52,54,51,50,50,55,
            63,50,59,54,60,50,56,68,66,71,
            82,68,78,66,70,66,78,69,71,69,
            78,66,68,71,69,77,76,71,43,47,
            48,37,40,42,38,49,43,46,34,46,
            46,48,47,43,52,53,61,60,53,53,
            50,53,54,61,61,61,64,53,53,54,
            61,60,51,50,53,64,64,53,60,54,
            55,58,62,62,54,53,61,54,51,62,
            57,50,64,63,65,71,71,73,66)

```

```
sys.sample = function(N,n,r){
  k = round(N/n)
  #ceiling(x) rounds to the nearest integer that's
     larger than x.
  sys.samp = seq(r, r + k*(n-1), k)
  print(sys.samp)
}

N = 185; n=10; r=4
sys = sys.sample(N, n, r)
age[sys]
```

Chapter 2

Strategies for understanding the meanings of data

R code Exa 2.2.1 Arrange ages from smallest to largest Page 20

```
##Example 2.2.1 Pg.20
##Arrange ages from smallest to largest

age <- c(48,35,46,44,43,42,39,44,49,49,
         44,39,38,49,49,53,56,57,51,61,
         53,66,71,75,72,65,67,38,37,46,
         44,44,48,49,30,45,47,45,48,47,
         47,44,48,43,45,40,48,49,38,44,
         43,47,46,57,52,54,56,53,64,53,
         58,54,59,56,62,50,64,53,61,53,
         62,57,52,54,61,59,57,52,54,53,
         62,52,62,57,59,59,56,57,53,59,
         61,55,61,56,52,54,51,50,50,55,
         63,50,59,54,60,50,56,68,66,71,
         82,68,78,66,70,66,78,69,71,69,
         78,66,68,71,69,77,76,71,43,47,
         48,37,40,42,38,49,43,46,34,46,
         46,48,47,43,52,53,61,60,53,53,
         50,53,54,61,61,61,64,53,53,54,
```

```
           61,60,51,50,53,64,64,53,60,54,
           55,58,62,62,54,53,61,54,51,62,
           57,50,64,63,65,71,71,73,66)
sort(age)
```

R code Exa 2.3.1 Form Class Intervals for the ages data Page 23

```
##Example 2.3.1 Pg.23
##Form Class Intervals for the ages data

age <- c(48,35,46,44,43,42,39,44,49,49,
           44,39,38,49,49,53,56,57,51,61,
           53,66,71,75,72,65,67,38,37,46,
           44,44,48,49,30,45,47,45,48,47,
           47,44,48,43,45,40,48,49,38,44,
           43,47,46,57,52,54,56,53,64,53,
           58,54,59,56,62,50,64,53,61,53,
           62,57,52,54,61,59,57,52,54,53,
           62,52,62,57,59,59,56,57,53,59,
           61,55,61,56,52,54,51,50,50,55,
           63,50,59,54,60,50,56,68,66,71,
           82,68,78,66,70,66,78,69,71,69,
           78,66,68,71,69,77,76,71,43,47,
           48,37,40,42,38,49,43,46,34,46,
           46,48,47,43,52,53,61,60,53,53,
           50,53,54,61,61,61,64,53,53,54,
           61,60,51,50,53,64,64,53,60,54,
           55,58,62,62,54,53,61,54,51,62,
           57,50,64,63,65,71,71,73,66)

breaks = seq(30,90,by=10)
breaks
CI = cut(age,breaks = breaks,right=F)
CI
table(CI) #gives class interevals along with
```

```
frequencies
hist(age,breaks = breaks,main = "Histogram of ages of 189 subjects") #Pg.27
par(new=T) ##overlaps new plot
plot(table(CI),type="b") #Pg.28
dev.off()
```

R code Exa 2.3.2 Stem and leaf plot for ages of 189 subjects Page 29

```

##Example 2.3.2 Pg.29
##Stem and leaf plot for ages of 189 subjects

age <- c(48,35,46,44,43,42,39,44,49,49,
         44,39,38,49,49,53,56,57,51,61,
         53,66,71,75,72,65,67,38,37,46,
         44,44,48,49,30,45,47,45,48,47,
         47,44,48,43,45,40,48,49,38,44,
         43,47,46,57,52,54,56,53,64,53,
         58,54,59,56,62,50,64,53,61,53,
         62,57,52,54,61,59,57,52,54,53,
         62,52,62,57,59,59,56,57,53,59,
         61,55,61,56,52,54,51,50,50,55,
         63,50,59,54,60,50,56,68,66,71,
         82,68,78,66,70,66,78,69,71,69,
         78,66,68,71,69,77,76,71,43,47,
         48,37,40,42,38,49,43,46,34,46,
         46,48,47,43,52,53,61,60,53,53,
         50,53,54,61,61,61,64,53,53,54,
         61,60,51,50,53,64,64,53,60,54,
         55,58,62,62,54,53,61,54,51,62,
         57,50,64,63,65,71,71,73,66)
stem(age,scale=0.5)
```

R code Exa 2.4.1 Obtain the mean age page 38

```
##Example 2.4.1 Pg.38
##Obtain the mean age

age <- c(48,35,46,44,43,42,39,44,49,49,
         44,39,38,49,49,53,56,57,51,61,
         53,66,71,75,72,65,67,38,37,46,
         44,44,48,49,30,45,47,45,48,47,
         47,44,48,43,45,40,48,49,38,44,
         43,47,46,57,52,54,56,53,64,53,
         58,54,59,56,62,50,64,53,61,53,
         62,57,52,54,61,59,57,52,54,53,
         62,52,62,57,59,59,56,57,53,59,
         61,55,61,56,52,54,51,50,50,55,
         63,50,59,54,60,50,56,68,66,71,
         82,68,78,66,70,66,78,69,71,69,
         78,66,68,71,69,77,76,71,43,47,
         48,37,40,42,38,49,43,46,34,46,
         46,48,47,43,52,53,61,60,53,53,
         50,53,54,61,61,61,64,53,53,54,
         61,60,51,50,53,64,64,53,60,54,
         55,58,62,62,54,53,61,54,51,62,
         57,50,64,63,65,71,71,73,66)
mean(age)
```

R code Exa 2.4.2 Mean age of 10 subjects Page 39

```
##Example 2.4.2 Pg.39
##Mean age of 10 subjects

age <- c(48,35,46,44,43,42,39,44,49,49,
```

```
            44,39,38,49,49,53,56,57,51,61,
            53,66,71,75,72,65,67,38,37,46,
            44,44,48,49,30,45,47,45,48,47,
            47,44,48,43,45,40,48,49,38,44,
            43,47,46,57,52,54,56,53,64,53,
            58,54,59,56,62,50,64,53,61,53,
            62,57,52,54,61,59,57,52,54,53,
            62,52,62,57,59,59,56,57,53,59,
            61,55,61,56,52,54,51,50,50,55,
            63,50,59,54,60,50,56,68,66,71,
            82,68,78,66,70,66,78,69,71,69,
            78,66,68,71,69,77,76,71,43,47,
            48,37,40,42,38,49,43,46,34,46,
            46,48,47,43,52,53,61,60,53,53,
            50,53,54,61,61,61,64,53,53,54,
            61,60,51,50,53,64,64,53,60,54,
            55,58,62,62,54,53,61,54,51,62,
            57,50,64,63,65,71,71,73,66)
set.seed(12)
srs =  sample(age,10)
srs
mean(srs)   ##different answer due to a different
   random sample
```

R code Exa 2.4.3 Median age Page 40

```
##Example 2.4.3 Pg.40
##Median age

age <- c(48,35,46,44,43,42,39,44,49,49,
            44,39,38,49,49,53,56,57,51,61,
            53,66,71,75,72,65,67,38,37,46,
            44,44,48,49,30,45,47,45,48,47,
            47,44,48,43,45,40,48,49,38,44,
            43,47,46,57,52,54,56,53,64,53,
```

```
            58,54,59,56,62,50,64,53,61,53,
            62,57,52,54,61,59,57,52,54,53,
            62,52,62,57,59,59,56,57,53,59,
            61,55,61,56,52,54,51,50,50,55,
            63,50,59,54,60,50,56,68,66,71,
            82,68,78,66,70,66,78,69,71,69,
            78,66,68,71,69,77,76,71,43,47,
            48,37,40,42,38,49,43,46,34,46,
            46,48,47,43,52,53,61,60,53,53,
            50,53,54,61,61,61,64,53,53,54,
            61,60,51,50,53,64,64,53,60,54,
            55,58,62,62,54,53,61,54,51,62,
            57,50,64,63,65,71,71,73,66)
median(age)
```

R code Exa 2.4.4 Median age of sample of size 10 Page 40

```
##Example 2.4.4 Pg.40
##Median age of sample of size 10

age <- c(48,35,46,44,43,42,39,44,49,49,
            44,39,38,49,49,53,56,57,51,61,
            53,66,71,75,72,65,67,38,37,46,
            44,44,48,49,30,45,47,45,48,47,
            47,44,48,43,45,40,48,49,38,44,
            43,47,46,57,52,54,56,53,64,53,
            58,54,59,56,62,50,64,53,61,53,
            62,57,52,54,61,59,57,52,54,53,
            62,52,62,57,59,59,56,57,53,59,
            61,55,61,56,52,54,51,50,50,55,
            63,50,59,54,60,50,56,68,66,71,
            82,68,78,66,70,66,78,69,71,69,
            78,66,68,71,69,77,76,71,43,47,
            48,37,40,42,38,49,43,46,34,46,
            46,48,47,43,52,53,61,60,53,53,
```

```
          50,53,54,61,61,61,64,53,53,54,
          61,60,51,50,53,64,64,53,60,54,
          55,58,62,62,54,53,61,54,51,62,
          57,50,64,63,65,71,71,73,66)
set.seed(12)
srs = sample(age,10)
srs
median(srs)
##different answer due to a different random sample
```

R code Exa 2.4.5 Modal age Page 41

```

##Example 2.4.5 Pg.41
##Modal age

age <- c(48,35,46,44,43,42,39,44,49,49,
          44,39,38,49,49,53,56,57,51,61,
          53,66,71,75,72,65,67,38,37,46,
          44,44,48,49,30,45,47,45,48,47,
          47,44,48,43,45,40,48,49,38,44,
          43,47,46,57,52,54,56,53,64,53,
          58,54,59,56,62,50,64,53,61,53,
          62,57,52,54,61,59,57,52,54,53,
          62,52,62,57,59,59,56,57,53,59,
          61,55,61,56,52,54,51,50,50,55,
          63,50,59,54,60,50,56,68,66,71,
          82,68,78,66,70,66,78,69,71,69,
          78,66,68,71,69,77,76,71,43,47,
          48,37,40,42,38,49,43,46,34,46,
          46,48,47,43,52,53,61,60,53,53,
          50,53,54,61,61,61,64,53,53,54,
          61,60,51,50,53,64,64,53,60,54,
          55,58,62,62,54,53,61,54,51,62,
          57,50,64,63,65,71,71,73,66)
```

```
table(age) ##53 is repeated 17 times is the mode

mode = names(which(table(age)==max(table(age))))  ##
    returns the modal value
mode
```

R code Exa 2.4.6 Skewness Page 42

```
##Example 2.4.6 Pg.42
##Skewness

##install.packages("moments",dependencies = T)
library(moments)

no_skew = c
    (5,5,6,6,6,7,7,7,7,8,8,8,8,8,9,9,9,9,10,10,10,11,11)

right_skew = c
    (5,5,5,5,5,5,5,6,6,6,6,6,6,7,7,7,7,7,8,8,8,8,9,9,9,10,10,11)

left_skew = c
    (5,6,6,7,7,7,8,8,8,8,9,9,9,9,9,10,10,10,10,10,10,11,11,11,11,11,1

par(mfrow=c(1,3))
hist(no_skew,main="No skew")
hist(right_skew,main="Right skew")
hist(left_skew,main="Left skew")
dev.off()

descriptive <- function(x)
{
  cat("Mean = " ,mean(x), "\n")
  cat("Median = " ,median(x),"\n")
  cat("Mode = " ,names(which(table(x)==max(table(x))
     )),"\n")
```

```
  cat("Skewness = " ,skewness(x))
}
d1 = descriptive(no_skew)
d2 = descriptive(right_skew)
d3 = descriptive(left_skew)
```

R code Exa 2.5.1 Range of ages Page 44

```
##Example 2.5.1 Pg.44
##Range of ages

age <- c(48,35,46,44,43,42,39,44,49,49,
         44,39,38,49,49,53,56,57,51,61,
         53,66,71,75,72,65,67,38,37,46,
         44,44,48,49,30,45,47,45,48,47,
         47,44,48,43,45,40,48,49,38,44,
         43,47,46,57,52,54,56,53,64,53,
         58,54,59,56,62,50,64,53,61,53,
         62,57,52,54,61,59,57,52,54,53,
         62,52,62,57,59,59,56,57,53,59,
         61,55,61,56,52,54,51,50,50,55,
         63,50,59,54,60,50,56,68,66,71,
         82,68,78,66,70,66,78,69,71,69,
         78,66,68,71,69,77,76,71,43,47,
         48,37,40,42,38,49,43,46,34,46,
         46,48,47,43,52,53,61,60,53,53,
         50,53,54,61,61,61,64,53,53,54,
         61,60,51,50,53,64,64,53,60,54,
         55,58,62,62,54,53,61,54,51,62,
         57,50,64,63,65,71,71,73,66)
Range = diff(range(age))
Range
```

R code Exa 2.5.2 Variance of ages Page 44

```
##Example 2.5.2 Pg.44
##Variance of ages
age <- c(48,35,46,44,43,42,39,44,49,49,
         44,39,38,49,49,53,56,57,51,61,
         53,66,71,75,72,65,67,38,37,46,
         44,44,48,49,30,45,47,45,48,47,
         47,44,48,43,45,40,48,49,38,44,
         43,47,46,57,52,54,56,53,64,53,
         58,54,59,56,62,50,64,53,61,53,
         62,57,52,54,61,59,57,52,54,53,
         62,52,62,57,59,59,56,57,53,59,
         61,55,61,56,52,54,51,50,50,55,
         63,50,59,54,60,50,56,68,66,71,
         82,68,78,66,70,66,78,69,71,69,
         78,66,68,71,69,77,76,71,43,47,
         48,37,40,42,38,49,43,46,34,46,
         46,48,47,43,52,53,61,60,53,53,
         50,53,54,61,61,61,64,53,53,54,
         61,60,51,50,53,64,64,53,60,54,
         55,58,62,62,54,53,61,54,51,62,
         57,50,64,63,65,71,71,73,66)
set.seed(12)
srs = sample(age,10)
srs
var(srs)

##Answers differ because of a different random
   sample
```

R code Exa 2.5.3 Coefficient of correlation Page 46

```
##Example 2.5.3 Pg.46
##Coefficient of correlation
```

```
mean1 = 145  #sample of 11 years old
sd1 = 10
mean2 = 80   #sample of 25 years old
sd2 = 10

cv<-function(mean,sd)  #user defined function for
    coefficient of variation
{cv = sd*100/mean
print(cv)}

cv1 = cv(mean1,sd1)
cv2 = cv(mean2,sd2)

##variation is much higher in 11 year old tan in 25
    year old
```

R code Exa 2.5.4 Kurtosis Page 49

```
##Example 2.5.4 Pg.49
##Kurtosis

##install.packages("moments",dependencies = T)
library(moments)

meso = c
    (1,2,2,3,3,3,4,4,4,4,5,5,5,5,5,6,6,6,6,7,7,7,8,8,9)

lepto = c
    (1,2,2,3,3,3,4,4,4,4,4,5,5,5,5,5,5,5,5,5,5,5,5,6,6,6,6,6,7,7,7,

platy = c
    (1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,5,5,5,5,5,6,6,6,7,7,7,8,8,8,9,9,9)

par(mfrow=c(1,3))
```

```
hist(meso,main="Mesokurtic",breaks = 9)
hist(lepto,main="Leptokurtic",breaks=9)
hist(platy,main="Platykurtic",breaks=9)
dev.off()

descriptive <- function(x)
{
  cat("Mean = " ,mean(x), "\n")
  cat("Median = " ,median(x),"\n")
  cat("Mode = " ,names(which(table(x)==max(table(x))
    )),"\n")
  cat("Kurtosis = " ,kurtosis(x))
}
d1 = descriptive(meso)
d2 = descriptive(lepto)   ##Kurtosis>3
d3 = descriptive(platy)   ##Kurtosis<3
```

R code Exa 2.5.5 Box and whisker plot Page 50

```
## Example 2.5.5, Pg.50
##Refer Table 2.5.1
##Box and whisker plot

grf = c
   (14.6,24.3,24.9,27,27.2,27.4,28.2,28.8,29.9,30.7,31.5,31.6,

        32.3,32.8,33.3,33.6,34.3,36.9,38.3,44.0)
boxplot(grf, main="Box and whisker plot for GRF
   measurements")
```

Chapter 3

Probability The basis of statistical inference

R code Exa 3.4.1 Probability of member being less than 18 years of age Page 69

```
##Example 3.4.1 Pg.69
##Probability of member being <18 years of age

disorder <- c("negative","bipolar","unipolar","both"
    )
early <- c(28,19,41,53)
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq
Prob_E = sum(early)/(sum(early)+sum(later))  #no.of
    early subjects/total no. of subjects
Prob_E
```

R code Exa 3.4.2 Conditional Probability Page 70

```
##Example 3.4.2 Pg.70
##Conditional Probability
##A = event that family has no history of mood
    disorders
##E = event that subject is <18 years

disorder <- c("negative","bipolar","unipolar","both"
   )
early <- c(28,19,41,53)
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq

condi_AE = freq[1,2]/sum(early)  #frequency of
   negative-young subjects/total young subjects
condi_AE
```

R code Exa 3.4.3 Joint Probability of early subjects and no history of mood disorders Page 71

```
##Exammple 3.4.3 Pg.71
##Joint Probability of early subjects and no history
     of mood disorders

disorder <- c("negative","bipolar","unipolar","both"
   )
early <- c(28,19,41,53)
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq
prob_AE = freq[1,2]/(sum(early)+sum(later))  #
   frequency of negative-young subjects/total
   subjects
prob_AE
```

R code Exa 3.4.4 Multiplication rule of early subjects and no history of mood disorders Page 71

```
##Exammple 3.4.4 Pg.71
##Multiplication rule of early subjects and no
   history of mood disorders

disorder <- c("negative","bipolar","unipolar","both"
   )
early <- c(28,19,41,53)
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq
Prob_E = sum(early)/(sum(early)+sum(later))  #no.of
   early subjects/total no. of subjects
condi_AE = freq[1,2]/sum(early)  #frequency of
   negative-young subjects/total young subjects

Prob_AE = Prob_E*condi_AE   #using multiplication
   rule
Prob_AE
```

R code Exa 3.4.5 Conditional Probability of early subjects and no history of mood disorders Page 72

```
##Exammple 3.4.5 Pg.72
##Conditional Probability of early subjects and no
   history of mood disorders

disorder <- c("negative","bipolar","unipolar","both"
   )
early <- c(28,19,41,53)
```

```
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq
Prob_E = sum(early)/(sum(early)+sum(later))  #no.of early subjects/total no. of subjects
prob_AE = freq[1,2]/(sum(early)+sum(later))  #frequency of negative-young subjects/total subjects

Condi_AE = prob_AE/Prob_E   #using multiplication rule
Condi_AE
```

R code Exa 3.4.6 Probability of early subjects OR no history of mood disorders Page 73

```
##Exammple 3.4.6 Pg.73
##Probability of early subjects OR no history of mood disorders

disorder <- c("negative","bipolar","unipolar","both")
early <- c(28,19,41,53)
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq

Prob_E = sum(early)/(sum(early)+sum(later))  #no.of early subjects/total no. of subjects
Prob_A = (freq[1,2]+freq[1,3])/(sum(early)+sum(later))  #no.of negative subjects/total no. of subjects
prob_AE = freq[1,2]/(sum(early)+sum(later))  #frequency of negative-young subjects/total subjects
```

```
Prob_AUE = Prob_A + Prob_E - prob_AE  #By addition rule
Prob_AUE
```

R code Exa 3.4.7 Conditional Probability of student wears eye glasses given he is a boy Page 74

```
##Exammple 3.4.7 Pg.74
##Conditional Probability of student wears eyeglassses given he is a boy

girls = 60
boys = 40
girls_glasses = 24
boys_glasses = 16
prob_E = 40/100        #prob that a student wears eyeglasses
prob_B = 40/100        #prob of a boy
prob_EB = 16/100

condi_EB = prob_EB / prob_B  #conditional prob of a student wearing glasss given he is a boy
condi_EB

prob_EB = prob_E*prob_B   #joint probability
prob_EB
```

R code Exa 3.4.8 Find probability of admissions that are NOT private Page 75

```
##Example 3.4.8 Pg.75
```

```
##Find probability of admissions that are NOT
    private

N = 1200  #total no.of admissions
A = 750   #no. of private admissions
ProbA = A/N
ProbA
ProbA_bar = 1 - ProbA
ProbA_bar
```

R code Exa 3.4.9 Marginal Probability of Early age Page 75

```
###Exammple 3.4.9 Pg.75
##Marginal Probability of Early age

disorder <- c("negative","bipolar","unipolar","both"
    )
early <- c(28,19,41,53)
later <- c(35,38,44,60)
freq <- data.frame(disorder,early,later)
freq
prob_EA = freq[1,2]/(sum(early)+sum(later))
prob_EA
prob_EB = freq[2,2]/(sum(early)+sum(later))
prob_EB
prob_EC = freq[3,2]/(sum(early)+sum(later))
prob_EC
prob_ED = freq[4,2]/(sum(early)+sum(later))
prob_ED

prob_E = prob_EA + prob_EB + prob_EC + prob_ED  #
    Marginal Probability
prob_E
```

R code Exa 3.5.1 Bayes theorem Refer Table for data Page 81

```
###Exammple 3.5.1 Pg.81
##Bayes theorem Refer Table for data

Yes_D = c(436,14)
No_Dbar = c(5,495)
dt = data.frame(Yes_D,No_Dbar,row.names = c("Positive_T","Negative_Tbar") )
dt
prob_D = 0.113
prob_Dbar = 1 - prob_D
condi_TD = dt[1,1]/sum(Yes_D)
condi_TDbar = dt[1,2]/sum(No_Dbar)
condi_TbarD = dt[2,1]/sum(Yes_D)
condi_TbarDbar = dt[2,2]/sum(No_Dbar)

condi_DT = (condi_TD*prob_D)/(condi_TD*prob_D + condi_TDbar*prob_Dbar) #Bayes theorem
condi_DT
##Predictive value of positive test result is very high

condi_DbarTbar = (condi_TbarDbar*prob_Dbar)/(condi_TbarDbar*prob_Dbar + condi_TbarD*prob_D) #Bayes theorem
condi_DbarTbar
##Predictive value of negative test result is very high
```

Chapter 4

Probabilistic features of certain data distributions

R code Exa 4.2.1 Probability distribution from frequency table Page 93

```
##Example 4.2.1 Pg.93
##Probability distribution from frequency table

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)

prob_dist = freq/N
prob_dist
sum(prob_dist)

pdf = data.frame(x,freq,prob_dist)
pdf

barplot(prob_dist,names.arg = x, xlab="x(no. of
   assistance programs)",ylab="Probability")
```

R code Exa 4.2.2 Probability distribution from frequency table Page 95

```
##Example 4.2.2 Pg.95
##Probability distribution from frequency table P(X
    =3)

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)

prob_dist = freq/N
prob_dist
sum(prob_dist)

pdf = data.frame(x,freq,prob_dist)
pdf

prob_3 = pdf$prob_dist[x==3]  #gives the prob value
    at x=3 from the data frame
prob_3
```

R code Exa 4.2.3 Prob that family used either one or two programs Page 95

```
##Example 4.2.3 Pg.95
##Prob that family used either one or two programs

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)

prob_dist = freq/N
prob_dist
sum(prob_dist)

```

```
pdf = data.frame(x,freq,prob_dist)
pdf

prob_1 = pdf$prob_dist[x==1]
prob_2 = pdf$prob_dist[x==2]
prob_1U2 = prob_1 + prob_2 #additive rule of
    mutually exclusive events
prob_1U2
```

R code Exa 4.2.4 Cumulative probability distribution Page 97

```
##Example 4.2.4 Pg.97
##Cumulative probability distribution and p(X<=2)

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)
prob_dist = freq/N
cum_dist = cumsum(prob_dist)
cdf = data.frame(x,freq,prob_dist,cum_dist)
cdf

cdf_2 = cdf$cum_dist[x==2]
cdf_2
```

R code Exa 4.2.5 Cumulative probability distribution Page 97

```
##Example 4.2.5 Pg.97
##Cumulative probability distribution and p(X<4)=P(X
    <=3)

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
```

```
N = sum(freq)
prob_dist = freq/N
cum_dist = cumsum(prob_dist)
cdf = data.frame(x,freq,prob_dist,cum_dist)
cdf

cdf_3 = cdf$cum_dist[x==3]
cdf_3
```

R code Exa 4.2.6 Cumulative probability distribution Page 97

```
##Example 4.2.6 Pg.97
##Cumulative probability distribution and p(X>=5) =
    1 - P(X<=4)

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)
prob_dist = freq/N
cum_dist = cumsum(prob_dist)
cdf = data.frame(x,freq,prob_dist,cum_dist)
cdf

ans = 1 - cdf$cum_dist[x==4]
ans
```

R code Exa 4.2.7 Cumulative probability distribution Page 97

```
##Example 4.2.7 Pg.97
##Cumulative probability distribution and p(3<=X<=5)
    = P(X<=5) - P(X<=2)

x = 1:8
```

```
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)
prob_dist = freq/N
cum_dist = cumsum(prob_dist)
cdf = data.frame(x,freq,prob_dist,cum_dist)
cdf

cdf_5 = cdf$cum_dist[x==5]
cdf_2 = cdf$cum_dist[x==2]
ans = cdf_5 - cdf_2
ans
```

R code Exa 4.2.8 mean and variance of prob distribution Page 98

```
##Example 4.2.8 Pg.98
##mean and variance of prob distribution

x = 1:8
freq = c(62,47,39,39,58,37,4,11)
N = sum(freq)
prob_dist = freq/N
cum_dist = cumsum(prob_dist)
cdf = data.frame(x,freq,prob_dist,cum_dist)
cdf

mean = sum(x*prob_dist)
mean

variance = sum(x^2 * prob_dist) - mean^2
variance

sd = sqrt(variance)
sd
```

R code Exa 4.3.1 Binomial distribution Page 99

```
##Example 4.3.1 Pg.99
##Binomial distribution P(X=3)

binom_3 = dbinom(3,5,0.858) #gives binomial density
    for x=3,n=5,p=0.858
binom_3
```

R code Exa 4.3.2 Binomial probability distribution Page 103

```
##Example 4.3.2 Pg.103
##Binomial probability distribution

n = 10
x = 4
p = 14/100
f = dbinom(x,n,p) #prob of success for a binomial
    distribution
f  #porb that exactly 4 mothers will be admitted to
    smoking
```

R code Exa 4.3.3 Binomial probabilities Page 103

```
##Example 4.3.3 Pg.103
##Binomial probabilities P(X<=5) , P(X>=6) , P(6<=X
    <=9) , P(2<=X<=4)

prob_a = pbinom(5,25,0.1)   #gives binomial
    cumulative dist for x=5,n=25,p=0.1
```

```
prob_a

prob_b = 1 - pbinom(5,25,0.1)
prob_b

prob_c = pbinom(9,25,0.1) - pbinom(5,25,0.1)
prob_c

prob_d = pbinom(4,25,0.1) - pbinom(1,25,0.1)
prob_d
```

R code Exa 4.3.4 Binomial probabilities Page 105

```
##Example 4.3.4 Pg.105
##Binomial probabilities P(X=5) , P(X<=5) , P(X>=8)

prob_a = dbinom(5,12,0.45)     #gives binomial density for x=5,n=12,p=0.45
prob_a

prob_b = pbinom(5,12,0.55)     #gives binomial cumulative dist for x=5,n=12,p=0.55
prob_b

prob_c = 1 - pbinom(7,12,0.55)
prob_c
```

R code Exa 4.4.1 Poisson distribution Page 110

```
##Example 4.4.1 Pg.110
##Poisson distribution P(X=3)

```

```
pois_3 = dpois(3,12) #gives poisson density for x=3,
    lambda = 12
pois_3
```

R code Exa 4.4.2 Poisson distribution Page 110

```
##Example 4.4.2 Pg.110
##Poisson distribution P(X>=3) = 1 - P(X<=2)

ans = 1 - ppois(2,12) #gives poisson cumulative
    distribution for x=2, lambda = 12
ans
```

R code Exa 4.4.3 Poisson distribution Page 110

```
##Example 4.4.3 Pg.110
##Poisson distribution P(X<=1) at lambda=2

cum = ppois(1,2)
cum
```

R code Exa 4.4.4 Poisson distribution Page 111

```
##Example 4.4.4 Pg.111
##Poisson distribution P(X=3) at lambda=2

pois_3 = dpois(3,2)
pois_3
```

R code Exa 4.4.5 Poisson distribution Page 112

```
##Example 4.4.5 Pg.112
##Poisson distribution P(X>5) at lambda=2

pois_5 = 1 - ppois(5,2)
pois_5
```

R code Exa 4.6.1 Standard Normal Distribution Page 119

```
##Example 4.6.1 Pg.119
##Standard Normal Distribution
## Find P(z<2)

p = pnorm(2,0,1) #gives probability of normal dist
    with mean 0 and variance 1 less than 2
p

#Generates a random normal densities and plots for P
    (z<2)
x = seq(-4,4,length=10000)
y = dnorm(x,0,1)
plot(x,y,type="l",lwd=2,col="red")
x = seq(-4,2,length=10000)
y = dnorm(x,0,1)
polygon(c(-4,x,2),c(0,y,0),col="gray")
```

R code Exa 4.6.2 Standard Normal Distribution Page 120

```

##Example 4.6.2 Pg 120
##Standard normal distribution P(−2.55<z<2.55)

p = pnorm(2.55,0,1) - pnorm(-2.55,0,1) #gives probability of normal dist with mean 0 and variance 1
p

#Generates a random normal densities and plots for P(−2.55<z<2.55)
x = seq(-4,4,length=10000)
y = dnorm(x,0,1)
plot(x,y,type="l",lwd=2,col="red")
x = seq(-2.55,2.55,length=10000)
y = dnorm(x,0,1)
polygon(c(-2.55,x,2.55),c(0,y,0),col="gray")
```

R code Exa 4.6.3 Standard Normal Distribution Page 121

```
##Example 4.6.3 Pg 121
##Standard normal distribution P(−2.74<z<1.53)

p = pnorm(1.53,0,1) - pnorm(-2.74,0,1) #gives probability of normal dist with mean 0 and variance 1
p

#Generates a random normal densities and plots for P(−2.55<z<2.55)
x = seq(-4,4,length=10000)
y = dnorm(x,0,1)
plot(x,y,type="l",lwd=2,col="red")
x = seq(-2.74,1.53,length=10000)
y = dnorm(x,0,1)
```

```
polygon(c(-2.74,x,1.53),c(0,y,0),col="gray")
```

R code Exa 4.6.4 Standard Normal Distribution Page 121

```
##Example 4.6.4 Pg.121
##Standard Normal Distribution
## Find P(z>=2.71)

p = 1 - pnorm(2.71,0,1) #gives probability of normal
    dist with mean 0 and variance 1 less than 2
p

#Generates a random normal densities and plots for P
    (z<2)
x = seq(-4,4,length=10000)
y = dnorm(x,0,1)
plot(x,y,type="l",lwd=2,col="red")
x = seq(2.71,4,length=10000)
y = dnorm(x,0,1)
polygon(c(2.71,x,4),c(0,y,0),col="gray")
```

R code Exa 4.6.5 Standard normal distribution Page 122

```
##Example 4.6.5 Pg 122
##Standard normal distribution P(0.84<z<2.45)

p = pnorm(2.45,0,1) - pnorm(0.84,0,1) #gives
    probability of normal dist with mean 0 and
    variance 1
p

#Generates a random normal densities and plots for P
    (0.84<z<2.45)
```

```
x = seq(-4,4,length=10000)
y = dnorm(x,0,1)
plot(x,y,type="l",lwd=2,col="red")
x = seq(0.84,2.45,length=10000)
y = dnorm(x,0,1)
polygon(c(0.84,x,2.45),c(0,y,0),col="gray")
```

R code Exa 4.7.1 Normal distribution Page 123

```
##Example 4.7.1 Pg.123
##Normal distribution P(x<3)

mean = 5.4; sd=1.3
p = pnorm(3,mean,sd) #gives distribution of normal with mean 5.4 and sd 1.3
p

#Generates a random normal densities and plots for P(z<2)
x = seq(0,10,length=10000)
y = dnorm(x,mean,sd)
plot(x,y,type="l",lwd=2,col="red")
x = seq(0,3,length=10000)
y = dnorm(x,mean,sd)
polygon(c(0,x,3),c(0,y,0),col="gray")
```

R code Exa 4.7.2 Normal distribution Page 125

```
##Example 4.7.2 Pg 125
##Normal distribution P(292<X<649)

mean = 491; sd=119
p = pnorm(649,mean,sd) - pnorm(292,mean,sd)
```

```
p

#Generates a random normal densities and plots for P
    (0.84<z<2.45)
x = seq(0,1000,length=10000)
y = dnorm(x,mean,sd)
plot(x,y,type="l",lwd=2,col="red")
x = seq(292,649,length=10000)
y = dnorm(x,mean,sd)
polygon(c(292,x,649),c(0,y,0),col="gray")
```

R code Exa 4.7.3 Normal distribution Page 126

```
##Example 4.7.3 Pg 126
##Normal distribution P(X>8.5)

mean = 5.4 ; sd = 1.3
p = 1 - pnorm(8.5,mean,sd)
p
```

Chapter 5

Probabilisitc features of the distributions of certain sample statistics

R code Exa 5.3.2 Sampling distribution of mean Page 142

```
##Example 5.3.2 Pg.142
##Sampling distribution of mean

xbar = 190 ;  mu = 185.6 ; sd = 12.7 ; n=10

z = (xbar-mu)/(sd/sqrt(n))
z

#Generates a random normal densities and plots for P
    (0.84<z<2.45)
x = seq(140,225,length=10000)
y = dnorm(x,mu,sd)
plot(x,y,type="l",lwd=2,col="red")
x = seq(190,250,length=10000)
y = dnorm(x,mu,sd)
polygon(c(190,x,250),c(0,y,0),col="gray")
```

R code Exa 5.3.3 Sampling distribution of mean Page 143

```
##Example 5.3.3 Pg.143
##Sampling distribution of mean P(115<xbar<125) = P(z1<z<z2)

xbar1 = 115 ; xbar2 = 125;  mu = 120 ; sd = 15 ; n=50

z1 = (xbar1-mu)/(sd/sqrt(n))
z2 = (xbar2-mu)/(sd/sqrt(n))

z = pnorm(z2,0,1)-pnorm(z1,0,1)
z
```

R code Exa 5.4.1 Sampling distribution of two means Page 145

```
##Example 5.4.1 Pg.145
##Sampling distribution of two means

xbar1 = 92 ;xbar2 = 105; mu1=0 ; mu2 = 0 ; sd1 = 20 ; sd2 = 20 ; n1=15 ; n2=15

z = ((xbar1-xbar2)-(mu1-mu2))/(sqrt((sd1^2/n1)+(sd2^2/n2)))
z

prob_z = pnorm(z,0,1)
prob_z
```

R code Exa 5.4.2 Sampling distribution of two means Page 148

```
##Example 5.4.2 Pg.148
##Sampling distribution of two means

xbar = 20 ; mu1=45 ; mu2 = 30 ; sd1 = 15 ; sd2 = 20
    ; n1=35 ; n2=40

z = (xbar-(mu1-mu2))/(sqrt((sd1^2/n1)+(sd2^2/n2)))
z = round(z,2)
z

prob_z = 1 - pnorm(z,0,1)
prob_z
```

R code Exa 5.5.2 Sampling distribution of Proportion Page 152

```
##Example 5.5.2 Pg.152
##Sampling distribution of Proportion

P = 0.45 ; p = 0.51 ; q = 1-p ; n=200

z = (P - p)/sqrt(p*q/n)
z = round(z,2)
z

prob_z = pnorm(z,0,1)
prob_z
```

R code Exa 5.6.1 Sampling distribution of Two Proportions Page 155

```
##Example 5.6.1 Pg.155
##Sampling distribution of Two Proportions
```

```
P = 0.10 ; p1 = 0.28 ; p2 = 0.21 ; n=100 ; q1 = 1-p1
    ; q2 = 1-p2

z = (P - (p1-p2))/sqrt((p1*q1/n)+(p2*q2/n))
z = round(z,2)
z

prob_z = 1 - pnorm(z,0,1)
prob_z
```

R code Exa 5.6.2 Sampling distribution of Two Proportions Page 155

```
##Example 5.6.2 Pg.155
##Sampling distribution of Two Proportions

P = 0.05 ; p1 = 0.34 ; p2 = 0.26 ; n1=250 ;n2=200;
   q1 = 1-p1 ; q2 = 1-p2

z = (P - (p1-p2))/sqrt((p1*q1/n1)+(p2*q2/n2))
z = round(z,2)
z

prob_z = pnorm(z,0,1)
prob_z
```

Chapter 6

Using sample data to make estimates about population parameters

R code Exa 6.2.1 Confidence interval for population mean mu Page 166

```
##Exaxmple 6.2.1 Pg.166
##Confidence iterval for population mean mu

xbar = 22 ; variance = 45; n = 10 ; alpha = 0.05
p = qnorm(1-alpha/2,0,1) #gives alpha level p value
p = round(p,1)
p

conf_l = xbar - p*sqrt(variance/n)
conf_u = xbar + p*sqrt(variance/n)

conf = c(conf_l,conf_u)
conf
```

R code Exa 6.2.2 Confidence interval for population mean mu Page 168

```
##Exaxmple 6.2.2 Pg.168
##Confidence iterval for population mean mu

xbar = 84.3 ; variance = 144; n = 15 ; alpha = 0.01
    ; p =qnorm(1-alpha/2,0,1) #gives alpha level p
    value

conf_l = xbar - p*sqrt(variance/n)
conf_u = xbar + p*sqrt(variance/n)

conf = c(conf_l,conf_u)
conf
```

R code Exa 6.2.3 Confidence interval for population mean mu Page 168

```
##Exaxmple 6.2.3 Pg.168
##Confidence iterval for population mean mu

xbar = 17.2 ; variance = 8^2; n = 35 ; alpha = 0.1 ;
     p =qnorm(1-alpha/2,0,1) #gives alpha level p
    value

conf_l = xbar - p*sqrt(variance/n)
conf_u = xbar + p*sqrt(variance/n)

conf = c(conf_l,conf_u)
conf
```

R code Exa 6.2.4 Confidence interval for population mean mu Page 169

```
##Exaxmple 6.2.4 Pg.169
```

```
##Confidence iterval for population mean mu

x = c(.360,1.827,.372,.610,.521,1.189,.537,.898,.319,.603,.614,.374,.4
xbar = mean(x) ; variance = 0.36; n = length(x) ; alpha = 0.05; p =qnorm(1-alpha/2,0,1) #gives alpha level p value

conf_l = xbar - p*sqrt(variance/n)
conf_u = xbar + p*sqrt(variance/n)

conf = c(conf_l,conf_u)
conf                          #95% confidence interval
xbar                          #Mean
sqrt(variance)                #Std dev
sqrt(variance/n)              #Standard error of mean
```

R code Exa 6.3.1 Confidence interval for mean T test Page 173

```
##Example 6.3.1 Pg.173
##Confidence interval for mean - T test

xbar = 250.8 ; s = 130.9 ; n = 19 ; alpha = 0.05; p =qt(1-alpha/2,n-1) #gives alpha level p value

conf_l = xbar - p*sqrt(s^2/n)
conf_u = xbar + p*sqrt(s^2/n)

conf = c(conf_l,conf_u)
conf                          #95% confidence interval
```

R code Exa 6.4.1 Confidence interval for two means Z test Page 177

```
##Example 6.4.1 Pg.177
##Confidence interval for two means - Z test

xbar1 = 4.5 ; xbar2=3.4; n1=12; n2=15; var1 = 1 ;
   var2=1.5
alpha = 0.05; p =qnorm(1-alpha/2,0,1) #gives alpha
   level p value

conf_l = (xbar1-xbar2) - p*sqrt((var1/n1)+(var2/n2))
conf_u = (xbar1-xbar2) + p*sqrt((var1/n1)+(var2/n2))

conf = c(conf_l,conf_u)
conf                          #95% confidence interval
```

R code Exa 6.4.2 Confidence interval for two means Z test Page 178

```
##Example 6.4.2 Pg.178
##Confidence interval for two means - Z test

xbar1 = 4.3 ; xbar2=13; n1=328; n2=64; s1 = 5.22 ;
   s2=8.97
alpha = 0.01; p =qnorm(1-alpha/2,0,1) #gives alpha
   level p value

conf_l = (xbar1-xbar2) - p*sqrt((s1^2/n1)+(s2^2/n2))
conf_u = (xbar1-xbar2) + p*sqrt((s1^2/n1)+(s2^2/n2))

conf = c(conf_l,conf_u)
conf                          #95% confidence interval
```

R code Exa 6.4.3 Confidence interval for two means T test Page 180

```
##Example 6.4.3 Pg.180
```

```
##Confidence interval for two means - T test

xbar1 = 4.7 ; xbar2= 8.8; n1=18; n2=10; var1 = 9.3^2
    ; var2=11.5^2
alpha = 0.05; p =qt(1-alpha/2,n1+n2-2) #gives alpha
    level p value
s_pooled = ((n1-1)*var1+(n2-1)*var2)/(n1+n2-2)

conf_l = (xbar1-xbar2) - p*sqrt((s_pooled/n1)+(s_
    pooled/n2))
conf_u = (xbar1-xbar2) + p*sqrt((s_pooled/n1)+(s_
    pooled/n2))

conf = c(conf_l,conf_u)
conf                            #95% confidence interval
```

R code Exa 6.4.4 Confidence interval for two means T test Page 181

```
##Example 6.4.4 Pg.181
##Confidence interval for two means - T test

xbar1 = 4.7 ; xbar2= 8.8; n1=18; n2=10; var1 = 9.3^2
    ; var2=11.5^2 ; alpha = 0.05
t1 = qt(1-alpha/2,n1-1); t2 = qt(1-alpha/2,n2-1)
t = ((t1*var1/n1)+(t2*var2/n2))/((var1/n1) + (var2/
    n2))


s_pooled = ((n1-1)*var1+(n2-1)*var2)/(n1+n2-2)

conf_l = (xbar1-xbar2) - t*sqrt((var1/n1)+(var2/n2))
conf_u = (xbar1-xbar2) + t*sqrt((var1/n1)+(var2/n2))

conf = c(conf_l,conf_u)
conf                            #95% confidence interval
```

R code Exa 6.5.1 Confidence interval for population proportion Page 185

```
##Example 6.5.1 Pg.185
##Confidence interval for population proportion

P = 0.18 ; Q=1-P ;  variance = 8^2; n =1220; alpha =
    0.05 ; p =qnorm(1-alpha/2,0,1) #gives alpha
   level p value

conf_l = P - p*sqrt(P*Q/n)
conf_u = P + p*sqrt(P*Q/n)

conf = c(conf_l,conf_u)
conf
```

R code Exa 6.6.1 Confidence interval for two population proportions Page 187

```
##Exaxmple 6.6.1 Pg.187
##Confidence iterval for two population proportions

P1 = 31/68; P2 = 53/255; Q1=1-P1 ; Q2 = 1-P2;  n1 =
   68 ; n2 = 255
alpha = 0.01 ; p =qnorm(1-alpha/2,0,1) #gives alpha
   level p value

conf_l = (P1-P2) - p*sqrt((P1*Q1/n1)+(P2*Q2/n2))
conf_u = (P1-P2) + p*sqrt((P1*Q1/n1)+(P2*Q2/n2))

conf = c(conf_l,conf_u)
conf
```

R code Exa 6.7.1 Determine the sample size n Page 190

```
##Example 6.7.1 Pg.190
##Determine the sample size n

z = 1.96 ; sd = 20 ; d = 5

n = z^2 * sd^2 / d^2
n
```

R code Exa 6.8.1 Determine the sample size n Page 192

```
##Example 6.8.1 Pg.192
##Determine the sample size n

z = qnorm(1-0.05/2,0,1) ; p=0.35; q=1-p; d=0.05
n = z^2 * p*q / d^2
n = round(n,0)
n
```

R code Exa 6.9.1 Chi square test Page 196

```
##Example 6.9.1 Pg.196
##Chi square test

x <- c(9.7,12.3,11.2,5.1,24.8,14.8,17.7)
conf.level= 0.95
df = length(x) - 1
chilower = round(qchisq((1 - conf.level)/2, df),3)
```

```
chiupper = round(qchisq((1 - conf.level)/2, df,
    lower.tail = FALSE),3)
v = var(x)
c(df * v/chiupper, df * v/chilower)  #95% conf
    interval for variance
c(sqrt(df * v/chiupper), sqrt(df * v/chilower))  #
    95% conf interval for sd
```

R code Exa 6.10.1 F test Page 200

```
##Example 6.10.1 Pg.200
##F test

n1=16 ; n2=4 ; s1 =8.1 ; s2 = 5.9; df1 = n1-1 ; df2=
    n2-1; alpha = 0.05

Flower = qf(alpha/2, df1,df2)
Fupper = qf(1 - (alpha/2), df1,df2)

conf = c(s1^2/(s2^2*Fupper), s1^2/(s2^2*Flower))  #
    95% conf interval for 2 variances
conf
```

Chapter 7

Using sample statistics to test hypothesis about population parameters

R code Exa 7.2.1 Test for mean when population variances are known Page 222

```
##Example 7.2.1 Pg.222
##Test for mean when population variances are known
   (two sided)

mu = 30 ; xbar = 27 ; var = 20 ; n=10
z = (xbar-mu)/sqrt(var/n)
z
z_critical = qnorm(0.05/2,0,1)
z_critical
pvalue = 2*pnorm(z,0,1)
pvalue
##Z > z_critical or pvalue<0.05 , hence significant
```

R code Exa 7.2.2 Test for mean when population variances are known Page 226

```r
##Example 7.2.2 Pg.226
##Test for mean when population variances are known (one sided)

mu = 30 ; xbar = 27 ; var = 20 ; n=10
z = (xbar-mu)/sqrt(var/n)
z
z_critical = qnorm(0.05,0,1)
z_critical
pvalue = pnorm(z,0,1)
pvalue
##Z > z_critical or pvalue<0.05 , hence significant
```

R code Exa 7.2.3 Test for mean when population variances are unknown Page 228

```r
##Example 7.2.3 Pg.228
##Test for mean when population variances are unknown (two sided)

days<- c(14,9,18,26,12,0,10,4,8,21,28,24,24,2,3,14,9)
mu = 15 ; xbar = mean(days) ; var = var(days) ; n=17 ; df = n-1
t = (xbar-mu)/sqrt(var/n)
t
t_critical = qt(0.05/2,df)
t_critical
pvalue = 2*pt(t,df)
pvalue

##T > t_critical or pvalue<0.05 , hence significant
```

R code Exa 7.2.4 Test for mean when population variances are known Page 231

```
##Example 7.2.4 Pg.231
##Test for mean when population variances are known
    (one sided)

mu = 140 ; xbar = 146 ; s = 27 ; n=157
z = (xbar-mu)/(s/sqrt(n))
z
z_critical = qnorm(1-0.05,0,1)
z_critical
pvalue = 1-pnorm(z,0,1)
pvalue
##Z > z_critical or pvalue<0.05 , hence significant
```

R code Exa 7.2.5 Test for mean when population variances are unknown Page 232

```
##Example 7.2.5 Pg.232
##Test for mean when population variances are
    unknown (two sided)

circ<- c
    (33.38,34.34,33.46,32.15,33.95,34.13,33.99,33.85,34.45,34.10,34.2

mu = 34.5 ; xbar = mean(circ) ; var = var(circ) ; n=
    length(circ) ; df = n-1
t = (xbar-mu)/sqrt(var/n)
t
t_critical = qt(0.05/2,df)
```

```
t_critical
pvalue = 2*pt(t,df)
pvalue

##T > t_critical or pvalue<0.05 , hence significant
```

R code Exa 7.3.1 Test for two means when population variances are known Page 237

```
##Example 7.3.1 Pg.237
##Test for two means when population variances are
    known (two sided)

mu1mu2 = 0 ; xbar1 =4.5 ;xbar2=3.4; var1 = 1; var2
    =1.5 ; n1=12; n2=15
z = ((xbar1-xbar2)-(mu1mu2))/sqrt((var1/n1)+(var2/n2
    ))
z
z_critical = qnorm(0.05/2,0,1)
z_critical
pvalue = 1-pnorm(z,0,1)
pvalue
##Z > z_critical or pvalue<0.05 , hence significant
```

R code Exa 7.3.2 Test for means when population variances are unknown Page 239

```
##Example 7.3.2 Pg.239
##Test for means when population variances are
    unknown (one sided)

control <-c(131,115,124,131,122,117,88,114,150,169)
sci <- c(60,150,130,180,163,130,121,119,130,148)
```

```
t.test(control,sci)
#pvalue >0.05, hence not significant
```

R code Exa 7.3.3 Test for means when population variances are unknown Page 240

```
##Example 7.3.3 Pg.240
##Test for means when population variances are
   unknown (two sided)

n1 = 15 ; n2 = 30 ; xbar1 = 19.16; xbar2 = 9.53 ; s1
    = 5.29; s2 = 2.69
t = (xbar1-xbar2)/(sqrt((s1^2/n1)+(s2^2/n2)))
t
alpha = 0.05; df = n1+n2-2
t_critical = qt(0.05/2,df)
t_critical


#T > Tcritical, hence significant
#Answer might slightly differ due to approximation
```

R code Exa 7.3.4 Test for two means when population variances are known Page 242

```
##Example 7.3.4 Pg.242
##Test for two means when population variances are
   known (two sided)

mu1mu2 = 0 ; xbar1 =59.01 ;xbar2=46.61; var1
   =44.89^2; var2 =34.85^2 ; n1=53; n2=54
```

```
z = ((xbar1-xbar2)-(mu1mu2))/sqrt((var1/n1)+(var2/n2))
z
z_critical = qnorm(1-0.01,0,1)
z_critical
pvalue = 1-pnorm(z,0,1)
pvalue
##Z < z_critical or pvalue >0.05 , hence not significant
```

R code Exa 7.4.1 Paired t test Page 251

```
##Example 7.4.1 Pg.251
##Paired t test (one sided)

preop <- c(22,63.3,96,9.2,3.1,50,33,69,64,18.8,0,34)
postop <-c(63.5,91.5,59,37.8,10.1,19.6,41,87.8,86,55,88,40)

t.test(postop,preop,paired = T,alternative = "greater")

#pvalue <0.05 , hence significant
```

R code Exa 7.5.1 Test for proportions Page 258

```
##Example 7.5.1 Pg.258
##Test for proportions (one sided)

P = 0.063 ; p =24/301 ; Q=1-P ; alpha = 0.05 ; n=301

z = (P - p)/(sqrt(P*Q/n))
z = round(z,2)
```

```
z
z_critical = qnorm(alpha,0,1)
z_critical
pvalue = pnorm(z,0,1)
pvalue

#Since p value >alpha, hence not significant
```

R code Exa 7.6.1 Test for two proportions Page 261

```
##Example 7.6.1 Pg.261
##Test for two proportions (one sided)

p1 = 24/44 ; p2 =11/29 ; q1=1-p1 ; q2=1-p2 ; alpha =
    0.05 ; n1=44; n2=29
z = (p1-p2)/(sqrt((p1*q1/n1)+(p2*q2/n2)))
z = round(z,2)
z
z_critical = qnorm(1-alpha,0,1)
z_critical
pvalue = 1 - pnorm(z,0,1)
pvalue

#Since p value >alpha, hence not significant
```

R code Exa 7.7.1 Test for single population variance Page 264

```
##Example 7.7.1 Pg.264
##Test for single population variance

var =600 ; n=16 ; df=n-1; s2 = 670.81 ; alpha = 0.05
chisq = s2*(n-1)/var
chisq
```

```
chi_critical1 = qchisq(alpha/2,df)
chi_critical2 = qchisq(alpha/2,df,lower.tail = F)
chi_critical = c(chi_critical1,chi_critical2)
chi_critical
pvalue = pchisq(chisq,df)
pvalue

#Since pvalue>alpha, hence not significant
```

R code Exa 7.8.1 Test for ratio of two population variances Page 268

```
##Example 7.8.1 Pg.268
##Test for ratio of two population variances

s1 =30.62 ;s2 = 11.37; n1=6 ;n2=6; df1=n1-1; df2 = n2-1 ; alpha = 0.05
f = s1^2/s2^2
f
f_critical = qf(alpha,df1,df2,lower.tail = F)
f_critical
pvalue = pf(f,df1,df2,lower.tail = F)
pvalue

#Since pvalue<alpha, hence significant
```

R code Exa 7.8.2 Test for ratio of two population variances Page 270

```
##Example 7.8.2 Pg.270
##Test for ratio of two population variances

control <-c(131,115,124,131,122,117,88,114,150,169)
sci <- c(60,150,130,180,163,130,121,119,130,148)

```

```
var.test(control,sci)

#pvalue>0.05, hence not significant
#Answer matches with minitab output
```

Chapter 8

Statistical inference and the analysis of data variability

R code Exa 8.2.1 One way ANOVA Page 318

```
##Example 8.2.1 Pg.318
##One way ANOVA

ven = c
   (26.72,28.58,29.71,26.95,10.97,21.97,14.35,32.21,19.19,30.92,10.42
         16.47,25.19,37.45,45.08,25.22,22.11,33.01,31.20,26.50,32.77,
squ = c
   (37.42,56.46,51.91,62.73,4.55,39.17,38.44,40.92,58.93,61.88,49.54
rrb = c
   (11.23,29.63,20.42,10.12,39.91,32.66,38.38,36.21,16.39,27.44,17.29
nrb = c
   (44.33,76.86,4.45,55.01,58.21,74.72,11.84,139.09,69.01,94.61,48.35

selenium = c(ven,squ,rrb,nrb)
```

```r
type = c(rep(1,length(ven)),rep(2,length(squ)),rep(3,length(rrb)),rep(4,length(nrb)))
type = factor(type,labels = c("ven","squ","rrb","nrb"))
dt = data.frame(type,selenium)
View(dt)

anova <- aov(selenium~type)  #anova model for selenium content and meat type
anova
summary(anova)

#pvalue <0.05, hence significant
#Answers might slightly differ due to approximation
```

R code Exa 8.2.2 One way ANOVA and Tukeys HSD Page 325

```r
##Example 8.2.2 Pg.325
##One way ANOVA and Tukeys HSD

ven = c(26.72,28.58,29.71,26.95,10.97,21.97,14.35,32.21,19.19,30.92,10.4
        16.47,25.19,37.45,45.08,25.22,22.11,33.01,31.20,26.50,32.77,
squ = c(37.42,56.46,51.91,62.73,4.55,39.17,38.44,40.92,58.93,61.88,49.54
rrb = c(11.23,29.63,20.42,10.12,39.91,32.66,38.38,36.21,16.39,27.44,17.2
nrb = c(44.33,76.86,4.45,55.01,58.21,74.72,11.84,139.09,69.01,94.61,48.3

```

```
selenium = c(ven,squ,rrb,nrb)
type = c(rep(1,length(ven)),rep(2,length(squ)),rep(3,length(rrb)),rep(4,length(nrb)))
type = factor(type,labels = c("ven","squ","rrb","nrb"))
dt = data.frame(type,selenium)
View(dt)

anova <- aov(selenium~type)  #anova model for selenium content and meat type
anova
summary(anova)

#pvalue<0.05, hence significant

posthoc <- TukeyHSD(anova, "type", conf.level=0.95)
posthoc

#Reject the null if pvalue<alpha
#Answers might slightly differ due to approximation
```

R code Exa 8.3.1 Ranadomized Block Design Two way ANOVA Page 339

```
##Example 8.3.1 Pg.339
##Ranadomized Block Design - Two way ANOVA

days <- c(7,8,9,10,11,9,9,9,9,12,10,10,12,12,14)
age <- rep(c(1,2,3,4,5),3)
age <- factor(age,labels = c("under 20","20-29","30-39","40-49","50 and over"))
teach <- c(rep(1,5),rep(2,5),rep(3,5))
teach <- factor(teach,labels=c("A","B","C"))
dt = data.frame(days,teach,age)
dt

```

```
anova <- aov(days~teach+age)
anova
summary(anova)

#Answers may slightly vary due to approximation
```

Chapter 9

Statistical inference and the relationship between two variables

R code Exa 9.3.1 Linear regression model Page 417

```
##Example 9.3.1 Pg.417
##Linear regression model

x <- c
   (74.75,72.60,81.80,83.95,74.65,71.85,80.90,83.40,63.50,73.2,71.9,7

y <- c
   (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32

plot(x,y,main="scatter plot of x and y",xlab="waist
   circumference",ylab="deep abdominal area")

reg = lm(y~x)  #Create a linear model
summary(reg)
resid(reg)   #List of residuals
plot(density(resid(reg))) #A density plot
```

```
qqnorm(resid(reg)) # A quantile normal plot - good for checking normality
qqline(resid(reg))

#Answers may differ due to approximations
```

R code Exa 9.4.1 Linear regression model Coefficient of determination Page 432

```
##Example 9.4.1 Pg.432
##Linear regression model - Coefficient of determination

x <-  c(74.75,72.60,81.80,83.95,74.65,71.85,80.90,83.40,63.50,73.2,71.9,7
y <- c(25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32

reg = lm(y~x)  #Create a linear model
summary(reg)

#Multiple R squared is 0.4556
#Answers may differ due to approximations
```

R code Exa 9.4.2 Linear regression model Test for slope Page 436

```
##Example 9.4.2 Pg.436
##Linear regression model - Test for slope

```

```
x <-  c
    (74.75,72.60,81.80,83.95,74.65,71.85,80.90,83.40,63.50,73.2,71.9,7
y <- c
    (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32

reg = lm(y~x)  #Create a linear model
summary(reg)

#pvalue of x is less than 0.05, hence significant
#Answers may differ due to approximations
```

R code Exa 9.4.3 Linear regression model Residual plot Page 440

```
##Example 9.4.3 Pg.440
##Linear regression model − Residual plot

x <-  c
    (74.75,72.60,81.80,83.95,74.65,71.85,80.90,83.40,63.50,73.2,71.9,7
y <- c
    (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32

reg = lm(y~x)  #Create a linear model
summary(reg)
res = resid(reg)
plot(x,res,ylab="residuals",xlab="abdominal area")
abline(0,0)

#plot may not be exact replicate due to a different
    scale
```

R code Exa 9.7.1 Correlation and linear regression model Page 447

```
##Example 9.7.1 Pg.447
##Correlation and linear regression model

height<-c
   (149,149,155,155,156,156,157,157,158,158,160,160,161,161,161,161,
   rep(179,9),180,180,181,181,181,181,181,rep(182,7)
   ,rep(184,6),185,185,
          187,187,187,187,188,188,189,189,190,190,190,190,191,191,19

cv <- c
   (14.4,13.4,13.5,13.5,13,13.6,14.3,14.9,14,14,15.4,14.7,15.5,15.7,


corr = cor(height,cv)
corr
reg <- lm(cv~height)
reg
summary(reg)
plot(height,cv)
abline(reg)
```

R code Exa 9.7.2 Correlation test Page 452

```
##Example 9.7.2 Pg.452
##Correlation test

height<-c
   (149,149,155,155,156,156,157,157,158,158,160,160,161,161,161,161,
   rep(179,9),180,180,181,181,181,181,181,rep(182,7)
   ,rep(184,6),185,185,
```

```
            187,187,187,187,188,188,189,189,190,190,190,190,191,191,19
cv <- c
   (14.4,13.4,13.5,13.5,13,13.6,14.3,14.9,14,14,15.4,14.7,15.5,15.7,

corr = cor(height,cv)
corr
reg <- lm(cv~height)
cor.test(height,cv)   #performs correlation test

#pvaue<0.05, hence significant
```

Chapter 10

Statistical inference and the relationships among three or more variables

R code Exa 10.3.1 Multiple regression equation Page 493

```

##Example 10.3.1 Pg.493
##Multiple regression equation

age<- c
   (72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
edlevel <- c
   (20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
cda <- c
   (4.57,-3.04,1.39,-3.55,-2.56,-4.66,-2.70,0.30,-4.46,-6.29,-4.43,0
dt = data.frame(age,edlevel,cda)

pairs(dt) #multiple scatter plots

```

```
reg <- lm(cda~age+edlevel)  #multiple regression model
reg
summary(reg)

#Answers might slightly differ due to approximation
```

R code Exa 10.4.1 Coefficient of multiple determination Page 502

```
##Example 10.4.1 Pg.502
##Coefficient of mutliple determination

age<- c(72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
edlevel <- c(20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
cda <- c(4.57,-3.04,1.39,-3.55,-2.56,-4.66,-2.70,0.30,-4.46,-6.29,-4.43,0
dt = data.frame(age,edlevel,cda)

reg <- lm(cda~age+edlevel)  #multiple regression model
reg
summary(reg)

#Multiple R squared value is 0.01807
#Answers might slightly differ due to approximation
```

R code Exa 10.4.2 Test for parameters Page 504

```
##Example 10.4.2 Pg.504
##Test for parameters

age<- c(72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
edlevel <- c(20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
cda <- c(4.57,-3.04,1.39,-3.55,-2.56,-4.66,-2.70,0.30,-4.46,-6.29,-4.43,0
dt = data.frame(age,edlevel,cda)

reg <- lm(cda~age+edlevel) #multiple regression model
summary(reg)
summary(aov(reg))

#pvalue > 0.05 , hence there is a significant relationship between three variables
#Answers might slightly differ due to approximation
```

R code Exa 10.4.3 Test for parameter of variable Age Page 505

```
##Example 10.4.3 Pg.505
##Test for parameter of variable Age

age<- c(72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
edlevel <- c(20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
cda <- c
```

```
(4.57,-3.04,1.39,-3.55,-2.56,-4.66,-2.70,0.30,-4.46,-6.29,-4.43,0
dt = data.frame(age,edlevel,cda)

reg <- lm(cda~age+edlevel)  #multiple regression model
reg
summary(reg)

#pvalue >0.05, hence significant relationship between cda and age
#Answers might slightly differ due to approximation
```

R code Exa 10.5.1 confidence interval for CDA Page 508

```
##Example 10.5.1 Pg.508
##95% confidencec interval for CDA

age<- c(72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
edlevel <- c(20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
cda <- c(4.57,-3.04,1.39,-3.55,-2.56,-4.66,-2.70,0.30,-4.46,-6.29,-4.43,0
dt = data.frame(age,edlevel,cda)

reg <- lm(cda~age+edlevel)  #multiple regression model
reg
summary(reg)

new.dat <- data.frame(age=68,edlevel=12)    #new
```

```
observation
predict(reg, newdata = new.dat, interval = 'confidence') #confidence interval
predict(reg, newdata = new.dat, interval = 'prediction') #prediction interval

#Answers might slightly differ due to approximation
```

R code Exa 10.6.1 Multiple correlation coefficient Page 511

```
##Example 10.6.1 Pg.511
##Multiple correlation coefficient

w<- c(193.6,137.5,145.4,117,105.4,99.9,74,74.4,112.8,125.4,126.5,115.9
p<- c(6.24,8.03,11.62,7.68,10.72,9.28,6.23,8.67,6.91,7.51,10.01,8.70,5
s<- c(30.1,22.2,25.7,28.9,27.3,33.4,26.4,17.2,15.9,12.2,30,24,22.6,18.

reg = lm(w~p+s)
reg
summary(reg)

#Multiple R sqaured = 0.2942
#Answers might slightly differ due to approximation
```

R code Exa 10.6.2 Partial correlation coefficient Page 515

```
##Example 10.6.2 Pg.515
```

```r
##Partial correlation coefficient

w<- c(193.6,137.5,145.4,117,105.4,99.9,74,74.4,112.8,125.4,126.5,115.9
p<- c(6.24,8.03,11.62,7.68,10.72,9.28,6.23,8.67,6.91,7.51,10.01,8.70,5
s<- c(30.1,22.2,25.7,28.9,27.3,33.4,26.4,17.2,15.9,12.2,30,24,22.6,18.

reg = lm(w~p+s)
reg
summary(reg)

res1 = residuals(lm(w~p))
res2 = residuals(lm(s~p))
res3 = residuals(lm(w~s))
res4 = residuals(lm(p~s))
res5 = residuals(lm(p~w))
res6 = residuals(lm(s~w))

# use Spearman correlation coefficient to calculate the all possible partial correlations
p1 = cor(res1,res2,method = "spearman")
p2 = cor(res1,res3,method = "spearman")
p3 = cor(res1,res4,method = "spearman")
p4 = cor(res1,res5,method = "spearman")
p5 = cor(res1,res6,method = "spearman")
p6 = cor(res2,res3,method = "spearman")
p7 = cor(res2,res4,method = "spearman")
p8 = cor(res2,res5,method = "spearman")
p9 = cor(res2,res6,method = "spearman")
p10 = cor(res3,res4,method = "spearman")
p11 = cor(res3,res5,method = "spearman")
p12 = cor(res3,res6,method = "spearman")
p13 = cor(res4,res5,method = "spearman")
```

```
p14 = cor(res4,res6,method = "spearman")
p15 = cor(res5,res6,method = "spearman")
p <- c(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12,p13,
   p14,p15)
p


#Answers might slightly differ due to approximation
```

Chapter 11

Additional techniques for the analysis of relationships among variables

R code Exa 11.1.1 Box and whiskers plot

```
##Example 11.1.1 Pg.540
##Box and whiskers plot

THC <- c
   (.30,2.75,2.27,2.37,1.12,.60,.61,.89,.33,.85,2.18,3.59,.28,1.90,1
log_THC <- log10(THC)
log_THC
data = data.frame(THC,log_THC)

#install.packages("car")
library(car)
Boxplot(data,ylab="concentration")
```

R code Exa 11.1.2 Correlation for 3 variables Page 542

```
##Example 11.1.2 Pg.542
##Correlation for 3 variables

sbp <- c
   (126,129,126,123,124,125,127,125,123,119,127,126,122,126,125)
weight <- c
   (125,130,132,200,321,100,138,138,149,180,184,251,197,107,125)
bmi <- c
   (24.41,23.77,20.07,27.12,39.07,20.90,22.96,24.44,23.33,25.82,26.4
dt<-data.frame(sbp,weight,bmi)

#install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)

chart.Correlation(dt)
##Shows correlation coefficient and significant
   values

cor.test(sbp,weight)
#result shows cor = -0.289, p value = 0.296
cor.test(sbp,bmi)
#result shows cor = -0.213, p value = 0.447
cor.test(bmi,weight)
#result shows cor = 0.962, p value = 0.000
```

R code Exa 11.2.1 Regression model for categorical data Page 545

```
##Example 11.2.1 Pg.545
##Regression model for categorical data

```

```
grams <- c(3147,2977,3119,3487,4111,3572,3487,
           3147,3345,2665,1559,3799,
           2750,3487,3317,3544,3459,2807,3856,
           3260,2183,3204,3005,3090,3430,3119,
           3912,3572,3884,3090,2977,3799,4054,
           3430,3459,3827,3147,3289,3629,3657,
           3175,3232,3175,3657,3600,3572,709,624,
           2778,3572,3232,3317,2863,3175,3317,3714,
           2240,3345,3119,2920,3430,3232,3430,4139,
           3714,1446,3147,2580,3374,3941,2070,3345,
           3600,3232,3657,3487,2948,2722,3771,3799,
           1871,3260,3969,3771,3600,2693,3062,2693,3033,3856,
           4111,3799,3147,2920,4054,2296,3402,1871,
           4167,3402)
weeks <- c
   (40,41,38,38,39,41,40,41,38,34,34,38,38,40,38,
           43,45,37,40,40,42,38,36,40,39,40,39,40,41,38,
           42,37,40,38,41,39,44,38,36,36,41,43,36,40,39,
           40,25,25,36,35,38,40,37,37,40,34,36,39,39,37,
           41,35,38,39,39,28,39,31,37,40,37,40,40,41,38,
           39,38,40,40,45,33,39,38,40,40,35,45,36,41,42,
           40,39,38,36,40,36,38,33,41,37)
smoke <- c
   (0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,0,0,
           1,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,0,
           0,0,0,rep(0,15),1,1,0,0,0,0,0,0,0,1,rep
              (0,10),
           1,rep(0,11),1,0,1)
smoke = factor(smoke, labels = c("nonsmoker","smoker
   "),levels=c(0,1))
```

```
plot(weeks, grams, pch=21,
     bg=c("red","green3")[unclass(smoke)])
##red for non smokers and green for smokers

reg = lm(grams~weeks+smoke)
summary(reg)
##Gives the estimates and corresponding p values
summary(aov(reg))
##Gives the Anova results (sum of squares and F statistic)


plot(weeks, grams, pch=21,
     bg=c("red","green3")[unclass(smoke)])
abline(reg)
```

R code Exa 11.2.2 Test for model parameter for categorical data

```
##Example 11.2.2 Pg.549
##Test for model parameter for categorical data

grams <- c(3147,2977,3119,3487,4111,3572,3487,
           3147,3345,2665,1559,3799,
           2750,3487,3317,3544,3459,2807,3856,
           3260,2183,3204,3005,3090,3430,3119,
           3912,3572,3884,3090,2977,3799,4054,
           3430,3459,3827,3147,3289,3629,3657,
           3175,3232,3175,3657,3600,3572,709,624,
           2778,3572,3232,3317,2863,3175,3317,3714,
           2240,3345,3119,2920,3430,3232,3430,4139,
           3714,1446,3147,2580,3374,3941,2070,3345,
           3600,3232,3657,3487,2948,2722,3771,3799,
           1871,3260,3969,3771,3600,2693,3062,2693,3033,3856,
           4111,3799,3147,2920,4054,2296,3402,1871,
```

```
            4167,3402)
weeks <- c(40,41,38,38,39,41,40,41,38,34,34,38,38,40,38,
            43,45,37,40,40,42,38,36,40,39,40,39,40,41,38,
            42,37,40,38,41,39,44,38,36,36,41,43,36,40,39,
            40,25,25,36,35,38,40,37,37,40,34,36,39,39,37,
            41,35,38,39,39,28,39,31,37,40,37,40,40,41,38,
            39,38,40,40,45,33,39,38,40,40,35,45,36,41,42,
            40,39,38,36,40,36,38,33,41,37)
smoke <- c(0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,0,0,
            1,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,0,
            0,0,0,rep(0,15),1,1,0,0,0,0,0,0,0,1,rep(0,10),
            1,rep(0,11),1,0,1)
smoke = factor(smoke, labels = c("nonsmoker","smoker"),levels=c(0,1))

reg = lm(grams~weeks+smoke)
summary(reg)

##We get the t test statistic for smokers as −2.17
##p value is 0.033 < 0.05, hence significant
##smoking mothers associateed with reduced birth weights of babies
```

R code Exa 11.3.1 Stepwise regression model Page 561

```
##Example 11.3.1 Pg.561
```

```
##Stepwise regression model

y = c(45,65,73,63,83,45,60,73,74,69,66,69,71,70,79,83,75,67,67,52,52,6
x1 = c(74,65,71,64,79,56,68,76,83,62,54,61,63,84,78,65,86,61,71,59,71,6
x2 = c(29,50,67,44,55,48,41,49,71,44,52,46,56,82,53,49,63,64,45,67,32,5
x3 = c(40,64,79,57,76,54,66,65,77,57,67,66,67,68,82,82,79,75,67,64,44,7
x4 = c(66,68,81,59,76,59,71,75,76,67,63,84,60,84,84,65,84,60,80,69,48,7
x5 = c(93,74,87,85,84,50,69,67,84,81,68,75,64,78,78,55,80,81,86,79,65,8
x6 = c(47,49,33,37,33,42,37,43,33,43,36,43,35,37,39,38,41,45,48,54,43,4

step(lm(y~x1+x2+x3+x4+x5+x6),direction = "both") # performs stepwise rergression

#x1,x2,x3,x6 variables are selected
```

R code Exa 11.4.1 Logistic regression page 572

```
## Example 11.4.1 Page 572
##Logistic regression

cases <- c(21,20,92,15)
```

```
disease <- c(1,1,0,0)
disease = factor(disease,labels = c("present","
   absent"))
sex <- c(1,2,1,2)
sex = factor(sex,labels=c("male","female"))
dt = data.frame(disease,sex,cases)
dt
xtabs(cases~.,dt)  #creates contingency table
fit <- glm(disease~sex, weights = cases, data = dt,
   family = "binomial")  #logistic regression
summary(fit)

##summary gives estimated calue for sex and
   intercept
##pvalue < 0.05, hence significant
```

R code Exa 11.4.2 Logistic regression page 573

```
##Example 11.4.2 Pg.573
##Logistic regression

age<- c(50,59,42,50,34,49,67,44,53,45,79,
        46,62,58,70,60,67,64,62,50,61,69,
        74,65,80,69,77,61,72,67,73,75,71,
        69,78,69,74,86,49,63,63,72,64,72,
        64,72,79,75,70,73,66,75,73,71,72,
        69,76,60,79,78,62,73,46,57,53,40,
        73,68,72,59,64,78,68,67,55,71,80,
        75,69,80,79,71,69,78,75,71,69,77,
        81,78,76,84,74,59,81,74,77,59,75,
        68,81,74,65,81,62,85,84,39,52,67,
        82,84,79,81,74,85,92,69,83,82,85,
        82,74,50,55,66,49,55,73,41,64,
        46,65,50,61,64,59,73,73,65,67,60,
        69,61,79,66,68,61,63,70,68,59,64,
```

```
        62,74,61,69,76,71,61,46,69,66,57,
        60,63,63,56,70,70,63,63,65,67,68,
        84,69,78,69,79,83,67,47,57,66)

status <- c(rep(0,122),rep(1,63))
status
status1 = factor(status,labels = c("nonparticipating","participating"),levels=c(0,1))
status1

fit <- glm(status1~1+age,family="binomial",control=glm.control(maxit=50))  #logistic regression
summary(fit)
##summary gives the estimates of intercept and age
##Also the p value to test the slope coefficient

##A function to estimate probabilities from logistic model
est_prob <- function(x)
{
  pred = predict(fit,newdata=data.frame(age=x))
  prob = exp(pred)/(1+exp(pred))
  print(prob)
}

est_prob(x=50)
est_prob(x=age)
plot(age,est_prob(x=age))
```

Chapter 12

Analysis of frequency data An introduction to the chi square distribution

R code Exa 12.3.2 Tests of goodness of fit Binomial distribution

```
##Exaxmple 12.3.2 Pg.609
## Tests of goodness of fit - Binomial distribution

x <- 0:10   #no. of paients out of 25 preferring new
    pain reliever
f <- c(5,6,8,10,10,15,17,10,10,9,0)  #no. of doctors
    reporting this number
N <- c(0,6,16,30,40,75,102,70,80,81,0) #total number
    of patients preferring new pain reliever by
   doctor

p = sum(N)/(25*sum(f)) ; p

prob = dbinom(x,25,p) ;prob

pooled1 = f[1] + f[2]   #pooling first two values
   since <5
```

```
Obs_f = c(pooled1,f[c(-1,-2)])  ;Obs_f
Exp_f = sum(f)*prob
pooled2 = Exp_f[1] + Exp_f[2]   #pooling first two values since <5
Exp_f = c(pooled2,Exp_f[c(-1,-2)]) ;Exp_f

dt = data.frame(Obs_f,Exp_f)
dt
sum(Obs_f)
sum(Exp_f)

chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
p_val = pchisq(0.005,length(x)-2)  ;p_val

#Since pval < 0.005, we conclude data came from binomial distribution
#Answers slightly differ by decimals due to aproximations
```

R code Exa 12.3.3 Test of goodness of fit Poisson distribution

```
##Exaxmple 12.3.3 Pg.611
## Tests of goodness of fit - Poisson distribution

x <- 0:10
f <- c(5,14,15,23,16,9,3,3,1,1,0)
lambda = sum(x*f)/sum(f) ;lambda    #mean=lambda
prob = dpois(x,lambda) ;prob

pooled1 = f[9] + f[10] + f[11]   #pooling last three values since <5
Obs_f = c(f[c(-9,-10,-11)],pooled1)  ;Obs_f
Exp_f = sum(f)*prob
pooled2 = Exp_f[9] + Exp_f[10] + Exp_f[11]   #pooling first two values since <5
```

```
Exp_f = c(Exp_f[c(-9,-10,-11)],pooled2) ;Exp_f

dt= data.frame(Obs_f,Exp_f)
dt
sum(Obs_f);sum(Exp_f)
chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
p_val = pchisq(0.05,length(x)-3)  ;p_val

#Since pval < 0.005, we conclude data came from
    poisson distribution
#Answer slightly differ by decimal due to
    approximation
```

R code Exa 12.3.4 Tests of goodness of fit Uniform distribution Page 614

```
##Exaxmple 12.3.4 Pg.614
## Tests of goodness of fit - Uniform distribution

x <- c("Dec","Jan","Feb","Mar","Apr")
f <- c(62,84,17,16,21)
prop <- 1/length(x)

Obs_f = f  ;Obs_f
Exp_f = prop*sum(f)  ;Exp_f

chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
p_val = pchisq(0.05,length(x)-1)  ;p_val

#Since pval < 0.005, we conclude data came from
    uniform distribution
```

R code Exa 12.3.5 Goodness of fit Distribution of traits Page 616

```
##Exaxmple 12.3.5 Pg.616
##Goodness of fit – Distribution of traits

n = 200
dominant = 43
heterozygous = 125
recessive = 32

Obs_f = c(43,125,32)
Exp_f = c(50,100,50)  #1:2:1 ratio

chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
chi_critical = qchisq(0.95,2) ;chi_critical
p_val = pchisq(0.05,2)  ;p_val

#Since pval < 0.005, we conclude data came from 1:2:1 ratio distribution
```

R code Exa 12.4.1 Test for independence Page 621

```

##Example 12.4.1 Pg.621
##Test for independence

x<-matrix(c(260,299,15,41,7,14),nrow=3,byrow=T)
x
rownames(x)<-c("White","Black","Other")
colnames(x)<-c("Yes","No")
print(x)

chisq.test(x)

#pvalue < 0.05 , hence there is relationship between race and folic acid
```

R code Exa 12.4.2 Test for independence Page 626

```
##Example 12.4.2 Pg.626
##Test for independence

x<-matrix(c(131,52,14,36),nrow=2,byrow=T)
x
rownames(x)<-c("Fallers","Non fallers")
colnames(x)<-c("Yes","No")
print(x)

chisq.test(x)

#pvalue < 0.05 , hence there is relationship between
    experiencing a fall and change in lifestyle
```

R code Exa 12.5.1 Test for homogeneity Page 631

```
##Example 12.5.1 Pg.631
##Test for homogeneity

x<-matrix(c(21,75,19,77),nrow=2,byrow=T)
x
rownames(x)<-c("Narcoleptic","Controls")
colnames(x)<-c("Yes","No")
print(x)

chisq.test(x)

#pvalue > 0.05 , hence two populations may be
   homogenous wrt migraine frequency
#Answer is slightly differing from the textbook
```

R code Exa 12.6.1 Fishers exact test Page 638

```
##Example 12.6.1 Pg.638
##Fishers exact test

x<-matrix(c(9,2,12,8),nrow=2,byrow=T)
x
rownames(x)<-c("Naive","Experienced")
colnames(x)<-c("Yes","No")
print(x)

fisher.test(x)

#pvalue > 0.05 , hence rate of remaining on regimen
    for 120 weeks is same for naive and experienced
    groups
#Answer is slightly differing from the textbook
```

R code Exa 12.7.1 Relative risk Page 644

```
##Example 12.7.1 Pg.644
##Relative risk

x<-matrix(c(18,199,22,216),nrow=2,byrow=T)
x
rownames(x)<-c("Not exercising","Extreme exercising"
    )
colnames(x)<-c("cases","non cases")
print(x)

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

```
library(mosaic)

relrisk(x)
```

R code Exa 12.7.2 Odds ratio Page 647

```
##Example 12.7.2 Pg.647
##Odds ratio

x<-matrix(c(68,3496,64,342),nrow=2,byrow=T)
x
rownames(x)<-c("Never Smoked"," smoked")
colnames(x)<-c("cases","non cases")
print(x)

#install.packages("mosaic")
library(mosaic)

oddsRatio(x)
```

Chapter 13

Special techniques for use when population parameters and population distributions are unknown

R code Exa 13.3.1 Sign test Page 673

```
##Example 13.3.1 Pg.673
##Sign test

score <- c(4,5,8,8,9,6,10,7,6,6)
md= median(score)
diff<-score-md
diff
sdiff<-sign(diff)
sdiff
s=length(sdiff[sdiff==1])
s
cv=pbinom(1,9,0.5)
cv
pval = 2*cv ;pval

```

```
#since pvalue<0.05, we conclude median score is not
    5
```

R code Exa 13.3.2 Sign test for paired data Page 677

```
##Example 13.3.2 Pg.677
##Sign test for paired data

score_x <- c(1.5,2,3.5,3,3.5,2.5,2,1.5,1.5,2,3,2)
score_y <- c(2,2,4,2.5,4,3,3.5,3,2.5,2.5,2.5,2.5)
diff<-score_x - score_y
diff
sdiff<-sign(diff)
sdiff
s=length(sdiff[sdiff==1])
s
cv=pbinom(s,11,0.5)
cv

#since pvalue<0.05, instruction was beneficial
```

R code Exa 13.4.1 Wilcoxon signed rank test page 683

```
##Example 13.4.1 Pg.683
##Wilcoxon signed rank test

cardiac_out <- c
    (4.91,4.10,6.74,7.27,7.42,7.50,6.56,4.64,5.98,3.14,3.23,5.80,6.17

wilcox.test(cardiac_out, alternative= "two.sided",
    conf.int=T)

```

```
#since pvalue >0.05, we conclude population mean may
   be 5.05
```

R code Exa 13.5.1 Median Test page 686

```

##Example 13.5.1 Pg.686
##Median Test

urban <-c
   (35,26,27,21,27,38,23,25,25,27,45,46,33,26,46,41)
rural <-c(29,50,43,22,42,47,42,32,50,37,34,31)
z<-c(urban,rural)
n<-length(z) ;n
u<-median(z)
a<-length(urban[urban>u]) ;a
b<-length(rural[rural>u]) ;b
c<-length(urban[urban<=u]) ;c
d<-length(rural[rural<=u]) ;d

chi<-(n*(a*d-b*c)^2)/((a+b)*(c+d)*(a+c)*(b+d))  ;chi

chi_critical <- qchisq(1-0.05,1,lower.tail = T) ;chi
   _critical
pval<- pchisq(chi,1) ;pval

#pval > 0.05, hence two saamples may have been drawn
   from populations with equal median
```

R code Exa 13.6.1 Mann Whitney test Page 691

```
##Example 13.6.1 Pg.691
##Mann Whitney test
```

```
exposed <-c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,15.7,16.7,13.7
unexposed <-c(17.4,16.2,17.1,17.5,15,16,16.9,15,16.3,16.8)

wilcox.test(exposed, unexposed, conf.level=0.95,conf.int=T)  #for mann whitney test
```

R code Exa 13.7.1 Kolmogorov Smirnov Goodness of fit test Page 699

```
##Example 13.7.1 Pg.699
##Kolmogorov Smirnov Goodness of fit test

values <- c(75,84,80,77,68,87,92,77,92,86,78,76,80,81,72,77,92,80,80,77,77,9
ks.test(values, "pnorm", mean(values), sd(values), alternative = "two.sided")

#pvalue > 0.05, hence sample would have come from normal distribution
```

R code Exa 13.8.1 Kruskal Wallis one way ANOVA Page 705

```
##Example 13.8.1 Pg.705
##Kruskal Wallis one way ANOVA

cell_count <-c(12.22,28.44,28.13,38.69,54.91,3.68,4.05,6.47,21.12,3.33,54.36,27
group<-c(rep(1,5),rep(2,5),rep(3,5))
```

```
group
group1<-factor(group,labels=c("air","benzaldehyde","acetaldehyde"))
group1
dt<-data.frame(group1,cell_count)
dt
kruskal.test(cell_count~group1)

#pval <0.05, hence there is a difference in the average cell count among three groups
```

R code Exa 13.8.2 Kruskal Wallis one way ANOVA page 708

```
##Example 13.8.2 Pg.708
##Kruskal Wallis one way ANOVA

book_value <-c(1735,1520,1476,1688,1702,2667,1575,1602,1530,1698,5260,4455,4480
group<-c(rep(1,10),rep(2,8),rep(3,9),rep(4,7),rep(5,7))
group
group1<-factor(group,labels=c("A","B","C","D","E"))
group1
dt<-data.frame(group1,book_value)
dt
kruskal.test(book_value~group1)

#pval <0.05, hence there is a difference in the average book value among five groups
```

R code Exa 13.9.1 Friedman test Page 713

```r
##Example 13.9.1 Pg.713
##Friedman test

ranks_A <-c(2,2,2,1,3,1,2,1,1)
ranks_B <-c(3,3,3,3,2,2,3,3,3)
ranks_C <-c(1,1,1,2,1,3,1,2,2)
n = 9 ; k=3
friedman = (12/(n*k*(k+1)))*(sum(ranks_A)^2+sum(ranks_B)^2+sum(ranks_C)^2) - 3*n*(k+1)
friedman
pval = pchisq(0.05/2,k-1) ;pval

#pval <0.05, hence the three models of low volt electrical stimulator are not equally preferred
```

R code Exa 13.9.2 Friedman test Page 715

```r
##Example 13.9.2 Pg.715
##Friedman test

salivaryflow <-
  matrix(c(29,48,75,100,72,30,100,100,70,100,86,96,54,35,90,99,5,43,32,81,
         nrow = 16,
         byrow = TRUE,
         dimnames = list(1 : 16,
                         c("A", "B", "C","D")))
friedman.test(salivaryflow)

#pval <0.05, hence there is a difference in the salivary flow among four groups
```

R code Exa 13.10.1 Test for correlation Page 720

```
##Example 13.10.1 Pg.720
##Test for correlation

age <- c
   (20,21,22,24,27,30,31,33,35,38,40,42,44,46,48,51,53,55,58,60)

eeg <- c
   (98,75,95,100,99,65,64,70,85,74,68,66,71,62,69,54,63,52,67,55)


cor.test(age,eeg)

#pvalue < 0.05, hence age and eeg are inversely
   related
```

R code Exa 13.10.2 Test for correlation Page 722

```
##Example 13.10.2 Pg.722
##Test for correlation

age <- c
   (82,85,83,64,82,53,26,47,37,49,65,40,32,50,62,33,36,53,50,71,54,6

conc <- c
   (169.62,48.94,41.16,63.95,21.09,5.40,6.33,4.26,3.62,4.82,108.22,1


cor.test(age,conc)

#pvalue < 0.05, hence age and mineral concentration
   are inversely related
```