

R Textbook Companion for
Biostatistics: Basic Concepts and
Methodology for the Health Sciences
by Daniel W. Wayne, Chad L. Cross¹

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

May 26, 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: 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	4
1 Getting acquainted with biostatistics	5
2 Strategies for understanding the meanings of data	8
3 Probability The basis of statistical inference	20
4 Probabilistic features of certain data distributions	27
5 Probabilistic features of the distributions of certain sample statistics	40
6 Using sample data to make estimates about population parameters	44
7 Using sample statistics to test hypothesis about population parameters	52
8 Statistical inference and the analysis of data variability	61
9 Statistical inference and the relationship between two variables	65
10 Statistical inference and the relationships among three or more variables	70
11 Additional techniques for the analysis of relationships among variables	77

12	Analysis of frequency data An introduction to the chi square distribution	85
13	Special techniques for use when population parameters and population distributions are unknown	92

List of R Codes

Exa 1.4.1	Simple random sampling of size 10 Page 8	5
Exa 1.4.2	Systematic sample of size 10 Page 11	6
Exa 2.2.1	Arrange ages from smallest to largest Page 20	8
Exa 2.3.1	Form Class Intervals for the ages data Page 23	9
Exa 2.3.2	Stem and leaf plot for ages of 189 subjects Page 29	10
Exa 2.4.1	Obtain the mean age page 38	11
Exa 2.4.2	Mean age of 10 subjects Page 39	11
Exa 2.4.3	Median age Page 40	12
Exa 2.4.4	Median age of sample of size 10 Page 40	13
Exa 2.4.5	Modal age Page 41	14
Exa 2.4.6	Skewness Page 42	15
Exa 2.5.1	Range of ages Page 44	16
Exa 2.5.2	Variance of ages Page 44	17
Exa 2.5.3	Coefficient of correlation Page 46	17
Exa 2.5.4	Kurtosis Page 49	18
Exa 2.5.5	Box and whisker plot Page 50	19
Exa 3.4.1	Probability of member being less than 18 years of age Page 69	20
Exa 3.4.2	Conditional Probability Page 70	20
Exa 3.4.3	Joint Probability of early subjects and no history of mood disorders Page 71	21
Exa 3.4.4	Multiplication rule of early subjects and no history of mood disorders Page 71	22
Exa 3.4.5	Conditional Probability of early subjects and no history of mood disorders Page 72	22
Exa 3.4.6	Probability of early subjects OR no history of mood disorders Page 73	23

Exa 3.4.7	Conditional Probability of student wears eye glasses given he is a boy Page 74	24
Exa 3.4.8	Find probability of admissions that are NOT private Page 75	24
Exa 3.4.9	Marginal Probability of Early age Page 75	25
Exa 3.5.1	Bayes theorem Refer Table for data Page 81	26
Exa 4.2.1	Probability distribution from frequency table Page 93	27
Exa 4.2.2	Probability distribution from frequency table Page 95	28
Exa 4.2.3	Prob that family used either one or two programs Page 95	28
Exa 4.2.4	Cumulative probability distribution Page 97	29
Exa 4.2.5	Cumulative probability distribution Page 97	29
Exa 4.2.6	Cumulative probability distribution Page 97	30
Exa 4.2.7	Cumulative probability distribution Page 97	30
Exa 4.2.8	mean and variance of prob distribution Page 98	31
Exa 4.3.1	Binomial distribution Page 99	32
Exa 4.3.2	Binomial probability distribution Page 103	32
Exa 4.3.3	Binomial probabilities Page 103	32
Exa 4.3.4	Binomial probabilities Page 105	33
Exa 4.4.1	Poisson distribution Page 110	33
Exa 4.4.2	Poisson distribution Page 110	34
Exa 4.4.3	Poisson distribution Page 110	34
Exa 4.4.4	Poisson distribution Page 111	34
Exa 4.4.5	Poisson distribution Page 112	35
Exa 4.6.1	Standard Normal Distribution Page 119	35
Exa 4.6.2	Standard Normal Distribution Page 120	35
Exa 4.6.3	Standard Normal Distribution Page 121	36
Exa 4.6.4	Standard Normal Distribution Page 121	37
Exa 4.6.5	Standard normal distribution Page 122	37
Exa 4.7.1	Normal distribution Page 123	38
Exa 4.7.2	Normal distribution Page 125	38
Exa 4.7.3	Normal distribution Page 126	39
Exa 5.3.2	Sampling distribution of mean Page 142	40
Exa 5.3.3	Sampling distribution of mean Page 143	41
Exa 5.4.1	Sampling distribution of two means Page 145	41
Exa 5.4.2	Sampling distribution of two means Page 148	42
Exa 5.5.2	Sampling distribution of Proportion Page 152	42
Exa 5.6.1	Sampling distribution of Two Proportions Page 155	42

Exa 5.6.2	Sampling distribution of Two Proportions Page 155 . .	43
Exa 6.2.1	Confidence interval for population mean μ Page 166	44
Exa 6.2.2	Confidence interval for population mean μ Page 168	44
Exa 6.2.3	Confidence interval for population mean μ Page 168	45
Exa 6.2.4	Confidence interval for population mean μ Page 169	45
Exa 6.3.1	Confidence interval for mean T test Page 173	46
Exa 6.4.1	Confidence interval for two means Z test Page 177 . .	46
Exa 6.4.2	Confidence interval for two means Z test Page 178 . .	47
Exa 6.4.3	Confidence interval for two means T test Page 180 . .	47
Exa 6.4.4	Confidence interval for two means T test Page 181 . .	48
Exa 6.5.1	Confidence interval for population proportion Page 185	49
Exa 6.6.1	Confidence interval for two population proportions Page 187	49
Exa 6.7.1	Determine the sample size n Page 190	50
Exa 6.8.1	Determine the sample size n Page 192	50
Exa 6.9.1	Chi square test Page 196	50
Exa 6.10.1	F test Page 200	51
Exa 7.2.1	Test for mean when population variances are known Page 222	52
Exa 7.2.2	Test for mean when population variances are known Page 226	52
Exa 7.2.3	Test for mean when population variances are unknown Page 228	53
Exa 7.2.4	Test for mean when population variances are known Page 231	54
Exa 7.2.5	Test for mean when population variances are unknown Page 232	54
Exa 7.3.1	Test for two means when population variances are known Page 237	55
Exa 7.3.2	Test for means when population variances are unknown Page 239	55
Exa 7.3.3	Test for means when population variances are unknown Page 240	56
Exa 7.3.4	Test for two means when population variances are known Page 242	56
Exa 7.4.1	Paired t test Page 251	57
Exa 7.5.1	Test for proportions Page 258	57
Exa 7.6.1	Test for two proportions Page 261	58

Exa 7.7.1	Test for single population variance Page 264	58
Exa 7.8.1	Test for ratio of two population variances Page 268 . .	59
Exa 7.8.2	Test for ratio of two population variances Page 270 . .	59
Exa 8.2.1	One way ANOVA Page 318	61
Exa 8.2.2	One way ANOVA and Tukeys HSD Page 325	62
Exa 8.3.1	Ranadomized Block Design Two way ANOVA Page 339	63
Exa 9.3.1	Linear regression model Page 417	65
Exa 9.4.1	Linear regression model Coefficient of determination Page 432	66
Exa 9.4.2	Linear regression model Test for slope Page 436	66
Exa 9.4.3	Linear regression model Residual plot Page 440	67
Exa 9.7.1	Correlation and linear regression model Page 447 . . .	68
Exa 9.7.2	Correlation test Page 452	68
Exa 10.3.1	Multiple regression equation Page 493	70
Exa 10.4.1	Coefficient of multiple determination Page 502	71
Exa 10.4.2	Test for parameters Page 504	71
Exa 10.4.3	Test for parameter of variable Age Page 505	72
Exa 10.5.1	confidence interval for CDA Page 508	73
Exa 10.6.1	Multiple correlation coefficient Page 511	74
Exa 10.6.2	Partial correlation coefficient Page 515	74
Exa 11.1.1	Box and whiskers plot	77
Exa 11.1.2	Correlation for 3 variables Page 542	77
Exa 11.2.1	Regression model for categorical data Page 545	78
Exa 11.2.2	Test for model parameter for categorical data	80
Exa 11.3.1	Stepwise regression model Page 561	81
Exa 11.4.1	Logistic regression page 572	82
Exa 11.4.2	Logistic regression page 573	83
Exa 12.3.2	Tests of goodness of fit Binomial distribution	85
Exa 12.3.3	Test of goodness of fit Poisson distribution	86
Exa 12.3.4	Tests of goodness of fit Uniform distribution Page 614	87
Exa 12.3.5	Goodness of fit Distribution of traits Page 616	87
Exa 12.4.1	Test for independence Page 621	88
Exa 12.4.2	Test for independence Page 626	89
Exa 12.5.1	Test for homogeneity Page 631	89
Exa 12.6.1	Fishers exact test Page 638	90
Exa 12.7.1	Relative risk Page 644	90
Exa 12.7.2	Odds ratio Page 647	91
Exa 13.3.1	Sign test Page 673	92

Exa 13.3.2 Sign test for paired data Page 677	93
Exa 13.4.1 Wilcoxon signed rank test page 683	93
Exa 13.5.1 Median Test page 686	94
Exa 13.6.1 Mann Whitney test Page 691	94
Exa 13.7.1 Kolmogorov Smirnov Goodness of fit test Page 699 . .	95
Exa 13.8.1 Kruskal Wallis one way ANOVA Page 705	95
Exa 13.8.2 Kruskal Wallis one way ANOVA page 708	96
Exa 13.9.1 Friedman test Page 713	96
Exa 13.9.2 Friedman test Page 715	97
Exa 13.10.1 Test for correlation Page 720	98
Exa 13.10.2 Test for correlation Page 722	98

Chapter 1

Getting acquainted with biostatistics

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

```
1 ##Example 1.4.1 Pg.8
2 ##Simple random sampling of size 10
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
5         44,39,38,49,49,53,56,57,51,61,
6         53,66,71,75,72,65,67,38,37,46,
7         44,44,48,49,30,45,47,45,48,47,
8         47,44,48,43,45,40,48,49,38,44,
9         43,47,46,57,52,54,56,53,64,53,
10        58,54,59,56,62,50,64,53,61,53,
11        62,57,52,54,61,59,57,52,54,53,
12        62,52,62,57,59,59,56,57,53,59,
13        61,55,61,56,52,54,51,50,50,55,
14        63,50,59,54,60,50,56,68,66,71,
15        82,68,78,66,70,66,78,69,71,69,
16        78,66,68,71,69,77,76,71,43,47,
17        48,37,40,42,38,49,43,46,34,46,
18        46,48,47,43,52,53,61,60,53,53,
19        50,53,54,61,61,61,64,53,53,54,
```

```

20         61,60,51,50,53,64,64,53,60,54,
21         55,58,62,62,54,53,61,54,51,62,
22         57,50,64,63,65,71,71,73,66)
23
24 length(age)
25 set.seed(12)
26 srs = sample(age,10)
27 srs
28 ##Answers change because of random sampling

```

R code Exa 1.4.2 Systematic sample of size 10 Page 11

```

1 ##Example 1.4.2 Pg.11
2 ##Systematic sample of size 10
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
5         44,39,38,49,49,53,56,57,51,61,
6         53,66,71,75,72,65,67,38,37,46,
7         44,44,48,49,30,45,47,45,48,47,
8         47,44,48,43,45,40,48,49,38,44,
9         43,47,46,57,52,54,56,53,64,53,
10        58,54,59,56,62,50,64,53,61,53,
11        62,57,52,54,61,59,57,52,54,53,
12        62,52,62,57,59,59,56,57,53,59,
13        61,55,61,56,52,54,51,50,50,55,
14        63,50,59,54,60,50,56,68,66,71,
15        82,68,78,66,70,66,78,69,71,69,
16        78,66,68,71,69,77,76,71,43,47,
17        48,37,40,42,38,49,43,46,34,46,
18        46,48,47,43,52,53,61,60,53,53,
19        50,53,54,61,61,61,64,53,53,54,
20        61,60,51,50,53,64,64,53,60,54,
21        55,58,62,62,54,53,61,54,51,62,
22        57,50,64,63,65,71,71,73,66)
23

```

```
24 sys.sample = function(N,n,r){
25   k = round(N/n)
26   #ceiling(x) rounds to the nearest integer that's
      larger than x.
27   sys.samp = seq(r, r + k*(n-1), k)
28   print(sys.samp)
29 }
30
31 N = 185; n=10; r=4
32 sys = sys.sample(N, n, r)
33 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

```
1 ##Example 2.2.1 Pg.20
2 ##Arrange ages from smallest to largest
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
5         44,39,38,49,49,53,56,57,51,61,
6         53,66,71,75,72,65,67,38,37,46,
7         44,44,48,49,30,45,47,45,48,47,
8         47,44,48,43,45,40,48,49,38,44,
9         43,47,46,57,52,54,56,53,64,53,
10        58,54,59,56,62,50,64,53,61,53,
11        62,57,52,54,61,59,57,52,54,53,
12        62,52,62,57,59,59,56,57,53,59,
13        61,55,61,56,52,54,51,50,50,55,
14        63,50,59,54,60,50,56,68,66,71,
15        82,68,78,66,70,66,78,69,71,69,
16        78,66,68,71,69,77,76,71,43,47,
17        48,37,40,42,38,49,43,46,34,46,
18        46,48,47,43,52,53,61,60,53,53,
19        50,53,54,61,61,61,64,53,53,54,
```

```

20         61,60,51,50,53,64,64,53,60,54,
21         55,58,62,62,54,53,61,54,51,62,
22         57,50,64,63,65,71,71,73,66)
23 sort(age)

```

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

```

1 ##Example 2.3.1 Pg.23
2 ##Form Class Intervals for the ages data
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
5         44,39,38,49,49,53,56,57,51,61,
6         53,66,71,75,72,65,67,38,37,46,
7         44,44,48,49,30,45,47,45,48,47,
8         47,44,48,43,45,40,48,49,38,44,
9         43,47,46,57,52,54,56,53,64,53,
10        58,54,59,56,62,50,64,53,61,53,
11        62,57,52,54,61,59,57,52,54,53,
12        62,52,62,57,59,59,56,57,53,59,
13        61,55,61,56,52,54,51,50,50,55,
14        63,50,59,54,60,50,56,68,66,71,
15        82,68,78,66,70,66,78,69,71,69,
16        78,66,68,71,69,77,76,71,43,47,
17        48,37,40,42,38,49,43,46,34,46,
18        46,48,47,43,52,53,61,60,53,53,
19        50,53,54,61,61,61,64,53,53,54,
20        61,60,51,50,53,64,64,53,60,54,
21        55,58,62,62,54,53,61,54,51,62,
22        57,50,64,63,65,71,71,73,66)
23
24 breaks = seq(30,90,by=10)
25 breaks
26 CI = cut(age,breaks = breaks,right=F)
27 CI
28 table(CI) #gives class intervals along with

```

```

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

```

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

```

1
2 ##Example 2.3.2 Pg.29
3 ##Stem and leaf plot for ages of 189 subjects
4
5 age <- c(48,35,46,44,43,42,39,44,49,49,
6          44,39,38,49,49,53,56,57,51,61,
7          53,66,71,75,72,65,67,38,37,46,
8          44,44,48,49,30,45,47,45,48,47,
9          47,44,48,43,45,40,48,49,38,44,
10         43,47,46,57,52,54,56,53,64,53,
11         58,54,59,56,62,50,64,53,61,53,
12         62,57,52,54,61,59,57,52,54,53,
13         62,52,62,57,59,59,56,57,53,59,
14         61,55,61,56,52,54,51,50,50,55,
15         63,50,59,54,60,50,56,68,66,71,
16         82,68,78,66,70,66,78,69,71,69,
17         78,66,68,71,69,77,76,71,43,47,
18         48,37,40,42,38,49,43,46,34,46,
19         46,48,47,43,52,53,61,60,53,53,
20         50,53,54,61,61,61,64,53,53,54,
21         61,60,51,50,53,64,64,53,60,54,
22         55,58,62,62,54,53,61,54,51,62,
23         57,50,64,63,65,71,71,73,66)
24 stem(age, scale=0.5)

```

R code Exa 2.4.1 Obtain the mean age page 38

```
1 ##Example 2.4.1 Pg.38
2 ##Obtain the mean age
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
5         44,39,38,49,49,53,56,57,51,61,
6         53,66,71,75,72,65,67,38,37,46,
7         44,44,48,49,30,45,47,45,48,47,
8         47,44,48,43,45,40,48,49,38,44,
9         43,47,46,57,52,54,56,53,64,53,
10        58,54,59,56,62,50,64,53,61,53,
11        62,57,52,54,61,59,57,52,54,53,
12        62,52,62,57,59,59,56,57,53,59,
13        61,55,61,56,52,54,51,50,50,55,
14        63,50,59,54,60,50,56,68,66,71,
15        82,68,78,66,70,66,78,69,71,69,
16        78,66,68,71,69,77,76,71,43,47,
17        48,37,40,42,38,49,43,46,34,46,
18        46,48,47,43,52,53,61,60,53,53,
19        50,53,54,61,61,61,64,53,53,54,
20        61,60,51,50,53,64,64,53,60,54,
21        55,58,62,62,54,53,61,54,51,62,
22        57,50,64,63,65,71,71,73,66)
23 mean(age)
```

R code Exa 2.4.2 Mean age of 10 subjects Page 39

```
1 ##Example 2.4.2 Pg.39
2 ##Mean age of 10 subjects
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
```

```

5      44,39,38,49,49,53,56,57,51,61,
6      53,66,71,75,72,65,67,38,37,46,
7      44,44,48,49,30,45,47,45,48,47,
8      47,44,48,43,45,40,48,49,38,44,
9      43,47,46,57,52,54,56,53,64,53,
10     58,54,59,56,62,50,64,53,61,53,
11     62,57,52,54,61,59,57,52,54,53,
12     62,52,62,57,59,59,56,57,53,59,
13     61,55,61,56,52,54,51,50,50,55,
14     63,50,59,54,60,50,56,68,66,71,
15     82,68,78,66,70,66,78,69,71,69,
16     78,66,68,71,69,77,76,71,43,47,
17     48,37,40,42,38,49,43,46,34,46,
18     46,48,47,43,52,53,61,60,53,53,
19     50,53,54,61,61,61,64,53,53,54,
20     61,60,51,50,53,64,64,53,60,54,
21     55,58,62,62,54,53,61,54,51,62,
22     57,50,64,63,65,71,71,73,66)
23 set.seed(12)
24 srs = sample(age,10)
25 srs
26 mean(srs)    ##different answer due to a different
               random sample

```

R code Exa 2.4.3 Median age Page 40

```

1 ##Example 2.4.3 Pg.40
2 ##Median age
3
4 age <- c(48,35,46,44,43,42,39,44,49,49,
5          44,39,38,49,49,53,56,57,51,61,
6          53,66,71,75,72,65,67,38,37,46,
7          44,44,48,49,30,45,47,45,48,47,
8          47,44,48,43,45,40,48,49,38,44,
9          43,47,46,57,52,54,56,53,64,53,

```

```

10      58,54,59,56,62,50,64,53,61,53,
11      62,57,52,54,61,59,57,52,54,53,
12      62,52,62,57,59,59,56,57,53,59,
13      61,55,61,56,52,54,51,50,50,55,
14      63,50,59,54,60,50,56,68,66,71,
15      82,68,78,66,70,66,78,69,71,69,
16      78,66,68,71,69,77,76,71,43,47,
17      48,37,40,42,38,49,43,46,34,46,
18      46,48,47,43,52,53,61,60,53,53,
19      50,53,54,61,61,61,64,53,53,54,
20      61,60,51,50,53,64,64,53,60,54,
21      55,58,62,62,54,53,61,54,51,62,
22      57,50,64,63,65,71,71,73,66)
23 median(age)

```

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

```

1  ##Example 2.4.4 Pg.40
2  ##Median age of sample of size 10
3
4  age <- c(48,35,46,44,43,42,39,44,49,49,
5           44,39,38,49,49,53,56,57,51,61,
6           53,66,71,75,72,65,67,38,37,46,
7           44,44,48,49,30,45,47,45,48,47,
8           47,44,48,43,45,40,48,49,38,44,
9           43,47,46,57,52,54,56,53,64,53,
10          58,54,59,56,62,50,64,53,61,53,
11          62,57,52,54,61,59,57,52,54,53,
12          62,52,62,57,59,59,56,57,53,59,
13          61,55,61,56,52,54,51,50,50,55,
14          63,50,59,54,60,50,56,68,66,71,
15          82,68,78,66,70,66,78,69,71,69,
16          78,66,68,71,69,77,76,71,43,47,
17          48,37,40,42,38,49,43,46,34,46,
18          46,48,47,43,52,53,61,60,53,53,

```

```

19         50,53,54,61,61,61,64,53,53,54,
20         61,60,51,50,53,64,64,53,60,54,
21         55,58,62,62,54,53,61,54,51,62,
22         57,50,64,63,65,71,71,73,66)
23 set.seed(12)
24 srs = sample(age,10)
25 srs
26 median(srs)
27 ##different answer due to a different random sample

```

R code Exa 2.4.5 Modal age Page 41

```

1
2 ##Example 2.4.5 Pg.41
3 ##Modal age
4
5 age <- c(48,35,46,44,43,42,39,44,49,49,
6         44,39,38,49,49,53,56,57,51,61,
7         53,66,71,75,72,65,67,38,37,46,
8         44,44,48,49,30,45,47,45,48,47,
9         47,44,48,43,45,40,48,49,38,44,
10        43,47,46,57,52,54,56,53,64,53,
11        58,54,59,56,62,50,64,53,61,53,
12        62,57,52,54,61,59,57,52,54,53,
13        62,52,62,57,59,59,56,57,53,59,
14        61,55,61,56,52,54,51,50,50,55,
15        63,50,59,54,60,50,56,68,66,71,
16        82,68,78,66,70,66,78,69,71,69,
17        78,66,68,71,69,77,76,71,43,47,
18        48,37,40,42,38,49,43,46,34,46,
19        46,48,47,43,52,53,61,60,53,53,
20        50,53,54,61,61,61,64,53,53,54,
21        61,60,51,50,53,64,64,53,60,54,
22        55,58,62,62,54,53,61,54,51,62,
23        57,50,64,63,65,71,71,73,66)

```



```

21   cat("Skewness = " ,skewness(x))
22 }
23 d1 = descriptive(no_skew)
24 d2 = descriptive(right_skew)
25 d3 = descriptive(left_skew)

```

R code Exa 2.5.1 Range of ages Page 44

```

1  ##Example 2.5.1 Pg.44
2  ##Range of ages
3
4  age <- c(48,35,46,44,43,42,39,44,49,49,
5           44,39,38,49,49,53,56,57,51,61,
6           53,66,71,75,72,65,67,38,37,46,
7           44,44,48,49,30,45,47,45,48,47,
8           47,44,48,43,45,40,48,49,38,44,
9           43,47,46,57,52,54,56,53,64,53,
10          58,54,59,56,62,50,64,53,61,53,
11          62,57,52,54,61,59,57,52,54,53,
12          62,52,62,57,59,59,56,57,53,59,
13          61,55,61,56,52,54,51,50,50,55,
14          63,50,59,54,60,50,56,68,66,71,
15          82,68,78,66,70,66,78,69,71,69,
16          78,66,68,71,69,77,76,71,43,47,
17          48,37,40,42,38,49,43,46,34,46,
18          46,48,47,43,52,53,61,60,53,53,
19          50,53,54,61,61,61,64,53,53,54,
20          61,60,51,50,53,64,64,53,60,54,
21          55,58,62,62,54,53,61,54,51,62,
22          57,50,64,63,65,71,71,73,66)
23 Range = diff(range(age))
24 Range

```

R code Exa 2.5.2 Variance of ages Page 44

```
1 ##Example 2.5.2 Pg.44
2 ##Variance of ages
3 age <- c(48,35,46,44,43,42,39,44,49,49,
4          44,39,38,49,49,53,56,57,51,61,
5          53,66,71,75,72,65,67,38,37,46,
6          44,44,48,49,30,45,47,45,48,47,
7          47,44,48,43,45,40,48,49,38,44,
8          43,47,46,57,52,54,56,53,64,53,
9          58,54,59,56,62,50,64,53,61,53,
10         62,57,52,54,61,59,57,52,54,53,
11         62,52,62,57,59,59,56,57,53,59,
12         61,55,61,56,52,54,51,50,50,55,
13         63,50,59,54,60,50,56,68,66,71,
14         82,68,78,66,70,66,78,69,71,69,
15         78,66,68,71,69,77,76,71,43,47,
16         48,37,40,42,38,49,43,46,34,46,
17         46,48,47,43,52,53,61,60,53,53,
18         50,53,54,61,61,61,64,53,53,54,
19         61,60,51,50,53,64,64,53,60,54,
20         55,58,62,62,54,53,61,54,51,62,
21         57,50,64,63,65,71,71,73,66)
22 set.seed(12)
23 srs = sample(age,10)
24 srs
25 var(srs)
26
27 ##Answers differ because of a different random
   sample
```

R code Exa 2.5.3 Coefficient of correlation Page 46

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

```

3
4 mean1 = 145    #sample of 11 years old
5 sd1 = 10
6 mean2 = 80     #sample of 25 years old
7 sd2 = 10
8
9 cv<-function(mean,sd) #user defined function for
    coefficient of variation
10 {cv = sd*100/mean
11 print(cv)}
12
13 cv1 = cv(mean1,sd1)
14 cv2 = cv(mean2,sd2)
15
16 ##variation is much higher in 11 year old than in 25
    year old

```

R code Exa 2.5.4 Kurtosis Page 49

```
1 ##Example 2.5.4 Pg.49
2 ##Kurtosis
3
4 ##install.packages("moments",dependencies = T)
5 library(moments)
6
7 meso = c
8       (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)
9
10 lepto = c
11         (1,2,2,3,3,3,4,4,4,4,4,5,5,5,5,5,5,5,5,5,5,5,5,5,6,6,6,6,6,6,7,7,7,8,8,8,9,9,9,9,9)
12
13 platy = c
14         (1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,5,5,5,5,5,5,6,6,6,7,7,7,8,8,8,9,9,9)
15
16 par(mfrow=c(1,3))
```



```

11 hist(meso,main="Mesokurtic",breaks = 9)
12 hist(lepto,main="Leptokurtic",breaks=9)
13 hist(platy,main="Platykurtic",breaks=9)
14 dev.off()
15
16 descriptive <- function(x)
17 {
18   cat("Mean = " ,mean(x), "\n")
19   cat("Median = " ,median(x),"\n")
20   cat("Mode = " ,names(which(table(x)==max(table(x))
    )), "\n")
21   cat("Kurtosis = " ,kurtosis(x))
22 }
23 d1 = descriptive(meso)
24 d2 = descriptive(lepto)    ##Kurtosis>3
25 d3 = descriptive(platy)    ##Kurtosis<3

```

R code Exa 2.5.5 Box and whisker plot Page 50

```

1 ## Example 2.5.5 , Pg.50
2 ##Refer Table 2.5.1
3 ##Box and whisker plot
4
5 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,
6
7     32.3,32.8,33.3,33.6,34.3,36.9,38.3,44.0)
7 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

```
1 ##Example 3.4.1 Pg.69
2 ##Probability of member being <18 years of age
3
4 disorder <- c("negative", "bipolar", "unipolar", "both"
5             )
6 early <- c(28, 19, 41, 53)
7 later <- c(35, 38, 44, 60)
8 freq <- data.frame(disorder, early, later)
9 Prob_E = sum(early)/(sum(early)+sum(later)) #no. of
10        early subjects/total no. of subjects
11 Prob_E
```

R code Exa 3.4.2 Conditional Probability Page 70

```

1 ##Example 3.4.2 Pg.70
2 ##Conditional Probability
3 ##A = event that family has no history of mood
  disorders
4 ##E = event that subject is <18 years
5
6 disorder <- c("negative","bipolar","unipolar","both"
  )
7 early <- c(28,19,41,53)
8 later <- c(35,38,44,60)
9 freq <- data.frame(disorder,early,later)
10 freq
11
12 condi_AE = freq[1,2]/sum(early) #frequency of
  negative-young subjects/total young subjects
13 condi_AE

```

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

```

1 ##Exammple 3.4.3 Pg.71
2 ##Joint Probability of early subjects and no history
  of mood disorders
3
4 disorder <- c("negative","bipolar","unipolar","both"
  )
5 early <- c(28,19,41,53)
6 later <- c(35,38,44,60)
7 freq <- data.frame(disorder,early,later)
8 freq
9 prob_AE = freq[1,2]/(sum(early)+sum(later)) #
  frequency of negative-young subjects/total
  subjects
10 prob_AE

```

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

```
1 ##Example 3.4.4 Pg.71
2 ##Multiplication rule of early subjects and no
  history of mood disorders
3
4 disorder <- c("negative", "bipolar", "unipolar", "both"
  )
5 early <- c(28,19,41,53)
6 later <- c(35,38,44,60)
7 freq <- data.frame(disorder,early,later)
8 freq
9 Prob_E = sum(early)/(sum(early)+sum(later)) #no.of
  early subjects/total no. of subjects
10 condi_AE = freq[1,2]/sum(early) #frequency of
  negative-young subjects/total young subjects
11
12 Prob_AE = Prob_E*condi_AE #using multiplication
  rule
13 Prob_AE
```

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

```
1 ##Example 3.4.5 Pg.72
2 ##Conditional Probability of early subjects and no
  history of mood disorders
3
4 disorder <- c("negative", "bipolar", "unipolar", "both"
  )
5 early <- c(28,19,41,53)
```

```

6 later <- c(35,38,44,60)
7 freq <- data.frame(disorder,early,late)
8 freq
9 Prob_E = sum(early)/(sum(early)+sum(late)) #no. of
    early subjects/total no. of subjects
10 prob_AE = freq[1,2]/(sum(early)+sum(late)) #
    frequency of negative-young subjects/total
    subjects
11
12 Condi_AE = prob_AE/Prob_E #using multiplication
    rule
13 Condi_AE

```

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

```

1 ##Example 3.4.6 Pg.73
2 ##Probability of early subjects OR no history of
    mood disorders
3
4 disorder <- c("negative","bipolar","unipolar","both"
    )
5 early <- c(28,19,41,53)
6 later <- c(35,38,44,60)
7 freq <- data.frame(disorder,early,late)
8 freq
9
10 Prob_E = sum(early)/(sum(early)+sum(late)) #no. of
    early subjects/total no. of subjects
11 Prob_A = (freq[1,2]+freq[1,3])/(sum(early)+sum(late)
    )) #no. of negative subjects/total no. of
    subjects
12 prob_AE = freq[1,2]/(sum(early)+sum(late)) #
    frequency of negative-young subjects/total
    subjects

```

```

13
14 Prob_AUE = Prob_A + Prob_E - prob_AE #By addition
    rule
15 Prob_AUE

```

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

```

1 ##Exammple 3.4.7 Pg.74
2 ##Conditional Probability of student wears
    eyeglassses given he is a boy
3
4 girls = 60
5 boys = 40
6 girls_glasses = 24
7 boys_glasses = 16
8 prob_E = 40/100      #prob that a student wears
    eyeglasses
9 prob_B = 40/100      #prob of a boy
10 prob_EB = 16/100
11
12 condi_EB = prob_EB / prob_B #conditional prob of a
    student wearing glassss given he is a boy
13 condi_EB
14
15 prob_EB = prob_E*prob_B    #joint probability
16 prob_EB

```

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

```

1 ##Example 3.4.8 Pg.75

```

```

2 ##Find probability of admissions that are NOT
   private
3
4 N = 1200 #total no.of admissions
5 A = 750  #no. of private admissions
6 ProbA = A/N
7 ProbA
8 ProbA_bar = 1 - ProbA
9 ProbA_bar

```

R code Exa 3.4.9 Marginal Probability of Early age Page 75

```

1 ####Exammple 3.4.9 Pg.75
2 ##Marginal Probability of Early age
3
4 disorder <- c("negative","bipolar","unipolar","both"
   )
5 early <- c(28,19,41,53)
6 later <- c(35,38,44,60)
7 freq <- data.frame(disorder,early,later)
8 freq
9 prob_EA = freq[1,2]/(sum(early)+sum(later))
10 prob_EA
11 prob_EB = freq[2,2]/(sum(early)+sum(later))
12 prob_EB
13 prob_EC = freq[3,2]/(sum(early)+sum(later))
14 prob_EC
15 prob_ED = freq[4,2]/(sum(early)+sum(later))
16 prob_ED
17
18 prob_E = prob_EA + prob_EB + prob_EC + prob_ED #
   Marginal Probability
19 prob_E

```

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

```
1 ###Example 3.5.1 Pg.81
2 ##Bayes theorem Refer Table for data
3
4 Yes_D = c(436,14)
5 No_Dbar = c(5,495)
6 dt = data.frame(Yes_D,No_Dbar,row.names = c("
      Positive_T","Negative_Tbar")) )
7 dt
8 prob_D = 0.113
9 prob_Dbar = 1 - prob_D
10 condi_TD = dt[1,1]/sum(Yes_D)
11 condi_TDbar = dt[1,2]/sum(No_Dbar)
12 condi_TbarD = dt[2,1]/sum(Yes_D)
13 condi_TbarDbar = dt[2,2]/sum(No_Dbar)
14
15 condi_DT = (condi_TD*prob_D)/(condi_TD*prob_D +
      condi_TDbar*prob_Dbar) #Bayes theorem
16 condi_DT
17 ##Predictive value of positive test result is very
      high
18
19 condi_DbarTbar = (condi_TbarDbar*prob_Dbar)/(condi_
      TbarDbar*prob_Dbar + condi_TbarD*prob_D) #Bayes
      theorem
20 condi_DbarTbar
21 ##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

```
1 ##Example 4.2.1 Pg.93
2 ##Probability distribution from frequency table
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7
8 prob_dist = freq/N
9 prob_dist
10 sum(prob_dist)
11
12 pdf = data.frame(x,freq,prob_dist)
13 pdf
14
15 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

```
1 ##Example 4.2.2 Pg.95
2 ##Probability distribution from frequency table P(X
  =3)
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7
8 prob_dist = freq/N
9 prob_dist
10 sum(prob_dist)
11
12 pdf = data.frame(x,freq,prob_dist)
13 pdf
14
15 prob_3 = pdf$prob_dist[x==3] #gives the prob value
  at x=3 from the data frame
16 prob_3
```

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

```
1 ##Example 4.2.3 Pg.95
2 ##Prob that family used either one or two programs
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7
8 prob_dist = freq/N
9 prob_dist
10 sum(prob_dist)
11
```

```

12 pdf = data.frame(x,freq,prob_dist)
13 pdf
14
15 prob_1 = pdf$prob_dist[x==1]
16 prob_2 = pdf$prob_dist[x==2]
17 prob_1U2 = prob_1 + prob_2 #additive rule of
    mutually exclusive events
18 prob_1U2

```

R code Exa 4.2.4 Cumulative probability distribution Page 97

```

1 ##Example 4.2.4 Pg.97
2 ##Cumulative probability distribution and p(X<=2)
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7 prob_dist = freq/N
8 cum_dist = cumsum(prob_dist)
9 cdf = data.frame(x,freq,prob_dist,cum_dist)
10 cdf
11
12 cdf_2 = cdf$cum_dist[x==2]
13 cdf_2

```

R code Exa 4.2.5 Cumulative probability distribution Page 97

```

1 ##Example 4.2.5 Pg.97
2 ##Cumulative probability distribution and p(X<4)=P(X
    <=3)
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)

```

```

6 N = sum(freq)
7 prob_dist = freq/N
8 cum_dist = cumsum(prob_dist)
9 cdf = data.frame(x,freq,prob_dist,cum_dist)
10 cdf
11
12 cdf_3 = cdf$cum_dist[x==3]
13 cdf_3

```

R code Exa 4.2.6 Cumulative probability distribution Page 97

```

1 ##Example 4.2.6 Pg.97
2 ##Cumulative probability distribution and  $p(X \geq 5) =$ 
    $1 - P(X \leq 4)$ 
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7 prob_dist = freq/N
8 cum_dist = cumsum(prob_dist)
9 cdf = data.frame(x,freq,prob_dist,cum_dist)
10 cdf
11
12 ans = 1 - cdf$cum_dist[x==4]
13 ans

```

R code Exa 4.2.7 Cumulative probability distribution Page 97

```

1 ##Example 4.2.7 Pg.97
2 ##Cumulative probability distribution and  $p(3 \leq X \leq 5)$ 
    $= P(X \leq 5) - P(X \leq 2)$ 
3
4 x = 1:8

```

```

5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7 prob_dist = freq/N
8 cum_dist = cumsum(prob_dist)
9 cdf = data.frame(x,freq,prob_dist,cum_dist)
10 cdf
11
12 cdf_5 = cdf$cum_dist[x==5]
13 cdf_2 = cdf$cum_dist[x==2]
14 ans = cdf_5 - cdf_2
15 ans

```

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

```

1 ##Example 4.2.8 Pg.98
2 ##mean and variance of prob distribution
3
4 x = 1:8
5 freq = c(62,47,39,39,58,37,4,11)
6 N = sum(freq)
7 prob_dist = freq/N
8 cum_dist = cumsum(prob_dist)
9 cdf = data.frame(x,freq,prob_dist,cum_dist)
10 cdf
11
12 mean = sum(x*prob_dist)
13 mean
14
15 variance = sum(x^2 * prob_dist) - mean^2
16 variance
17
18 sd = sqrt(variance)
19 sd

```

R code Exa 4.3.1 Binomial distribution Page 99

```
1 ##Example 4.3.1 Pg.99
2 ##Binomial distribution P(X=3)
3
4 binom_3 = dbinom(3,5,0.858) #gives binomial density
      for x=3,n=5,p=0.858
5 binom_3
```

R code Exa 4.3.2 Binomial probability distribution Page 103

```
1 ##Example 4.3.2 Pg.103
2 ##Binomial probability distribution
3
4 n = 10
5 x = 4
6 p = 14/100
7 f = dbinom(x,n,p) #prob of success for a binomial
      distribution
8 f #porb that exactly 4 mothers will be admitted to
      smoking
```

R code Exa 4.3.3 Binomial probabilities Page 103

```
1 ##Example 4.3.3 Pg.103
2 ##Binomial probabilities P(X<=5) , P(X>=6) , P(6<=X
      <=9) , P(2<=X<=4)
3
4 prob_a = pbinom(5,25,0.1) #gives binomial
      cumulative dist for x=5,n=25,p=0.1
```

```

5 prob_a
6
7 prob_b = 1 - pbinom(5,25,0.1)
8 prob_b
9
10 prob_c = pbinom(9,25,0.1) - pbinom(5,25,0.1)
11 prob_c
12
13 prob_d = pbinom(4,25,0.1) - pbinom(1,25,0.1)
14 prob_d

```

R code Exa 4.3.4 Binomial probabilities Page 105

```

1 ##Example 4.3.4 Pg.105
2 ##Binomial probabilities P(X=5) , P(X<=5) , P(X>=8)
3
4 prob_a = dbinom(5,12,0.45) #gives binomial
   density for x=5,n=12,p=0.45
5 prob_a
6
7 prob_b = pbinom(5,12,0.55) #gives binomial
   cumulative dist for x=5,n=12,p=0.55
8 prob_b
9
10 prob_c = 1 - pbinom(7,12,0.55)
11 prob_c

```

R code Exa 4.4.1 Poisson distribution Page 110

```

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

```

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

R code Exa 4.4.2 Poisson distribution Page 110

```
1 ##Example 4.4.2 Pg.110  
2 ##Poisson distribution  $P(X \geq 3) = 1 - P(X \leq 2)$   
3  
4 ans = 1 - ppois(2,12) #gives poisson cumulative  
    distribution for x=2, lambda = 12  
5 ans
```

R code Exa 4.4.3 Poisson distribution Page 110

```
1 ##Example 4.4.3 Pg.110  
2 ##Poisson distribution  $P(X \leq 1)$  at lambda=2  
3  
4 cum = ppois(1,2)  
5 cum
```

R code Exa 4.4.4 Poisson distribution Page 111

```
1 ##Example 4.4.4 Pg.111  
2 ##Poisson distribution  $P(X=3)$  at lambda=2  
3  
4 pois_3 = dpois(3,2)  
5 pois_3
```

R code Exa 4.4.5 Poisson distribution Page 112

```
1 ##Example 4.4.5 Pg.112
2 ##Poisson distribution P(X>5) at lambda=2
3
4 pois_5 = 1 - ppois(5,2)
5 pois_5
```

R code Exa 4.6.1 Standard Normal Distribution Page 119

```
1 ##Example 4.6.1 Pg.119
2 ##Standard Normal Distribution
3 ## Find P(z<2)
4
5 p = pnorm(2,0,1) #gives probability of normal dist
   with mean 0 and variance 1 less than 2
6 p
7
8 #Generates a random normal densities and plots for P
   (z<2)
9 x = seq(-4,4,length=10000)
10 y = dnorm(x,0,1)
11 plot(x,y,type="l",lwd=2,col="red")
12 x = seq(-4,2,length=10000)
13 y = dnorm(x,0,1)
14 polygon(c(-4,x,2),c(0,y,0),col="gray")
```

R code Exa 4.6.2 Standard Normal Distribution Page 120

```

1
2 ##Example 4.6.2 Pg 120
3 ##Standard normal distribution  $P(-2.55 < z < 2.55)$ 
4
5 p = pnorm(2.55,0,1) - pnorm(-2.55,0,1) #gives
      probability of normal dist with mean 0 and
      variance 1
6 p
7
8 #Generates a random normal densities and plots for P
       $(-2.55 < z < 2.55)$ 
9 x = seq(-4,4,length=10000)
10 y = dnorm(x,0,1)
11 plot(x,y,type="l",lwd=2,col="red")
12 x = seq(-2.55,2.55,length=10000)
13 y = dnorm(x,0,1)
14 polygon(c(-2.55,x,2.55),c(0,y,0),col="gray")

```

R code Exa 4.6.3 Standard Normal Distribution Page 121

```

1 ##Example 4.6.3 Pg 121
2 ##Standard normal distribution  $P(-2.74 < z < 1.53)$ 
3
4 p = pnorm(1.53,0,1) - pnorm(-2.74,0,1) #gives
      probability of normal dist with mean 0 and
      variance 1
5 p
6
7 #Generates a random normal densities and plots for P
       $(-2.55 < z < 2.55)$ 
8 x = seq(-4,4,length=10000)
9 y = dnorm(x,0,1)
10 plot(x,y,type="l",lwd=2,col="red")
11 x = seq(-2.74,1.53,length=10000)
12 y = dnorm(x,0,1)

```

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

R code Exa 4.6.4 Standard Normal Distribution Page 121

```
1 ##Example 4.6.4 Pg.121
2 ##Standard Normal Distribution
3 ## Find  $P(z \geq 2.71)$ 
4
5 p = 1 - pnorm(2.71,0,1) #gives probability of normal
    dist with mean 0 and variance 1 less than 2
6 p
7
8 #Generates a random normal densities and plots for P
    (z<2)
9 x = seq(-4,4,length=10000)
10 y = dnorm(x,0,1)
11 plot(x,y,type="l",lwd=2,col="red")
12 x = seq(2.71,4,length=10000)
13 y = dnorm(x,0,1)
14 polygon(c(2.71,x,4),c(0,y,0),col="gray")
```

R code Exa 4.6.5 Standard normal distribution Page 122

```
1 ##Example 4.6.5 Pg 122
2 ##Standard normal distribution  $P(0.84 < z < 2.45)$ 
3
4 p = pnorm(2.45,0,1) - pnorm(0.84,0,1) #gives
    probability of normal dist with mean 0 and
    variance 1
5 p
6
7 #Generates a random normal densities and plots for P
    (0.84<z<2.45)
```

```

8 x = seq(-4,4,length=10000)
9 y = dnorm(x,0,1)
10 plot(x,y,type="l",lwd=2,col="red")
11 x = seq(0.84,2.45,length=10000)
12 y = dnorm(x,0,1)
13 polygon(c(0.84,x,2.45),c(0,y,0),col="gray")

```

R code Exa 4.7.1 Normal distribution Page 123

```

1 ##Example 4.7.1 Pg.123
2 ##Normal distribution P(x<3)
3
4 mean = 5.4; sd=1.3
5 p = pnorm(3,mean,sd) #gives distribution of normal
   with mean 5.4 and sd 1.3
6 p
7
8 #Generates a random normal densities and plots for P
   (z<2)
9 x = seq(0,10,length=10000)
10 y = dnorm(x,mean,sd)
11 plot(x,y,type="l",lwd=2,col="red")
12 x = seq(0,3,length=10000)
13 y = dnorm(x,mean,sd)
14 polygon(c(0,x,3),c(0,y,0),col="gray")

```

R code Exa 4.7.2 Normal distribution Page 125

```

1 ##Example 4.7.2 Pg 125
2 ##Normal distribution P(292<X<649)
3
4 mean = 491; sd=119
5 p = pnorm(649,mean,sd) - pnorm(292,mean,sd)

```

```

6 p
7
8 #Generates a random normal densities and plots for P
  (0.84<z<2.45)
9 x = seq(0,1000,length=10000)
10 y = dnorm(x,mean,sd)
11 plot(x,y,type="l",lwd=2,col="red")
12 x = seq(292,649,length=10000)
13 y = dnorm(x,mean,sd)
14 polygon(c(292,x,649),c(0,y,0),col="gray")

```

R code Exa 4.7.3 Normal distribution Page 126

```

1 ##Example 4.7.3 Pg 126
2 ##Normal distribution P(X>8.5)
3
4 mean = 5.4 ; sd = 1.3
5 p = 1 - pnorm(8.5,mean,sd)
6 p

```

Chapter 5

Probabilistic features of the distributions of certain sample statistics

R code Exa 5.3.2 Sampling distribution of mean Page 142

```
1 ##Example 5.3.2 Pg.142
2 ##Sampling distribution of mean
3
4 xbar = 190 ; mu = 185.6 ; sd = 12.7 ; n=10
5
6 z = (xbar-mu)/(sd/sqrt(n))
7 z
8
9 #Generates a random normal densities and plots for P
  (0.84<z<2.45)
10 x = seq(140,225,length=10000)
11 y = dnorm(x,mu,sd)
12 plot(x,y,type="l",lwd=2,col="red")
13 x = seq(190,250,length=10000)
14 y = dnorm(x,mu,sd)
15 polygon(c(190,x,250),c(0,y,0),col="gray")
```

R code Exa 5.3.3 Sampling distribution of mean Page 143

```
1 ##Example 5.3.3 Pg.143
2 ##Sampling distribution of mean  $P(115 < \bar{x} < 125) = P(z_1 < z < z_2)$ 
3
4 xbar1 = 115 ; xbar2 = 125; mu = 120 ; sd = 15 ; n
  =50
5
6 z1 = (xbar1-mu)/(sd/sqrt(n))
7 z2 = (xbar2-mu)/(sd/sqrt(n))
8
9 z = pnorm(z2,0,1)-pnorm(z1,0,1)
10 z
```

R code Exa 5.4.1 Sampling distribution of two means Page 145

```
1 ##Example 5.4.1 Pg.145
2 ##Sampling distribution of two means
3
4 xbar1 = 92 ; xbar2 = 105; mu1=0 ; mu2 = 0 ; sd1 = 20
  ; sd2 = 20 ; n1=15 ; n2=15
5
6 z = ((xbar1-xbar2)-(mu1-mu2))/(sqrt((sd1^2/n1)+(sd2
  ^2/n2)))
7 z
8
9 prob_z = pnorm(z,0,1)
10 prob_z
```

R code Exa 5.4.2 Sampling distribution of two means Page 148

```
1 ##Example 5.4.2 Pg.148
2 ##Sampling distribution of two means
3
4 xbar = 20 ; mu1=45 ; mu2 = 30 ; sd1 = 15 ; sd2 = 20
   ; n1=35 ; n2=40
5
6 z = (xbar-(mu1-mu2))/(sqrt((sd1^2/n1)+(sd2^2/n2)))
7 z = round(z,2)
8 z
9
10 prob_z = 1 - pnorm(z,0,1)
11 prob_z
```

R code Exa 5.5.2 Sampling distribution of Proportion Page 152

```
1 ##Example 5.5.2 Pg.152
2 ##Sampling distribution of Proportion
3
4 P = 0.45 ; p = 0.51 ; q = 1-p ; n=200
5
6 z = (P - p)/sqrt(p*q/n)
7 z = round(z,2)
8 z
9
10 prob_z = pnorm(z,0,1)
11 prob_z
```

R code Exa 5.6.1 Sampling distribution of Two Proportions Page 155

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



```

3
4 P = 0.10 ; p1 = 0.28 ; p2 = 0.21 ; n=100 ; q1 = 1-p1
   ; q2 = 1-p2
5
6 z = (P - (p1-p2))/sqrt((p1*q1/n)+(p2*q2/n))
7 z = round(z,2)
8 z
9
10 prob_z = 1 - pnorm(z,0,1)
11 prob_z

```

R code Exa 5.6.2 Sampling distribution of Two Proportions Page 155

```

1 ##Example 5.6.2 Pg.155
2 ##Sampling distribution of Two Proportions
3
4 P = 0.05 ; p1 = 0.34 ; p2 = 0.26 ; n1=250 ;n2=200;
   q1 = 1-p1 ; q2 = 1-p2
5
6 z = (P - (p1-p2))/sqrt((p1*q1/n1)+(p2*q2/n2))
7 z = round(z,2)
8 z
9
10 prob_z = pnorm(z,0,1)
11 prob_z

```

Chapter 6

Using sample data to make estimates about population parameters

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

```
1 ##Exaxmple 6.2.1 Pg.166
2 ##Confidence interval for population mean mu
3
4 xbar = 22 ; variance = 45; n = 10 ; alpha = 0.05
5 p = qnorm(1-alpha/2,0,1) #gives alpha level p value
6 p = round(p,1)
7 p
8
9 conf_l = xbar - p*sqrt(variance/n)
10 conf_u = xbar + p*sqrt(variance/n)
11
12 conf = c(conf_l,conf_u)
13 conf
```

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

```
1 ##Exaxmple 6.2.2 Pg.168
2 ##Confidence interval for population mean mu
3
4 xbar = 84.3 ; variance = 144; n = 15 ; alpha = 0.01
   ; p =qnorm(1-alpha/2,0,1) #gives alpha level p
   value
5
6 conf_l = xbar - p*sqrt(variance/n)
7 conf_u = xbar + p*sqrt(variance/n)
8
9 conf = c(conf_l,conf_u)
10 conf
```

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

```
1 ##Exaxmple 6.2.3 Pg.168
2 ##Confidence interval for population mean mu
3
4 xbar = 17.2 ; variance = 8^2; n = 35 ; alpha = 0.1 ;
   p =qnorm(1-alpha/2,0,1) #gives alpha level p
   value
5
6 conf_l = xbar - p*sqrt(variance/n)
7 conf_u = xbar + p*sqrt(variance/n)
8
9 conf = c(conf_l,conf_u)
10 conf
```

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

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

```

2 ##Confidence interval for population mean mu
3
4 x = c
      (.360,1.827,.372,.610,.521,1.189,.537,.898,.319,.603,.614,.374,.4

5 xbar = mean(x) ; variance = 0.36; n = length(x) ;
      alpha = 0.05; p =qnorm(1-alpha/2,0,1) #gives
      alpha level p value
6
7 conf_l = xbar - p*sqrt(variance/n)
8 conf_u = xbar + p*sqrt(variance/n)
9
10 conf = c(conf_l,conf_u)
11 conf                                     #95% confidence interval
12 xbar                                     #Mean
13 sqrt(variance)                          #Std dev
14 sqrt(variance/n)                        #Standard error of mean

```

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

```

1 ##Example 6.3.1 Pg.173
2 ##Confidence interval for mean – T test
3
4 xbar = 250.8 ; s = 130.9 ; n = 19 ; alpha = 0.05; p
      =qt(1-alpha/2,n-1) #gives alpha level p value
5
6 conf_l = xbar - p*sqrt(s^2/n)
7 conf_u = xbar + p*sqrt(s^2/n)
8
9 conf = c(conf_l,conf_u)
10 conf                                     #95% confidence interval

```

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

```

1 ##Example 6.4.1 Pg.177
2 ##Confidence interval for two means – Z test
3
4 xbar1 = 4.5 ; xbar2=3.4; n1=12; n2=15; var1 = 1 ;
   var2=1.5
5 alpha = 0.05; p =qnorm(1-alpha/2,0,1) #gives alpha
   level p value
6
7 conf_l = (xbar1-xbar2) - p*sqrt((var1/n1)+(var2/n2))
8 conf_u = (xbar1-xbar2) + p*sqrt((var1/n1)+(var2/n2))
9
10 conf = c(conf_l,conf_u)
11 conf                                     #95% confidence interval

```

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

```

1 ##Example 6.4.2 Pg.178
2 ##Confidence interval for two means – Z test
3
4 xbar1 = 4.3 ; xbar2=13; n1=328; n2=64; s1 = 5.22 ;
   s2=8.97
5 alpha = 0.01; p =qnorm(1-alpha/2,0,1) #gives alpha
   level p value
6
7 conf_l = (xbar1-xbar2) - p*sqrt((s1^2/n1)+(s2^2/n2))
8 conf_u = (xbar1-xbar2) + p*sqrt((s1^2/n1)+(s2^2/n2))
9
10 conf = c(conf_l,conf_u)
11 conf                                     #95% confidence interval

```

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

```

1 ##Example 6.4.3 Pg.180

```

```

2 ##Confidence interval for two means – T test
3
4 xbar1 = 4.7 ; xbar2= 8.8; n1=18; n2=10; var1 = 9.3^2
  ; var2=11.5^2
5 alpha = 0.05; p =qt(1-alpha/2,n1+n2-2) #gives alpha
  level p value
6 s_pooled = ((n1-1)*var1+(n2-1)*var2)/(n1+n2-2)
7
8 conf_l = (xbar1-xbar2) - p*sqrt((s_pooled/n1)+(s_
  pooled/n2))
9 conf_u = (xbar1-xbar2) + p*sqrt((s_pooled/n1)+(s_
  pooled/n2))
10
11 conf = c(conf_l,conf_u)
12 conf                                     #95% confidence interval

```

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

```

1 ##Example 6.4.4 Pg.181
2 ##Confidence interval for two means – T test
3
4 xbar1 = 4.7 ; xbar2= 8.8; n1=18; n2=10; var1 = 9.3^2
  ; var2=11.5^2 ; alpha = 0.05
5 t1 = qt(1-alpha/2,n1-1); t2 = qt(1-alpha/2,n2-1)
6 t = ((t1*var1/n1)+(t2*var2/n2))/((var1/n1) + (var2/
  n2))
7
8
9 s_pooled = ((n1-1)*var1+(n2-1)*var2)/(n1+n2-2)
10
11 conf_l = (xbar1-xbar2) - t*sqrt((var1/n1)+(var2/n2))
12 conf_u = (xbar1-xbar2) + t*sqrt((var1/n1)+(var2/n2))
13
14 conf = c(conf_l,conf_u)
15 conf                                     #95% confidence interval

```

R code Exa 6.5.1 Confidence interval for population proportion Page 185

```
1 ##Example 6.5.1 Pg.185
2 ##Confidence interval for population proportion
3
4 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
5
6 conf_l = P - p*sqrt(P*Q/n)
7 conf_u = P + p*sqrt(P*Q/n)
8
9 conf = c(conf_l,conf_u)
10 conf
```

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

```
1 ##Exaxmple 6.6.1 Pg.187
2 ##Confidence interval for two population proportions
3
4 P1 = 31/68; P2 = 53/255; Q1=1-P1 ; Q2 = 1-P2; n1 =
    68 ; n2 = 255
5 alpha = 0.01 ; p =qnorm(1-alpha/2,0,1) #gives alpha
    level p value
6
7 conf_l = (P1-P2) - p*sqrt((P1*Q1/n1)+(P2*Q2/n2))
8 conf_u = (P1-P2) + p*sqrt((P1*Q1/n1)+(P2*Q2/n2))
9
10 conf = c(conf_l,conf_u)
11 conf
```

R code Exa 6.7.1 Determine the sample size n Page 190

```
1 ##Example 6.7.1 Pg.190
2 ##Determine the sample size n
3
4 z = 1.96 ; sd = 20 ; d = 5
5
6 n = z^2 * sd^2 / d^2
7 n
```

R code Exa 6.8.1 Determine the sample size n Page 192

```
1 ##Example 6.8.1 Pg.192
2 ##Determine the sample size n
3
4 z = qnorm(1-0.05/2,0,1) ; p=0.35; q=1-p; d=0.05
5 n = z^2 * p*q / d^2
6 n = round(n,0)
7 n
```

R code Exa 6.9.1 Chi square test Page 196

```
1 ##Example 6.9.1 Pg.196
2 ##Chi square test
3
4 x <- c(9.7,12.3,11.2,5.1,24.8,14.8,17.7)
5 conf.level= 0.95
6 df = length(x) - 1
7 chilower = round(qchisq((1 - conf.level)/2, df),3)
```



```

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

```

R code Exa 6.10.1 F test Page 200

```

1 ##Example 6.10.1 Pg.200
2 ##F test
3
4 n1=16 ; n2=4 ; s1 =8.1 ; s2 = 5.9; df1 = n1-1 ; df2=
  n2-1; alpha = 0.05
5
6 Flower = qf(alpha/2, df1,df2)
7 Fupper = qf(1 - (alpha/2), df1,df2)
8
9 conf = c(s1^2/(s2^2*Fupper), s1^2/(s2^2*Flower)) #
  95% conf interval for 2 variances
10 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

```
1 ##Example 7.2.1 Pg.222
2 ##Test for mean when population variances are known
  (two sided)
3
4 mu = 30 ; xbar = 27 ; var = 20 ; n=10
5 z = (xbar-mu)/sqrt(var/n)
6 z
7 z_critical = qnorm(0.05/2,0,1)
8 z_critical
9 pvalue = 2*pnorm(z,0,1)
10 pvalue
11 ##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

```
1 ##Example 7.2.2 Pg.226
2 ##Test for mean when population variances are known
  (one sided)
3
4 mu = 30 ; xbar = 27 ; var = 20 ; n=10
5 z = (xbar-mu)/sqrt(var/n)
6 z
7 z_critical = qnorm(0.05,0,1)
8 z_critical
9 pvalue = pnorm(z,0,1)
10 pvalue
11 ##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

```
1 ##Example 7.2.3 Pg.228
2 ##Test for mean when population variances are
  unknown (two sided)
3
4 days<- c
  (14,9,18,26,12,0,10,4,8,21,28,24,24,2,3,14,9)
5 mu = 15 ; xbar = mean(days) ; var = var(days) ; n=17
  ; df = n-1
6 t = (xbar-mu)/sqrt(var/n)
7 t
8 t_critical = qt(0.05/2,df)
9 t_critical
10 pvalue = 2*pt(t,df)
11 pvalue
12
13 ##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

```
1 ##Example 7.2.4 Pg.231
2 ##Test for mean when population variances are known
  (one sided)
3
4 mu = 140 ; xbar = 146 ; s = 27 ; n=157
5 z = (xbar-mu)/(s/sqrt(n))
6 z
7 z_critical = qnorm(1-0.05,0,1)
8 z_critical
9 pvalue = 1-pnorm(z,0,1)
10 pvalue
11 ##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

```
1 ##Example 7.2.5 Pg.232
2 ##Test for mean when population variances are
  unknown (two sided)
3
4 circ<- c
  (33.38,34.34,33.46,32.15,33.95,34.13,33.99,33.85,34.45,34.10,34.2)

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

```

9  t_critical
10 pvalue = 2*pt(t,df)
11 pvalue
12
13 ##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

```

1  ##Example 7.3.1 Pg.237
2  ##Test for two means when population variances are
   known (two sided)
3
4  mu1mu2 = 0 ; xbar1 =4.5 ;xbar2=3.4; var1 = 1; var2
   =1.5 ; n1=12; n2=15
5  z = ((xbar1-xbar2)-(mu1mu2))/sqrt((var1/n1)+(var2/n2
   ))
6  z
7  z_critical = qnorm(0.05/2,0,1)
8  z_critical
9  pvalue = 1-pnorm(z,0,1)
10 pvalue
11 ##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

```

1  ##Example 7.3.2 Pg.239
2  ##Test for means when population variances are
   unknown (one sided)
3
4  control <-c(131,115,124,131,122,117,88,114,150,169)
5  sci <- c(60,150,130,180,163,130,121,119,130,148)

```

```

6
7 t.test(control,sci)
8 #pvalue>0.05, hence not significant

```

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

```

1 ##Example 7.3.3 Pg.240
2 ##Test for means when population variances are
   unknown (two sided)
3
4 n1 = 15 ; n2 = 30 ; xbar1 = 19.16; xbar2 = 9.53 ; s1
   = 5.29; s2 = 2.69
5 t = (xbar1-xbar2)/(sqrt((s1^2/n1)+(s2^2/n2)))
6 t
7 alpha = 0.05; df = n1+n2-2
8 t_critical = qt(0.05/2,df)
9 t_critical
10
11
12 #T > Tcritical, hence significant
13 #Answer might slightly differ due to approximation

```

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

```

1 ##Example 7.3.4 Pg.242
2 ##Test for two means when population variances are
   known (two sided)
3
4 mu1mu2 = 0 ; xbar1 =59.01 ;xbar2=46.61; var1
   =44.89^2; var2 =34.85^2 ; n1=53; n2=54

```

```

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

```

R code Exa 7.4.1 Paired t test Page 251

```

1 ##Example 7.4.1 Pg.251
2 ##Paired t test (one sided)
3
4 preop <- c(22,63.3,96,9.2,3.1,50,33,69,64,18.8,0,34)
5 postop <-c
    (63.5,91.5,59,37.8,10.1,19.6,41,87.8,86,55,88,40)
6
7 t.test(postop,preop,paired = T,alternative = "
    greater")
8
9 #pvalue<0.05 , hence significant

```

R code Exa 7.5.1 Test for proportions Page 258

```

1 ##Example 7.5.1 Pg.258
2 ##Test for proportions (one sided)
3
4 P = 0.063 ; p =24/301 ; Q=1-P ; alpha = 0.05 ; n=301
5
6 z = (P - p)/(sqrt(P*Q/n))
7 z = round(z,2)

```

```

8 z
9 z_critical = qnorm(alpha,0,1)
10 z_critical
11 pvalue = pnorm(z,0,1)
12 pvalue
13
14 #Since p value >alpha, hence not significant

```

R code Exa 7.6.1 Test for two proportions Page 261

```

1 ##Example 7.6.1 Pg.261
2 ##Test for two proportions (one sided)
3
4 p1 = 24/44 ; p2 =11/29 ; q1=1-p1 ; q2=1-p2 ; alpha =
    0.05 ; n1=44; n2=29
5 z = (p1-p2)/(sqrt((p1*q1/n1)+(p2*q2/n2)))
6 z = round(z,2)
7 z
8 z_critical = qnorm(1-alpha,0,1)
9 z_critical
10 pvalue = 1 - pnorm(z,0,1)
11 pvalue
12
13 #Since p value >alpha, hence not significant

```

R code Exa 7.7.1 Test for single population variance Page 264

```

1 ##Example 7.7.1 Pg.264
2 ##Test for single population variance
3
4 var =600 ; n=16 ; df=n-1; s2 = 670.81 ; alpha = 0.05
5 chisq = s2*(n-1)/var
6 chisq

```



```

7 chi_critical1 = qchisq(alpha/2,df)
8 chi_critical2 = qchisq(alpha/2,df,lower.tail = F)
9 chi_critical = c(chi_critical1,chi_critical2)
10 chi_critical
11 pvalue = pchisq(chisq,df)
12 pvalue
13
14 #Since pvalue>alpha, hence not significant

```

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

```

1 ##Example 7.8.1 Pg.268
2 ##Test for ratio of two population variances
3
4 s1 =30.62 ;s2 = 11.37; n1=6 ;n2=6; df1=n1-1; df2 =
   n2-1 ; alpha = 0.05
5 f = s1^2/s2^2
6 f
7 f_critical = qf(alpha,df1,df2,lower.tail = F)
8 f_critical
9 pvalue = pf(f,df1,df2,lower.tail = F)
10 pvalue
11
12 #Since pvalue<alpha, hence significant

```

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

```

1 ##Example 7.8.2 Pg.270
2 ##Test for ratio of two population variances
3
4 control <-c(131,115,124,131,122,117,88,114,150,169)
5 sci <- c(60,150,130,180,163,130,121,119,130,148)
6

```

```
7 var.test(control,sci)
8
9 #pvalue>0.05, hence not significant
10 #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

```
1 ##Example 8.2.1 Pg.318
2 ##One way ANOVA
3
4 ven = c
      (26.72,28.58,29.71,26.95,10.97,21.97,14.35,32.21,19.19,30.92,10.42,
5      16.47,25.19,37.45,45.08,25.22,22.11,33.01,31.20,26.50,32.77,
6
7 squ = c
      (37.42,56.46,51.91,62.73,4.55,39.17,38.44,40.92,58.93,61.88,49.54,
8
9 rrb = c
      (11.23,29.63,20.42,10.12,39.91,32.66,38.38,36.21,16.39,27.44,17.23,
10
11 nrb = c
      (44.33,76.86,4.45,55.01,58.21,74.72,11.84,139.09,69.01,94.61,48.33,
12
13 selenium = c(ven,squ,rrb,nrb)
```

```

11 type = c(rep(1,length(ven)),rep(2,length(squ)),rep
           (3,length(rrb)),rep(4,length(nrb)))
12 type = factor(type,labels = c("ven","squ","rrb","nrb"
                                "))
13 dt = data.frame(type,selenium)
14 View(dt)
15
16 anova <- aov(selenium~type) #anova model for
    selenium content and meat type
17 anova
18 summary(anova)
19
20 #pvalue<0.05, hence significant
21 #Answers might slightly differ due to approximation

```

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

```

1 ##Example 8.2.2 Pg.325
2 ##One way ANOVA and Tukeys HSD
3
4 ven = c
    (26.72,28.58,29.71,26.95,10.97,21.97,14.35,32.21,19.19,30.92,10.42,
5
    16.47,25.19,37.45,45.08,25.22,22.11,33.01,31.20,26.50,32.77,
6
6 squ = c
    (37.42,56.46,51.91,62.73,4.55,39.17,38.44,40.92,58.93,61.88,49.54,
7
7 rrb = c
    (11.23,29.63,20.42,10.12,39.91,32.66,38.38,36.21,16.39,27.44,17.23,
8
8 nrb = c
    (44.33,76.86,4.45,55.01,58.21,74.72,11.84,139.09,69.01,94.61,48.30,
9

```

```

10 selenium = c(ven,squ,rrb,nrb)
11 type = c(rep(1,length(ven)),rep(2,length(squ)),rep
           (3,length(rrb)),rep(4,length(nrb)))
12 type = factor(type,labels = c("ven","squ","rrb","nrb"
                                "))
13 dt = data.frame(type,selenium)
14 View(dt)
15
16 anova <- aov(selenium~type) #anova model for
    selenium content and meat type
17 anova
18 summary(anova)
19
20 #pvalue<0.05, hence significant
21
22 posthoc <- TukeyHSD(anova, "type", conf.level=0.95)
23 posthoc
24
25 #Reject the null if pvalue<alpha
26 #Answers might slightly differ due to approximation

```

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

```

1 ##Example 8.3.1 Pg.339
2 ##Ranadomized Block Design – Two way ANOVA
3
4 days <- c(7,8,9,10,11,9,9,9,9,12,10,10,12,12,14)
5 age <- rep(c(1,2,3,4,5),3)
6 age <- factor(age,labels = c("under 20","20–29","
    30–39","40–49","50 and over"))
7 teach <- c(rep(1,5),rep(2,5),rep(3,5))
8 teach <- factor(teach,labels=c("A","B","C"))
9 dt = data.frame(days,teach,age)
10 dt
11

```

```
12 anova <- aov(days~teach+age)
13 anova
14 summary(anova)
15
16 #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

```
1 ##Example 9.3.1 Pg.417
2 ##Linear regression model
3
4 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
5 y <- c
   (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32
6
7 plot(x,y,main="scatter plot of x and y",xlab="waist
   circumference",ylab="deep abdominal area")
8
9 reg = lm(y~x) #Create a linear model
10 summary(reg)
11 resid(reg) #List of residuals
12 plot(density(resid(reg))) #A density plot
```

```

13 qqnorm(resid(reg)) # A quantile normal plot – good
    for checking normality
14 qqline(resid(reg))
15
16 #Answers may differ due to approximations

```

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

```

1 ##Example 9.4.1 Pg.432
2 ##Linear regression model – Coefficient of
    determination
3
4 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
5 y <- c
    (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32
6
7 reg = lm(y~x) #Create a linear model
8 summary(reg)
9
10 #Multiple R squared is 0.4556
11 #Answers may differ due to approximations

```

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

```

1 ##Example 9.4.2 Pg.436
2 ##Linear regression model – Test for slope
3

```



```

4 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
5 y <- c
   (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32
6
7 reg = lm(y~x) #Create a linear model
8 summary(reg)
9
10 #pvalue of x is less than 0.05, hence significant
11 #Answers may differ due to approximations

```

R code Exa 9.4.3 Linear regression model Residual plot Page 440

```

1 ##Example 9.4.3 Pg.440
2 ##Linear regression model – Residual plot
3
4 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
5 y <- c
   (25.72,25.89,42.60,42.80,29.84,21.68,29.08,32.98,11.44,32.22,28.32
6
7 reg = lm(y~x) #Create a linear model
8 summary(reg)
9 res = resid(reg)
10 plot(x,res,ylab="residuals",xlab="abdominal area")
11 abline(0,0)
12
13 #plot may not be exact replicate due to a different
   scale

```

R code Exa 9.7.1 Correlation and linear regression model Page 447

```
1 ##Example 9.7.1 Pg.447
2 ##Correlation and linear regression model
3
4 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,
5      187,187,187,187,188,188,189,189,190,190,190,190,191,191,19
6
7 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,
7
8 corr = cor(height,cv)
9 corr
10 reg <- lm(cv~height)
11 reg
12 summary(reg)
13 plot(height,cv)
14 abline(reg)
```

R code Exa 9.7.2 Correlation test Page 452

```
1 ##Example 9.7.2 Pg.452
2 ##Correlation test
3
4 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,
```

```

5           187,187,187,187,188,188,189,189,190,190,190,190,191,191,19
6  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,
7
8  corr = cor(height,cv)
9  corr
10 reg <- lm(cv~height)
11 cor.test(height,cv)    #performs correlation test
12
13 #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

```
1
2 ##Example 10.3.1 Pg.493
3 ##Multiple regression equation
4
5 age<- c
  (72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
6
7 edlevel <- c
  (20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
8
9 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
10
11 dt = data.frame(age,edlevel,cda)
12
13 pairs(dt) #multiple scatter plots
```

```

12 reg <- lm(cda~age+edlevel) #multiple regression
    model
13 reg
14 summary(reg)
15
16 #Answers might slightly differ due to approximation

```

R code Exa 10.4.1 Coefficient of multiple determination Page 502

```

1 ##Example 10.4.1 Pg.502
2 ##Coefficient of mutliple determination
3
4 age<- c
    (72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
5
5 edlevel <- c
    (20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
6
6 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
7
7 dt = data.frame(age,edlevel,cda)
8
9 reg <- lm(cda~age+edlevel) #multiple regression
    model
10 reg
11 summary(reg)
12
13 #Multiple R squared value is 0.01807
14 #Answers might slightly differ due to approximation

```

R code Exa 10.4.2 Test for parameters Page 504

```

1 ##Example 10.4.2 Pg.504
2 ##Test for parameters
3
4 age<- c
      (72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
5 edlevel <- c
      (20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
6 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
7 dt = data.frame(age,edlevel,cda)
8
9 reg <- lm(cda~age+edlevel) #multiple regression
      model
10 summary(reg)
11 summary(aov(reg))
12
13 #pvalue > 0.05 , hence there is a significant
      relationship between three variables
14 #Answers might slightly differ due to approximation

```

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

```

1 ##Example 10.4.3 Pg.505
2 ##Test for parameter of variable Age
3
4 age<- c
      (72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
5 edlevel <- c
      (20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
6 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

7 dt = data.frame(age, edlevel, cda)
8
9 reg <- lm(cda~age+edlevel) #multiple regression
  model
10 reg
11 summary(reg)
12
13 #pvalue >0.05, hence significant relationship
  between cda and age
14 #Answers might slightly differ due to approximation

```

R code Exa 10.5.1 confidence interval for CDA Page 508

```

1 ##Example 10.5.1 Pg.508
2 ##95% confidence interval for CDA
3
4 age<- c
  (72,68,65,85,84,90,79,74,69,87,84,79,71,76,73,86,69,66,79,87,71,8
5
5 edlevel <- c
  (20,12,13,14,13,15,12,10,12,15,12,12,12,14,14,12,17,11,12,12,14,1
6
6 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
7
7 dt = data.frame(age, edlevel, cda)
8
9 reg <- lm(cda~age+edlevel) #multiple regression
  model
10 reg
11 summary(reg)
12
13 new.dat <- data.frame(age=68, edlevel=12) #new

```

```

      observation
14 predict(reg, newdata = new.dat, interval = '
      confidence') #confidence interval
15 predict(reg, newdata = new.dat, interval = '
      prediction') #prediction interval
16
17 #Answers might slightly differ due to approximation

```

R code Exa 10.6.1 Multiple correlation coefficient Page 511

```

1 ##Example 10.6.1 Pg.511
2 ##Multiple correlation coefficient
3
4 w<- c
      (193.6,137.5,145.4,117,105.4,99.9,74,74.4,112.8,125.4,126.5,115.9
5
6 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
7
8 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.
9
10 reg = lm(w~p+s)
11 reg
12 summary(reg)
13
14 #Multiple R squared = 0.2942
15 #Answers might slightly differ due to approximation

```

R code Exa 10.6.2 Partial correlation coefficient Page 515

```

1 ##Example 10.6.2 Pg.515

```



```

2  ##Partial correlation coefficient
3
4  w<- c
      (193.6,137.5,145.4,117,105.4,99.9,74,74.4,112.8,125.4,126.5,115.9
5
6  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
7
8  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.
9
10 reg = lm(w~p+s)
11 reg
12 summary(reg)
13
14 res1 = residuals(lm(w~p))
15 res2 = residuals(lm(s~p))
16 res3 = residuals(lm(w~s))
17 res4 = residuals(lm(p~s))
18 res5 = residuals(lm(p~w))
19 res6 = residuals(lm(s~w))
20
21 # use Spearman correlation coefficient to calculate
22   the all possible partial correlations
23
24 p1 = cor(res1,res2,method = "spearman")
25 p2 = cor(res1,res3,method = "spearman")
26 p3 = cor(res1,res4,method = "spearman")
27 p4 = cor(res1,res5,method = "spearman")
28 p5 = cor(res1,res6,method = "spearman")
29 p6 = cor(res2,res3,method = "spearman")
30 p7 = cor(res2,res4,method = "spearman")
31 p8 = cor(res2,res5,method = "spearman")
32 p9 = cor(res2,res6,method = "spearman")
33 p10 = cor(res3,res4,method = "spearman")
34 p11 = cor(res3,res5,method = "spearman")
35 p12 = cor(res3,res6,method = "spearman")
36 p13 = cor(res4,res5,method = "spearman")

```

```
33 p14 = cor(res4,res6,method = "spearman")
34 p15 = cor(res5,res6,method = "spearman")
35 p <- c(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12,p13,
        p14,p15)
36 p
37
38
39 #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

```
1 ##Example 11.1.1 Pg.540
2 ##Box and whiskers plot
3
4 THC <- c
      (.30,2.75,2.27,2.37,1.12,.60,.61,.89,.33,.85,2.18,3.59,.28,1.90,1
5 log_THC <- log10(THC)
6 log_THC
7 data = data.frame(THC,log_THC)
8
9 #install.packages("car")
10 library(car)
11 Boxplot(data,ylab="concentration")
```

R code Exa 11.1.2 Correlation for 3 variables Page 542

```
1 ##Example 11.1.2 Pg.542
2 ##Correlation for 3 variables
3
4 sbp <- c
   (126,129,126,123,124,125,127,125,123,119,127,126,122,126,125)

5 weight <- c
   (125,130,132,200,321,100,138,138,149,180,184,251,197,107,125)

6 bmi <- c
   (24.41,23.77,20.07,27.12,39.07,20.90,22.96,24.44,23.33,25.82,26.4

7 dt<-data.frame(sbp,weight,bmi)
8
9 #install.packages("PerformanceAnalytics")
10 library(PerformanceAnalytics)
11
12 chart.Correlation(dt)
13 ##Shows correlation coefficient and significant
   values
14
15 cor.test(sbp,weight)
16 #result shows cor = -0.289, p value = 0.296
17 cor.test(sbp,bmi)
18 #result shows cor = -0.213, p value = 0.447
19 cor.test(bmi,weight)
20 #result shows cor = 0.962, p value = 0.000
```

R code Exa 11.2.1 Regression model for categorical data Page 545

```
1 ##Example 11.2.1 Pg.545
2 ##Regression model for categorical data
3
```

```

4 grams <- c(3147,2977,3119,3487,4111,3572,3487,
5           3147,3345,2665,1559,3799,
6           2750,3487,3317,3544,3459,2807,3856,
7           3260,2183,3204,3005,3090,3430,3119,
8           3912,3572,3884,3090,2977,3799,4054,
9           3430,3459,3827,3147,3289,3629,3657,
10          3175,3232,3175,3657,3600,3572,709,624,
11          2778,3572,3232,3317,2863,3175,3317,3714,
12          2240,3345,3119,2920,3430,3232,3430,4139,
13          3714,1446,3147,2580,3374,3941,2070,3345,
14          3600,3232,3657,3487,2948,2722,3771,3799,
15          1871,3260,3969,3771,3600,2693,3062,2693,3033,3856,

16          4111,3799,3147,2920,4054,2296,3402,1871,
17          4167,3402)
18 weeks <- c
    (40,41,38,38,39,41,40,41,38,34,34,38,38,40,38,
19     43,45,37,40,40,42,38,36,40,39,40,39,40,41,38,

20     42,37,40,38,41,39,44,38,36,36,41,43,36,40,39,

21     40,25,25,36,35,38,40,37,37,40,34,36,39,39,37,

22     41,35,38,39,39,28,39,31,37,40,37,40,40,41,38,

23     39,38,40,40,45,33,39,38,40,40,35,45,36,41,42,

24     40,39,38,36,40,36,38,33,41,37)
25 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,
26     1,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0,1,1,0,1,0,0,0,

27     0,0,0,rep(0,15),1,1,0,0,0,0,0,0,0,1,rep
        (0,10),
28     1,rep(0,11),1,0,1)
29 smoke = factor(smoke, labels = c("nonsmoker","smoker
    "),levels=c(0,1))
30

```

```

31 plot(weeks, grams, pch=21,
32       bg=c("red", "green3")[unclass(smoke)])
33 ##red for non smokers and green for smokers
34
35 reg = lm(grams~weeks+smoke)
36 summary(reg)
37 ##Gives the estimates and corresponding p values
38 summary(aov(reg))
39 ##Gives the Anova results (sum of squares and F
    statistic)
40
41
42 plot(weeks, grams, pch=21,
43       bg=c("red", "green3")[unclass(smoke)])
44 abline(reg)

```

R code Exa 11.2.2 Test for model parameter for categorical data

```

1 ##Example 11.2.2 Pg.549
2 ##Test for model parameter for categorical data
3
4 grams <- c(3147,2977,3119,3487,4111,3572,3487,
5            3147,3345,2665,1559,3799,
6            2750,3487,3317,3544,3459,2807,3856,
7            3260,2183,3204,3005,3090,3430,3119,
8            3912,3572,3884,3090,2977,3799,4054,
9            3430,3459,3827,3147,3289,3629,3657,
10           3175,3232,3175,3657,3600,3572,709,624,
11           2778,3572,3232,3317,2863,3175,3317,3714,
12           2240,3345,3119,2920,3430,3232,3430,4139,
13           3714,1446,3147,2580,3374,3941,2070,3345,
14           3600,3232,3657,3487,2948,2722,3771,3799,
15           1871,3260,3969,3771,3600,2693,3062,2693,3033,3856,
16
17           4111,3799,3147,2920,4054,2296,3402,1871,

```

```

17         4167,3402)
18 weeks <- c
    (40,41,38,38,39,41,40,41,38,34,34,38,38,40,38,
19         43,45,37,40,40,42,38,36,40,39,40,39,40,41,38,
20         42,37,40,38,41,39,44,38,36,36,41,43,36,40,39,
21         40,25,25,36,35,38,40,37,37,40,34,36,39,39,37,
22         41,35,38,39,39,28,39,31,37,40,37,40,40,41,38,
23         39,38,40,40,45,33,39,38,40,40,35,45,36,41,42,
24         40,39,38,36,40,36,38,33,41,37)
25 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,
26         1,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,0,
27         0,0,0,rep(0,15),1,1,0,0,0,0,0,0,0,0,1,rep
            (0,10),
28         1,rep(0,11),1,0,1)
29 smoke = factor(smoke, labels = c("nonsmoker","smoker
    "),levels=c(0,1))
30
31 reg = lm(grams~weeks+smoke)
32 summary(reg)
33
34 ##We get the t test statistic for smokers as -2.17
35 ##p value is 0.033 < 0.05, hence significant
36 ##smoking mothers associateed with reduced birth
    weights of babies

```

R code Exa 11.3.1 Stepwise regression model Page 561

1 [##Example 11.3.1 Pg.561](#)

```

2 ##Stepwise regression model
3
4 y = c
      (45,65,73,63,83,45,60,73,74,69,66,69,71,70,79,83,75,67,67,52,52,6
5 x1 = c
      (74,65,71,64,79,56,68,76,83,62,54,61,63,84,78,65,86,61,71,59,71,6
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
7 x3 = c
      (40,64,79,57,76,54,66,65,77,57,67,66,67,68,82,82,79,75,67,64,44,7
8 x4 = c
      (66,68,81,59,76,59,71,75,76,67,63,84,60,84,84,65,84,60,80,69,48,7
9 x5 = c
      (93,74,87,85,84,50,69,67,84,81,68,75,64,78,78,55,80,81,86,79,65,8
10 x6 = c
      (47,49,33,37,33,42,37,43,33,43,36,43,35,37,39,38,41,45,48,54,43,4
11
12 step(lm(y~x1+x2+x3+x4+x5+x6),direction = "both") #
      performs stepwise regression
13
14 #x1,x2,x3,x6 variables are selected

```

R code Exa 11.4.1 Logistic regression page 572

```

1 ## Example 11.4.1 Page 572
2 ##Logistic regression
3
4 cases <- c(21,20,92,15)

```



```

5 disease <- c(1,1,0,0)
6 disease = factor(disease, labels = c("present", "
      absent"))
7 sex <- c(1,2,1,2)
8 sex = factor(sex, labels=c("male", "female"))
9 dt = data.frame(disease, sex, cases)
10 dt
11 xtabs(cases~., dt) #creates contingency table
12 fit <- glm(disease~sex, weights = cases, data = dt,
      family = "binomial") #logistic regression
13 summary(fit)
14
15 ##summary gives estimated value for sex and
      intercept
16 ##pvalue < 0.05, hence significant

```

R code Exa 11.4.2 Logistic regression page 573

```

1 ##Example 11.4.2 Pg.573
2 ##Logistic regression
3
4 age<- c(50,59,42,50,34,49,67,44,53,45,79,
5         46,62,58,70,60,67,64,62,50,61,69,
6         74,65,80,69,77,61,72,67,73,75,71,
7         69,78,69,74,86,49,63,63,72,64,72,
8         64,72,79,75,70,73,66,75,73,71,72,
9         69,76,60,79,78,62,73,46,57,53,40,
10        73,68,72,59,64,78,68,67,55,71,80,
11        75,69,80,79,71,69,78,75,71,69,77,
12        81,78,76,84,74,59,81,74,77,59,75,
13        68,81,74,65,81,62,85,84,39,52,67,
14        82,84,79,81,74,85,92,69,83,82,85,
15        82,74,50,55,66,49,55,73,41,64,
16        46,65,50,61,64,59,73,73,65,67,60,
17        69,61,79,66,68,61,63,70,68,59,64,

```

```

18         62,74,61,69,76,71,61,46,69,66,57,
19         60,63,63,56,70,70,63,63,65,67,68,
20         84,69,78,69,79,83,67,47,57,66)
21
22 status <- c(rep(0,122),rep(1,63))
23 status
24 status1 = factor(status,labels = c("nonparticipating
      ","participating"),levels=c(0,1))
25 status1
26
27 fit <- glm(status1~1+age,family="binomial",control=
      glm.control(maxit=50)) #logistic regression
28 summary(fit)
29 ##summary gives the estimates of intercept and age
30 ##Also the p value to test the slope coefficient
31
32 ##A function to estimate probabilities from logistic
      model
33 est_prob <- function(x)
34 {
35   pred = predict(fit,newdata=data.frame(age=x))
36   prob = exp(pred)/(1+exp(pred))
37   print(prob)
38 }
39
40 est_prob(x=50)
41 est_prob(x=age)
42 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

```
1 ##Exaxmple 12.3.2 Pg.609
2 ## Tests of goodness of fit – Binomial distribution
3
4 x <- 0:10    #no. of paients out of 25 preferring new
    pain reliever
5 f <- c(5,6,8,10,10,15,17,10,10,9,0) #no. of doctors
    reporting this number
6 N <- c(0,6,16,30,40,75,102,70,80,81,0) #total number
    of patients preferring new pain reliever by
    doctor
7
8 p = sum(N)/(25*sum(f)) ; p
9
10 prob = dbinom(x,25,p) ;prob
11
12 pooled1 = f[1] + f[2]    #pooling first two values
    since <5
```

```

13 Obs_f = c(pooled1,f[c(-1,-2)]) ;Obs_f
14 Exp_f = sum(f)*prob
15 pooled2 = Exp_f[1] + Exp_f[2] #pooling first two
    values since <5
16 Exp_f = c(pooled2,Exp_f[c(-1,-2)]) ;Exp_f
17
18 dt = data.frame(Obs_f,Exp_f)
19 dt
20 sum(Obs_f)
21 sum(Exp_f)
22
23 chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
24 p_val = pchisq(0.005,length(x)-2) ;p_val
25
26 #Since pval < 0.005, we conclude data came from
    binomial distribution
27 #Answers slightly differ by decimals due to
    aproximations

```

R code Exa 12.3.3 Test of goodness of fit Poisson distribution

```

1 ##Exaxmple 12.3.3 Pg.611
2 ## Tests of goodness of fit – Poisson distribution
3
4 x <- 0:10
5 f <- c(5,14,15,23,16,9,3,3,1,1,0)
6 lambda = sum(x*f)/sum(f) ;lambda #mean=lambda
7 prob = dpois(x,lambda) ;prob
8
9 pooled1 = f[9] + f[10] + f[11] #pooling last three
    values since <5
10 Obs_f = c(f[c(-9,-10,-11)],pooled1) ;Obs_f
11 Exp_f = sum(f)*prob
12 pooled2 = Exp_f[9] + Exp_f[10] + Exp_f[11] #
    pooling first two values since <5

```

```

13 Exp_f = c(Exp_f[c(-9,-10,-11)],pooled2) ;Exp_f
14
15 dt= data.frame(Obs_f,Exp_f)
16 dt
17 sum(Obs_f);sum(Exp_f)
18 chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
19 p_val = pchisq(0.05,length(x)-3) ;p_val
20
21 #Since pval < 0.005, we conclude data came from
    poisson distribution
22 #Answer slightly differ by decimal due to
    approximation

```

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

```

1 ##Exaxmple 12.3.4 Pg.614
2 ## Tests of goodness of fit – Uniform distribution
3
4 x <- c("Dec","Jan","Feb","Mar","Apr")
5 f <- c(62,84,17,16,21)
6 prop <- 1/length(x)
7
8 Obs_f = f ;Obs_f
9 Exp_f = prop*sum(f) ;Exp_f
10
11 chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
12 p_val = pchisq(0.05,length(x)-1) ;p_val
13
14 #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

```

1 ##Exaxmple 12.3.5 Pg.616
2 ##Goodness of fit – Distribution of traits
3
4 n = 200
5 dominant = 43
6 heterozygous = 125
7 recessive = 32
8
9 Obs_f = c(43,125,32)
10 Exp_f = c(50,100,50) #1:2:1 ratio
11
12 chi_sq = sum((Obs_f-Exp_f)^2/Exp_f) ; chi_sq
13 chi_critical = qchisq(0.95,2) ;chi_critical
14 p_val = pchisq(0.05,2) ;p_val
15
16 #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

```

1
2 ##Example 12.4.1 Pg.621
3 ##Test for independence
4
5 x<-matrix(c(260,299,15,41,7,14),nrow=3,byrow=T)
6 x
7 rownames(x)<-c("White","Black","Other")
8 colnames(x)<-c("Yes","No")
9 print(x)
10
11 chisq.test(x)
12
13 #pvalue < 0.05 , hence there is relationship between
    race and folic acid

```

R code Exa 12.4.2 Test for independence Page 626

```
1 ##Example 12.4.2 Pg.626
2 ##Test for independence
3
4 x<-matrix(c(131,52,14,36),nrow=2,byrow=T)
5 x
6 rownames(x)<-c("Fallers","Non fallers")
7 colnames(x)<-c("Yes","No")
8 print(x)
9
10 chisq.test(x)
11
12 #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

```
1 ##Example 12.5.1 Pg.631
2 ##Test for homogeneity
3
4 x<-matrix(c(21,75,19,77),nrow=2,byrow=T)
5 x
6 rownames(x)<-c("Narcoleptic","Controls")
7 colnames(x)<-c("Yes","No")
8 print(x)
9
10 chisq.test(x)
11
12 #pvalue > 0.05 , hence two populations may be
    homogenous wrt migraine frequency
13 #Answer is slightly differing from the textbook
```

R code Exa 12.6.1 Fishers exact test Page 638

```
1 ##Example 12.6.1 Pg.638
2 ##Fishers exact test
3
4 x<-matrix(c(9,2,12,8),nrow=2,byrow=T)
5 x
6 rownames(x)<-c("Naive","Experienced")
7 colnames(x)<-c("Yes","No")
8 print(x)
9
10 fisher.test(x)
11
12 #pvalue > 0.05 , hence rate of remaining on regimen
    for 120 weeks is same for naive and experienced
    groups
13 #Answer is slightly differing from the textbook
```

R code Exa 12.7.1 Relative risk Page 644

```
1 ##Example 12.7.1 Pg.644
2 ##Relative risk
3
4 x<-matrix(c(18,199,22,216),nrow=2,byrow=T)
5 x
6 rownames(x)<-c("Not exercising","Extreme exercising")
7 colnames(x)<-c("cases","non cases")
8 print(x)
9
10 #install.packages("mosaic")
```



```
11 library(mosaic)
12
13 relrisk(x)
```

R code Exa 12.7.2 Odds ratio Page 647

```
1 ##Example 12.7.2 Pg.647
2 ##Odds ratio
3
4 x<-matrix(c(68,3496,64,342),nrow=2,byrow=T)
5 x
6 rownames(x)<-c("Never Smoked","smoked")
7 colnames(x)<-c("cases","non cases")
8 print(x)
9
10 #install.packages("mosaic")
11 library(mosaic)
12
13 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

```
1 ##Example 13.3.1 Pg.673
2 ##Sign test
3
4 score <- c(4,5,8,8,9,6,10,7,6,6)
5 md= median(score)
6 diff<-score-md
7 diff
8 sdiff<-sign(diff)
9 sdiff
10 s=length(sdiff[sdiff==1])
11 s
12 cv=pbinom(1,9,0.5)
13 cv
14 pval = 2*cv ;pval
15
```

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

R code Exa 13.3.2 Sign test for paired data Page 677

```
1 ##Example 13.3.2 Pg.677
2 ##Sign test for paired data
3
4 score_x <- c(1.5,2,3.5,3,3.5,2.5,2,1.5,1.5,2,3,2)
5 score_y <- c(2,2,4,2.5,4,3,3.5,3,2.5,2.5,2.5,2.5)
6 diff<-score_x - score_y
7 diff
8 sdiff<-sign(diff)
9 sdiff
10 s=length(sdiff[sdiff==1])
11 s
12 cv=pbinom(s,11,0.5)
13 cv
14
15 #since pvalue<0.05, instruction was beneficial
```

R code Exa 13.4.1 Wilcoxon signed rank test page 683

```
1 ##Example 13.4.1 Pg.683
2 ##Wilcoxon signed rank test
3
4 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
5 wilcox.test(cardiac_out, alternative= "two.sided",
    conf.int=T)
6
```

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

R code Exa 13.5.1 Median Test page 686

```
1  
2 ##Example 13.5.1 Pg.686  
3 ##Median Test  
4  
5 urban <-c  
   (35,26,27,21,27,38,23,25,25,27,45,46,33,26,46,41)  
6 rural <-c(29,50,43,22,42,47,42,32,50,37,34,31)  
7 z<-c(urban,rural)  
8 n<-length(z) ;n  
9 u<-median(z)  
10 a<-length(urban[urban>u]) ;a  
11 b<-length(rural[rural>u]) ;b  
12 c<-length(urban[urban<=u]) ;c  
13 d<-length(rural[rural<=u]) ;d  
14  
15 chi<-(n*(a*d-b*c)^2)/((a+b)*(c+d)*(a+c)*(b+d)) ;chi  
16  
17 chi_critical <- qchisq(1-0.05,1,lower.tail = T) ;chi  
   _critical  
18 pval<- pchisq(chi,1) ;pval  
19  
20 #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

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

```

3
4 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
5 unexposed <-c
    (17.4,16.2,17.1,17.5,15,16,16.9,15,16.3,16.8)
6
7 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

```

1 ##Example 13.7.1 Pg.699
2 ##Kolmogorov Smirnov Goodness of fit test
3
4 values <- c
    (75,84,80,77,68,87,92,77,92,86,78,76,80,81,72,77,92,80,80,77,77,9
5 ks.test(values, "pnorm", mean(values), sd(values),
    alternative = "two.sided")
6
7 #pvalue > 0.05, hence sample would have come from
    normal distribution

```

R code Exa 13.8.1 Kruskal Wallis one way ANOVA Page 705

```

1 ##Example 13.8.1 Pg.705
2 ##Kruskal Wallis one way ANOVA
3
4 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
5 group<-c(rep(1,5),rep(2,5),rep(3,5))

```

```

6 group
7 group1<-factor(group,labels=c("air","benzaldehyde","
  acetaldehyde"))
8 group1
9 dt<-data.frame(group1,cell_count)
10 dt
11 kruskal.test(cell_count~group1)
12
13 #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

```

1 ##Example 13.8.2 Pg.708
2 ##Kruskal Wallis one way ANOVA
3
4 book_value <-c
  (1735,1520,1476,1688,1702,2667,1575,1602,1530,1698,5260,4455,4480
5 group<-c(rep(1,10),rep(2,8),rep(3,9),rep(4,7),rep
  (5,7))
6 group
7 group1<-factor(group,labels=c("A","B","C","D","E"))
8 group1
9 dt<-data.frame(group1,book_value)
10 dt
11 kruskal.test(book_value~group1)
12
13 #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

```

1 ##Example 13.9.1 Pg.713
2 ##Friedman test
3
4 ranks_A <-c(2,2,2,1,3,1,2,1,1)
5 ranks_B <-c(3,3,3,3,2,2,3,3,3)
6 ranks_C <-c(1,1,1,2,1,3,1,2,2)
7 n = 9 ; k=3
8 friedman = (12/(n*k*(k+1)))*(sum(ranks_A)^2+sum(
    ranks_B)^2+sum(ranks_C)^2) - 3*n*(k+1)
9 friedman
10 pval = pchisq(0.05/2,k-1) ;pval
11
12 #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

```

1 ##Example 13.9.2 Pg.715
2 ##Friedman test
3
4 salivaryflow <-
5   matrix(c
        (29,48,75,100,72,30,100,100,70,100,86,96,54,35,90,99,5,43,32,81
        ,
6         nrow = 16,
7         byrow = TRUE,
8         dimnames = list(1 : 16,
9                           c("A", "B", "C", "D"))))
10 friedman.test(salivaryflow)
11
12 #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

```
1 ##Example 13.10.1 Pg.720
2 ##Test for correlation
3
4 age <- c
      (20,21,22,24,27,30,31,33,35,38,40,42,44,46,48,51,53,55,58,60)
5 eeg <- c
      (98,75,95,100,99,65,64,70,85,74,68,66,71,62,69,54,63,52,67,55)
6
7 cor.test(age,eeg)
8
9 #pvalue < 0.05, hence age and eeg are inversely
  related
```

R code Exa 13.10.2 Test for correlation Page 722

```
1 ##Example 13.10.2 Pg.722
2 ##Test for correlation
3
4 age <- c
      (82,85,83,64,82,53,26,47,37,49,65,40,32,50,62,33,36,53,50,71,54,61)
5 conc <- c
      (169.62,48.94,41.16,63.95,21.09,5.40,6.33,4.26,3.62,4.82,108.22,110.4,
6
7 cor.test(age,conc)
8
9 #pvalue < 0.05, hence age and mineral concentration
   are inversely related
```