

R Textbook Companion for
Probability and Statistics for Engineers
by Richard L. Scheaffer, Madhuri S. Mulekar,
James T. McClave¹

Created by
Shikha Vyas
B.E.
Information Technology
Institute of Engineering and Technology, DAVV, Indore
Cross-Checked by
R TBC Team

May 25, 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: Probability and Statistics for Engineers

Author: Richard L. Scheaffer, Madhuri S. Mulekar, James T. McClave

Publisher: Cengage Learning, USA

Edition: 5

Year: 2011

ISBN: 978-0-534-40302-7

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 Data Collection and Exploring Univariate Distributions	5
2 Exploring Bivariate Distributions and Estimating Relations	13
4 Probability	25
5 Discrete Probability Distributions	37
6 Continuous Probability Distributions	47
7 Multivariate Probability Distributions	61
8 Statistics Sampling Distributions and Control Charts	71
9 Estimate	87
10 Hypothesis Testing	101
11 Inference for Regression Parameters	118
12 Analysis of Variance	137

List of R Codes

Exa 1.1	Auto paint shop relative freq	5
Exa 1.2	Air quality relative frequency	5
Exa 1.3	histograms	6
Exa 1.4	Dotplot	7
Exa 1.8	AQI mean and median	9
Exa 1.9	accidents mean and median	9
Exa 1.10	AQI quantile	9
Exa 1.12	AQI sd and variance	10
Exa 1.13	histogram of cars	10
Exa 1.14	lawnmower mean and sd	11
Exa 1.16	summary of CO and SO2	11
Exa 1.17	Boxplots	12
Exa 2.1	Projections on net new workers	13
Exa 2.3	Timeplot of CO and VOC emission	14
Exa 2.4	Average number of days with AQI greater than 100 . .	14
Exa 2.5	Fuel consumption and efficiency of cars and vans . . .	14
Exa 2.6	Annual temperatures at Newnan	15
Exa 2.7	scatterplot of heat exchangers	16
Exa 2.8	correlation coeff for power plant	17
Exa 2.9	correlation coeff for O3 and SO2	17
Exa 2.12	equation of the least squares regression line	17
Exa 2.13	linear fit OT and WR	18
Exa 2.14	Predict peak power load	19
Exa 2.15	Predict wall reduction	20
Exa 2.16	coeff of determination for the peak power load	20
Exa 2.18	Least squares fit and residual plot for WR OT data . .	20
Exa 2.19	linear regression model for population density	21
Exa 2.20	linear regression modal for length and wt	22

Exa 4.1	prob of eastern and western cities	25
Exa 4.2	die example	25
Exa 4.5	Venn diag EE	26
Exa 4.6	Venn diag labor statistics	26
Exa 4.7	Venn diag electric motors	26
Exa 4.12	product rule example	27
Exa 4.13	Product rule for cities	27
Exa 4.16	Permutation of employees	27
Exa 4.17	Permutation of operations	28
Exa 4.18	Permutation of divisions	28
Exa 4.19.a	selection of employees	28
Exa 4.19.b	Probability of female candidate	28
Exa 4.20	Combination of applicants	29
Exa 4.21.a	Partitioning of employees	29
Exa 4.21.b	Partitioning of employees specifically	29
Exa 4.25	conditional probability of motors	30
Exa 4.28	independent event of job	30
Exa 4.29	Complementary Events	31
Exa 4.31	additive rule	31
Exa 4.32	multiplicative rule for defects	31
Exa 4.33	multiplicative rule for relays	32
Exa 4.35	example of ballpoint pen	32
Exa 4.37	Bayes Rule	33
Exa 4.39	flight accident	33
Exa 4.40	Bays rule for tower malfunction	34
Exa 4.41	Odd ratio	34
Exa 4.42	Odd ratio	35
Exa 5.1	Relay example	37
Exa 5.2	multiplicative rule of probability	37
Exa 5.3	distribution function	38
Exa 5.4	mean and sd	39
Exa 5.5	expected daily demand and variance	39
Exa 5.6	mean and variance of the daily costs	39
Exa 5.7	mean score and the standard deviation	40
Exa 5.8.a	Tchebyshefs Theorem	40
Exa 5.8.b	shortest interval	41
Exa 5.11	binomial distribution of fuses	41
Exa 5.12	binomial distribution of battery	42

Exa 5.13	binomial distribution of chemicals	42
Exa 5.15	binomial distribution of contracts	43
Exa 5.17	geomatric distribution of interviews	44
Exa 5.18	expected time	44
Exa 5.19	geomatric distribution of interviews	45
Exa 5.20	industrial accidents	45
Exa 5.22	The Hypergeometric Distribution	46
Exa 5.23.a	The Hypergeometric Distribution of boxes	46
Exa 6.1	continuous random variable	47
Exa 6.2	probability density function	47
Exa 6.4	distribution function	48
Exa 6.5	Expected value and variance	48
Exa 6.6	expected weekly demand	49
Exa 6.7	Expected interval	50
Exa 6.8.a	probability of uniform distribution	50
Exa 6.9.a	Probability of exponential distribution	50
Exa 6.9.b	cumilative probability	51
Exa 6.10	The Gamma Distribution of components	51
Exa 6.11	The mean and variance for the length of maintenance times	52
Exa 6.12	z value	52
Exa 6.13	Normal distribution	52
Exa 6.14	standard normal distribution of bottles	53
Exa 6.15	Normal distribution	53
Exa 6.16	Normal distribution of maths score	54
Exa 6.17.a	percentile score	54
Exa 6.17.b	probability	55
Exa 6.18	octane rating	56
Exa 6.20	QQ plot	56
Exa 6.22	CO z score	57
Exa 6.23	lognormal distribution	57
Exa 6.24	beta distribution	58
Exa 6.25	Relative frequency histograms and densities	58
Exa 6.26	Weibull Distribution	59
Exa 6.27	Weibull Distribution	60
Exa 7.2	joint probability distribution of X1 and X2	61
Exa 7.3	Joint probability distribution of imprities	62
Exa 7.4	marginal probability density functions	63

Exa 7.5	The marginal probability density functions	63
Exa 7.6	conditional probability	64
Exa 7.7	conditional probability density function	64
Exa 7.10	covariance between two random variables	65
Exa 7.12	mean and variance	67
Exa 7.13	mean and variance of the total weekly amount	68
Exa 7.15	multinomial probability distribution	68
Exa 7.19	conditional expectation	69
Exa 7.20	conditional distribution of X_1 given X_2	69
Exa 8.2	central limit theorem	71
Exa 8.3	central limit theorem for average fracture strength	72
Exa 8.4	Probability of sample mean	73
Exa 8.5	Normal distribution of failure strengths	74
Exa 8.6	The Sampling Distribution of large samples	74
Exa 8.7	The Sampling Distribution of S^2	75
Exa 8.8	The Sampling Distribution of S^2	75
Exa 8.9	General distribution large samples	76
Exa 8.10	Small samples case equal variances	77
Exa 8.11	The sampling distribution of \bar{X}	77
Exa 8.12	sampling distribution of p_1 minus p_2	78
Exa 8.13.a	sampling distribution of S_1^2 divided S_2^2	79
Exa 8.13.b	probability of sampling distribution	79
Exa 8.14	\bar{X} and R charts	80
Exa 8.15	\bar{X} and S charts	81
Exa 8.16	p chart	83
Exa 8.17	c chart	83
Exa 8.18	u chart	84
Exa 8.19	total proportion out of specification	85
Exa 9.4	Large Sample Confidence Interval for a Mean	87
Exa 9.6	Determining Sample Size to Estimate Mean	87
Exa 9.7	lower limit for the mean lifetime of batteries	88
Exa 9.8	Confidence Interval for a Mean Based on t distribution	88
Exa 9.9	Large Sample Confidence Interval for a Proportion	89
Exa 9.10	number of workers	89
Exa 9.11	Confidence Interval for a Variance	90
Exa 9.12	Confidence Interval for a Difference in Means	90
Exa 9.13	Confidence Interval for a Linear Function of Means	91
Exa 9.14	Normal Distributions with Common Variance	92

Exa 9.15	confidence interval for difference in mean denier	92
Exa 9.16	95 percent confidence level	93
Exa 9.17	95 percent confidence interval for normal distribution	94
Exa 9.18	Normal Distributions with Unequal Variances	95
Exa 9.19	Two sample T test for Chemical vs Atmospheric	95
Exa 9.20	Large Sample Confidence Interval for a Difference in Proportions	96
Exa 9.21	95 percent confidence interval	97
Exa 9.22	Confidence Interval for a Ratio of Population Variances	97
Exa 9.23	A Prediction Interval	98
Exa 9.24	Tolerance Intervals	98
Exa 9.25	the confidence coefficient	99
Exa 9.29	a 95 percent confidence interval for theeta	99
Exa 9.30	confidence interval	100
Exa 10.8	Testing for mean	101
Exa 10.9	Hypothesis testing at 5 percent significance level	102
Exa 10.10	Observed Significance Level or p value	102
Exa 10.11	p value for the situation	103
Exa 10.12	hypothesis about the population mean	103
Exa 10.13	the probability of a type II error	104
Exa 10.14	Determining Sample Size	105
Exa 10.15	Testing a Mean Normal Distribution Case	105
Exa 10.16	Hypothesis about contradicting the manufacturers claim	106
Exa 10.18	Testing for proportion Large sample case	106
Exa 10.19	Testing for variance Normal distribution case	107
Exa 10.20	Testing the difference between two means	107
Exa 10.21	Checking the condition of equal variances	108
Exa 10.22	Checking the condition of equal variances	109
Exa 10.23	Testing the Difference between 2 Means Unequal Variances Case	109
Exa 10.24	Testing the Difference between Means for Paired Samples	110
Exa 10.25	Testing the Difference between Means for Paired Samples	111
Exa 10.26	Testing the ratio of variances Normal distributions case	111
Exa 10.27	Testing Parameters of the Multinomial Distribution ChiSquare Test	112
Exa 10.28	Testing Equality among Binomial Parameters ChiSquare Test	113
Exa 10.29	Test of Independence ChiSq test	114

Exa 10.30	ChiSq test	115
Exa 10.31	Kolmogorov Smirnov test	116
Exa 10.32	Kolmogrov Smirnov Normality Test	116
Exa 10.33	Kolmogrov Smirnov Normality Test	117
Exa 11.2	SSE for the least squares line	118
Exa 11.3	95 percent confidence interval for the slope β_1 . . .	118
Exa 11.4	Testing the Slope of a Straight Line Model T test . . .	119
Exa 11.5	fitting a line	119
Exa 11.6	association between the test strength	120
Exa 11.7	tool life and the cutting speeds	121
Exa 11.8.a	confidence interval for the mean peak power load . . .	121
Exa 11.8.b	Predict the peak power load for a day	122
Exa 11.9.a	Estimate the mean wall reduction	123
Exa 11.9.b	Predict the amount of wall reduction	123
Exa 11.10	Polynomial Regression of degree 2	124
Exa 11.14	Fitting the model The least squares approach	124
Exa 11.15	Estimation of error variance s^2	125
Exa 11.16	Testing the Utility of a Multiple Regression Model The Global F test	125
Exa 11.17	least square fit of the modal	126
Exa 11.18	Estimating and testing hypotheses about β_2	127
Exa 11.19	model for mean lost work hours	128
Exa 11.20	multiple regression model for estimation and prediction	129
Exa 11.21	least squares equation to predict tool life	129
Exa 11.22	A Test for a Portion of a Model	131
Exa 11.23.b	Testing a Portion of a Model F test	132
Exa 11.25	Representation of Mean Profit in the Additive Model .	132
Exa 11.26	Response surface method	134
Exa 11.27	Modeling a time trend	135
Exa 11.28	Logistic regression	135
Exa 12.2	Test to Compare k Treatment Means for a Completely Randomized Design	137
Exa 12.3	mean score for the three groups of managers	138
Exa 12.4	test for mean counts show significant differences . . .	138
Exa 12.5	Equivalence between a t test and an F test	139
Exa 12.6	common variance using a pooled sample variance . . .	140
Exa 12.7	modal for the test score of one manager	140

Exa 12.8	Confidence Intervals for Means in the Completely Randomized Design Bonferroni Method	141
Exa 12.9	confidence intervals for the pairwise difference in mean stop times	142
Exa 12.10	Test to Compare k Treatment Means for a Randomized Block Design	143
Exa 12.11	difference in mean gains	145
Exa 12.12	ANOVA for RBD and regression analysis	145
Exa 12.13	Confidence Intervals for Means in the Randomized Block Design	146
Exa 12.14	Analysis of variance for the factorial experiment	148
Exa 12.15	confidence intervals for the six possible differences between treatment means	148
Exa 12.16.a	Fitting higher order models	149
Exa 12.17	Factorial Design	150
Exa 12.18	effect on defrosted fish by freezing method defrosting method and duration	151
Exa 12.19	yield differ significantly by temperature pressure and reaction time	152

Chapter 1

Data Collection and Exploring Univariate Distributions

R code Exa 1.1 Auto paint shop relative freq

```
1 complain = c(32,17,5,4,3,3,1,1)
2 n= sum(complain)
3 rf=complain/n
4 cat("Relative frequency :", rf)
5
6 a =function(x){
7   sum( rf[1:x])
8 }
9
10 cat("CRF:")
11 for (i in 1:8) {
12   cat( a(i)," ")
13 }
```

R code Exa 1.2 Air quality relative frequency

```

1 LA_days= c(155,138,36,5)
2 O_days= c(233,39,1,1)
3
4 n1=sum(LA_days)
5 rf1 =LA_days/n1
6 n2=sum(O_days)
7 rf2=O_days/n2
8 cat("Relative frequency for LA:", rf1)
9 cat("Relative frequency for Oriando :",rf2)
10
11
12 a =function(x){
13   sum( rf1[1:x])
14 }
15 cat("Cumilative frequency for LA: ")
16 for (i in 1:4) {
17   cat( a(i)," ")
18 }
19
20 b =function(x){
21   sum( rf2[1:x])
22 }
23 cat("Cumilative frequency for Ontario: ")
24 for (i in 1:4) {
25   cat( b(i)," ")
26 }

```

R code Exa 1.3 histograms

```

1 source =c("Fuel","Industrial","Transport","Misc")
2 C01990 =c(5.510,5.582,76.635,11.122)
3 C02000 =c(4.500,7.521,76.383,20.806)
4 data1 <- data.frame(C01990, source)
5 data2 <- data1[order(data1[,1],decreasing=TRUE),]
6 barplot(data2[,1],names.arg=data2[,2],ylim = c

```

```

      (0,100),space = 0)
7
8 data3 <- data.frame(CO2000,source)
9 data4 <- data3[order(data3[,1],decreasing=TRUE),]
10 barplot(data4[,1],names.arg=data4[,2],ylim = c
      (0,100),space = 0)
11
12 VOC1990 =c(1.005,10.000,8.988,1.059)
13 VOC2000 =c(1.206,8.033,8.396,2.710)
14
15 data5 <- data.frame(VOC1990, source)
16 data6 <- data5[order(data5[,1],decreasing=TRUE),]
17 barplot(data6[,1],names.arg=data6[,2],ylim = c
      (0,100),space = 0)
18
19 data7 <- data.frame(VOC2000,source)
20 data8 <- data7[order(data7[,1],decreasing=TRUE),]
21 barplot(data8[,1],names.arg=data8[,2],ylim = c
      (0,100),space = 0)
22
23 #Graph for VOC compounds plotted in book is
      incorrect

```

R code Exa 1.4 Dotplot

```

1 #Number of AQI exceedences for 1990, 1998, and 2006
2
3 year1 =c(42,0,5,0,9,11,51,2,161,15,39,18,0,2,14)
4 year2 =c(50,0,10,24,7,17,38,14,49,14,37,39,0,3,44)
5 year3 =c(18,1,5,13,6,6,18,11,34,11,18,36,2,5,18)
6
7
8 dat=data.frame(year1,year2,year3)
9 library(ggplot2)
10

```

```

11 plot1= ggplot(dat,aes(x=year1)) + geom_dotplot(
      dotsize = 0.75,binwidth = 3.5) +xlab("1990")
12 plot2= ggplot(dat,aes(x=year2)) + geom_dotplot(
      dotsize = 0.75,binwidth = 3.5) + xlim(0,150) +
      xlab("1998")
13 plot3= ggplot(dat,aes(x=year3)) + geom_dotplot(
      dotsize = 0.75,binwidth = 4.0) + xlim(0,150) +
      xlab("2006")
14
15 library(grid)
16 grid.newpage()
17 grid.draw(rbind(ggplotGrob(plot1), ggplotGrob(plot2)
      ,ggplotGrob(plot3), size = "last"))
18
19 #Number of AQI exceedences by city
20
21
22 boston =c(0,0,0,0,0,0,0,0,0,4,0,3,9,8,1,4,1)
23 houston =c
      (51,36,32,27,38,65,26,46,38,51,42,28,21,31,22,28,18)
24
25 Newyork =c
      (15,30,4,11,13,17,11,22,14,22,19,19,27,11,6,15,11)
26
27
28
29 plot4= ggplot(dat1,aes(x=boston)) + geom_dotplot(
      dotsize = 0.75,binwidth = 1.0) +xlab("Boston") +
      xlim(0,70)
30 plot5= ggplot(dat1,aes(x=houston)) + geom_dotplot(
      dotsize = 0.5,binwidth = 1.5) +xlab("Houston") +
      xlim(0,70)
31 plot6= ggplot(dat1,aes(x=Newyork)) + geom_dotplot(
      dotsize = 0.5,binwidth = 1.5) +xlab("New York") +
      xlim(0,70)
32

```

```
33 grid.newpage()
34 grid.draw(rbind(ggplotGrob(plot4), ggplotGrob(plot5)
,ggplotGrob(plot6), size = "last"))
```

R code Exa 1.8 AQI mean and median

```
1 data =c(12, 8, 10, 5, 17, 19, 31, 11, 88, 11, 19,
37, 1, 2, 12 )
2 cat("Mean =",mean(data))
3 cat("Median =",median(data))
```

R code Exa 1.9 accidents mean and median

```
1 n=c(3133,2065)
2 x=c(13.67,8.97)
3 cat(" the mean transportation time in Alabama is ",
weighted.mean(x,n),"min")
```

R code Exa 1.10 AQI quantile

```
1 #The summary of AQI data for year 2003 are as
follows
2
3 data=c(1, 2, 5, 8, 10, 11, 11, 12, 12, 17, 19, 19,
31, 37, 88)
4 quantile(data)
```

R code Exa 1.12 AQI sd and variance

```
1 data=c(1, 2, 5, 8, 10, 11, 11, 12, 12, 17, 19, 19,
        31, 37, 88)
2 cat("Mean :", mean(data))
3 cat("Variance :", var(data))
4 cat("Standard deviation :", sd(data))
```

R code Exa 1.13 histogram of cars

```
1 domestic =c(1995 , 2001 , 2004, 2000 , 2002 , 2002
, 2000 , 2001, 1999 , 2002 , 2004, 2001, 2002
,1996,
2          1990, 1995 , 1992 ,1995 , 1999 , 1996
, 1999 , 1999 ,1998 , 2001, 2002 ,
, 2004 ,2004 , 2004,
3          2001, 1997 , 2002 , 2001 , 2002 ,2001,
, 2000 , 2002 , 1999 , 2001, 2002
,2000 , 2003, 2001 ,
4          2001 , 1999 , 2002 , 2001 , 2002 , 2001
, 2000 , 2002 ,2001 , 2002, 2000 ,
, 2000 , 2002 , 2001 ,
5          2002 , 2002 , 2001 , 2001 ,2002 ,2002
, 2003 , 2003 , 2002, 2001 , 2002 ,
, 2001, 2002 , 2003 ,
6          2002
7 )
8
9 summary(domestic)
10
11 foriegn =c(1997 ,2000, 2002 ,2002 ,2001 , 2003 ,
, 1995 , 1990 , 1992 ,1991 , 1997 , 2000 , 2000,
, 1998 ,
12          2000 , 1998, 2001 , 2004 , 2001 , 2000,
, 2001 , 2000 , 2002 , 2003 ,2003 ,
```

```

      2002 , 2002 , 2002 ,
13      2003 , 1999 , 2000 , 2001 , 2003 , 2003 ,
      2000 , 2001)
14 summary(foreign)
15
16 hist(domestic,xlab = "Year of car advertised",main =
      "Domestic",xlim = c(1988,2006),breaks = 16)
17 hist(foreign,xlab = "Year of car advertised",main =
      "Foreign",xlim = c(1988,2006),breaks = 16)

```

R code Exa 1.14 lawnmower mean and sd

```

1 prev_mean = 500
2 prev_Sd = 125
3
4 #a
5 # Increase the price of each lawnmower by $50.00.
6 cat("Changed mean= ",prev_mean+50)
7 cat("SD remains unchanged")
8
9 #b
10 # Increase the price of each lawnmower by 10%.
11 cat("Changed mean= ",(1.1)*prev_mean)
12 cat("Changed sd= ",prev_Sd*(1.1))

```

R code Exa 1.16 summary of CO and SO2

```

1 # Mean,sd and no. of observations for CO and SO2 are
  as follows:
2 CO =c( 37298.013 , 84369.21 ,3143)
3 SO2 =c( 5616.0483 , 18869.243 ,3143)
4
5 z_CO = (189966.99-CO[1]) /CO[2]

```

```

6 z_S0 = (91310.67-S02[1]) /S02[2]
7
8 cat("The z-score when carbon monoxide emission
    estimates at 189,966.99 :",z_C0)
9 cat("The z-score when sulfur dioxide emission
    estimates at 91,310.67 :",z_S0)

```

R code Exa 1.17 Boxplots

```

1 #Boxplots
2
3 boston =c(0,0,0,0,0,0,0,0,0,4,0,3,9,8,1,4,1)
4 houston =c
    (51,36,32,27,38,65,26,46,38,51,42,28,21,31,22,28,18)
5 Newyork =c
    (15,30,4,11,13,17,11,22,14,22,19,19,27,11,6,15,11)
6
7 boxplot(boston,houston,Newyork,names = c("Boston","
    Houston","New YOrk"))

```

Chapter 2

Exploring Bivariate Distributions and Estimating Relations

R code Exa 2.1 Projections on net new workers

```
1 White =c(23,24)
2 Black =c(9,6)
3 Asian =c(7,6)
4 Hispanic =c(13,12)
5 gender =c(2,1)
6 df =data.frame(White,Black,Asian,Hispanic,gender)
7 means<-aggregate(df,by=list(df$gender),mean)
8 means<-means[,2:length(means)]
9 library(reshape2)
10 means.long<-melt(means,id.vars="gender")
11 library(ggplot2)
12 ggplot(means.long,aes(x=variable,y=value,fill=factor
    (gender)))+
13   geom_bar(stat="identity",position="dodge")+
14   scale_fill_discrete(name="Gender",
15                       breaks=c(1,2),
16                       labels=c("Men", "Women"))+
```

```
17   xlab("Category")+ylab("Mean Percentage")
```

R code Exa 2.3 Timeplot of CO and VOC emission

```
1  year =c
      (1989,1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000)

2  CO =c
      (106.439,99.119,101.797,99.007,99.791,103.713,94.058,104.600,105.4

3  VOC =c
      (22.513,21.053,21.249,20.862,21.099,21.683,20.918,19.924,20.325,1

4  plot(year,CO)
5  lines(year,CO)
6  plot(year,VOC)
7  lines(year,VOC)
```

R code Exa 2.4 Average number of days with AQI greater than 100

```
1  x =c
      (24.60,28.00,19.93,22.40,20.93,25.73,14.46,18.13,23.06,26.53,20.5

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

3  plot(year,x)
4  lines(year,x)
```

R code Exa 2.5 Fuel consumption and efficiency of cars and vans

```

1 year=c(1970,1975,1980:1999)
2 Car =c
    (67.8,74.3,70.2,69.3,69.3,70.5,70.8,71.7,73.4,73.5,73.5,74.1,69.8

3 Van =c
    (12.3,19.1,23.8,23.7,22.7,23.9,25.6,27.4,29.1,30.6,32.7,33.3,35.6

4
5 #Graph for fuel consumption
6 plot(year,Car,ylim = c(10,80),ylab = "Fuel
    Consumption",xlab = "Year" )
7 lines(year,Car)
8 par(new =TRUE)
9 plot(year,Van,ylim = c(10,80),ylab = "Fuel
    Consumption",xlab = "Year")
10 lines(year,Van)
11
12 Car1 =c
    (13.5,14.0,16.0,16.5,16.9,17.1,17.4,17.5,17.4,18.0,18.8,18.0,20.3

13 Van1 =c
    (10.0,10.5,12.2,12.5,13.5,13.7,14.0,14.3,14.6,14.9,15.4,16.1,16.1

14
15 #Graph for fuel efficiency
16 plot(year,Car1,ylim = c(10,30),ylab = "Miles per
    gallon",xlab = "Year" )
17 lines(year,Car1)
18 par(new =TRUE)
19 plot(year,Van1,ylim = c(10,30),ylab = "Miles per
    gallon",xlab = "Year")
20 lines(year,Van1)

```

R code Exa 2.6 Annual temperatures at Newnan

```

1 temp =c(59.64 , 61.98 , 60.78 , 61.61 , 61.57 ,
          63.02 , 63.20 , 63.49 , 62.81 , 62.22 , 65.00 ,
          61.57 , 63.33 , 62.78 , 63.43 , 64.10 , 62.39 ,
          63.55 , 63.87 ,
2          61.91 , 65.24 , 64.27 , 62.07 , 60.89 ,
          64.03 , 62.79 , 65.14 , 62.39 , 63.03 ,
          62.76 , 65.06 , 64.20 , 65.10 , 62.51 ,
          62.89 , 62.74 , 61.40 , 62.43 ,
3          62.38 , 59.32 , 61.92 , 60.52 , 61.63 ,
          63.09 , 62.39 , 63.32 , 61.19 , 62.85 ,
          62.90 , 61.65 , 62.04 , 62.71 , 62.53 ,
          62.92 , 62.32 , 62.47 , 62.64 ,
4          60.39 , 61.92 , 60.72 , 61.12 , 62.23 ,
          60.58 , 61.10 , 61.09 , 59.78 , 60.53 ,
          60.00 , 60.25 , 61.69 , 61.29 , 61.47 ,
          61.48 , 61.13 , 61.41 , 59.57 ,
5          60.70 , 60.49 , 60.15 , 61.08 , 60.38 ,
          60.38 , 58.65 , 60.22 , 60.71 , 61.92 ,
          60.60 , 60.05 , 60.17 , 62.93 , 61.90 ,
          60.13 , 60.41 , 60.88 , 61.16 ,
6          60.59 , 59.98 , 61.48 , 61.34 , 59.06
7 )
8 year=c(1901:2000)
9 plot(year,temp)
10 lines(year,temp)

```

R code Exa 2.7 scatterplot of heat exchangers

```

1 OT =c
  (-0.0010 , -0.0010 , -0.0005 , -0.0005 , -0.0005 , -0.0005 , 0.0005 , 0.0010 , 0.
2 WR =c
  (1.30 , 3.00 , 1.60 , 3.50 , 4.25 , 4.30 , 3.80 , 3.80 , 2.70 , 4.40 , 4.90 , 2.80 , 3.50
3

```

```
4 plot(OT,WR)
```

R code Exa 2.8 correlation coeff for power plant

```
1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3
4 r= cor(x,y)
5 cat("The fact that the value of r i.e.,"r,"is
      positive and near 1 indicates that the peak power
      load is very strongly associated with the daily
      maximum temperature")
```

R code Exa 2.9 correlation coeff for O3 and SO2

```
1 # Mean,sd and no. of observations for CO and SO2 are
   as follows:
2 O3 =c
      (0.023,0.029,0.032,0.014,0.016,0.024,0.021,0.017,0.044,0.038,0.023)
3 SO2 =c
      (0.019,0.030,0.075,0.047,0.009,0.043,0.031,0.039,0.010,0.046,0.023)
4
5 r=cor(O3,SO2)
6 cat("The value of r is very small and doesn't shw
      any association between the O3 and SO2 levels
      measured.")
```

R code Exa 2.12 equation of the least squares regression line


```

1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3
4 pol <- lm(y~x)
5 coef = coefficients(pol)
6
7 #a
8 cat("beta1 : ",coef[2])
9 cat("y intercept i.e beta0: ",coef[1])
10
11 #b
12 cat("Peak power load increased by :", 5*coef[2],"
      megawatts when the maximum temperature increases
      by 5 F. ")
13
14 #c
15 cat("y = ",coef[1]," + ",coef[2],"x")
16
17 # Straight-line fit to power load and temperature
    data
18 abline(pol)

```

R code Exa 2.13 linear fit OT and WR

```

1 OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
2 WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50

3 pol <- lm(WR~OT)
4 summary(pol)
5 coef=coefficients(pol)
6 cat("WR = ",coef[1]," + ",coef[2],"OT")

```

R code Exa 2.14 Predict peak power load

```
1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3
4 pol <- data.frame(x,y)
5
6 line <- lm(y~x,data = pol)
7
8 #a
9 # here x0 <- 95 F
10 data_a <- data.frame(x=95)
11 res_a <- predict(line,data_a)
12 cat(" the fitted relation tells us that the likely
      peak load will be around ",res_a," megawatts. ")
13
14 #b
15 # here x0 <- 98 F
16 data_b <- data.frame(x=98)
17 res_b <- predict(line,data_b)
18 cat(" It predicts the peak power load of ",res_b, "
      megawatts for the day with maximum temperature 98
      F. ")
19
20 #c
21 # here x= <- 102 F
22 data_c <- data.frame(x=102)
23 res_c <-predict(line,data_c)
24 cat(" It predicts the peak power load of ",res_c, "
      megawatts for the day with maximum temperature
      102 F.")
```

R code Exa 2.15 Predict wall reduction

```
1 OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
2 WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50

3 pol <- lm(WR~OT)
4 data <- data.frame(OT=0.010)
5 res <- predict(pol,data)
6 cat("A wall reduction of",res,"is predicted for OT =
    0.010. ")
```

R code Exa 2.16 coeff of determination for the peak power load

```
1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3
4 pol <- data.frame(x,y)
5
6 line <- lm(y~x,data = pol)
7 r_sq <- summary(line)$r.squared
8 cat(" the sample variability of the peak load about
    their mean is reduced by ",r_sq*100," when the
    mean peak loads
9     is modeled as a linear function of daily high
    temperature")
```

R code Exa 2.18 Least squares fit and residual plot for WR OT data

```

1 OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
2 WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
3
4 #Least square fit
5
6 fit =lm(WR~OT)
7 summary(fit)
8 plot(WR~OT)
9 plot(OT,WR)
10 abline(reg = fit)
11
12 #Residual Plot
13 plot(OT,resid(fit))
14 abline(h=0)

```

R code Exa 2.19 linear regression model for population density

```

1 year=c
    (1790,1800,1810,1820,1830,1840,1850,1860,1870,1880,1890,1900,1910
2 pop =c
    (4.5,6.1,4.3,5.5,7.4,9.8,7.9,10.6,10.9,14.2,17.8,21.5,26.0,29.9,3
3
4 pol <- data.frame(x,y)
5
6 line <- lm(pop~year,data = pol)
7 summary(line)
8 coef<- coefficients(line)
9
10 #Regression Analysis: PopDens versus Year

```

```

11 #The regression equation is :
12
13 cat("Pop_Dens = ",coef[1]," + ",coef[2]," year")
14
15
16 line2 <- lm(log(pop)~year,data = pol)
17 summary(line2)
18 coef2<- coefficients(line2)
19
20 #Regression Analysis: ln(PopDens) versus Year
21 #The regression equation is :
22
23 cat("log(y) = ",coef2[1]," + ",coef2[2]," year")

```

R code Exa 2.20 linear regression modal for length and wt

```

1 wt =c
    (130,51,640,28,80,110,33,90,36,38,366,84,80,83,70,61,54,44,106,84

2 len =c
    (94,74,147,58,86,94,63,86,69,72,128,85,82,86,88,72,74,61,90,89,68

3
4 pol <- data.frame(len,wt)
5
6 #Regression Analysis: Weight versus Length
7 #The regression equation is :
8
9 line1 <- lm(wt~len,data = pol)
10 summary(line1)
11 coef1<- coefficients(line1)
12
13 cat("Weight = ",coef1[1] ," + ",coef1[2]," length")
14
15 # Output for the linear growth model

```

```

16 plot(len,wt)
17
18 #Residual plot
19 plot(len,resid(line1))
20
21
22 #Regression Analysis: ln(Weight) versus Length
23 #The regression equation is :
24
25 line2 <- lm(log(wt)~len,data = pol)
26 summary(line2)
27 coef2<- coefficients(line2)
28
29 cat("log (Weight) = ",coef2[1] ," + ",coef2[2] ,"
      length")
30
31 #Fit of ln(weight) versus length of alligators
32 plot(len,log(wt))
33
34 #Residual plot
35 plot(len,resid(line2))
36
37
38 #Regression Analysis: ln(Weight) versus ln(Length )
39 #The regression equation is :
40
41 line3 <- lm(log(wt)~log(len),data = pol)
42 summary(line3)
43 coef3<- coefficients(line3)
44
45 cat("log (Weight) = ",coef3[1] ," + ",coef3[2] ,"log (
      length)")
46
47 #Fit of ln(weight) versus ln(length) for alligators
48 plot(log(len),log(wt))
49
50 #Residual plot
51 plot(log(len),resid(line3))

```


Chapter 4

Probability

R code Exa 4.1 prob of eastern and western cities

```
1 prob_east <- 4
2 prob_west <- 2
3 prob_east_and_west <- prob_east * prob_west
4 prob_east_and_west
```

R code Exa 4.2 die example

```
1 dice_outcomes = c(1,2,3,4,5,6)
2 A= c(2,4,6)
3 B= c(1,3,5)
4 C= c(5,6)
5 prob_even = length(A)/length(dice_outcomes)
6 prob_even
7 prob_odd = length(B)/length(dice_outcomes)
8 prob_odd
9 prob_greater_than_4 = length(C)/length(dice_outcomes
)
10 prob_greater_than_4
```

R code Exa 4.5 Venn diag EE

```
1 total_students = 100
2 calculus = 30
3 signal_processing = 25
4 calculus_and_signal_processing = 10
5 not_calculus = total_students - calculus
6 not_calculus
7 calculus_or_signal_processing = calculus + signal_
  processing - calculus_and_signal_processing
8 calculus_or_signal_processing
9 not_calculus_and_not_signal_processing = total_
  students - calculus_or_signal_processing
10 not_calculus_and_not_signal_processing
```

R code Exa 4.6 Venn diag labor statistics

```
1 all_woman = 52
2 all_white = 37
3 white_woman = 23
4 woman_or_white = all_woman + all_white - white_woman
5 woman_or_white
```

R code Exa 4.7 Venn diag electric motors

```
1 total_electric_motors = 20
2 defect_free = 11
3 finish_defect = 8
4 assembly_defect = 3
```

```

5 defected_motors = total_elecric_motors - defect_free
6 finish_defect_and_assembly_defect = finish_defect +
  assembly_defect - defected_motors
7 finish_defect_and_assembly_defect
8 finish_defect_or_assembly_defect = defected_motors
9 finish_defect_or_assembly_defect
10 only_one_defect = finish_defect_or_assembly_defect -
  finish_defect_and_assembly_defect
11 only_one_defect

```

R code Exa 4.12 product rule example

```

1 ways_to_select_I_item <- 2
2 ways_to_select_II_item <- 2
3 ways_to_select_both_items <- ways_to_select_I_item *
  ways_to_select_II_item
4 ways_to_select_both_items

```

R code Exa 4.13 Product rule for cities

```

1 prob_E_selected_in_west <- 0.5
2 prob_E_gets_selected <- prob_E_selected_in_west
3 prob_E_gets_selected

```

R code Exa 4.16 Permutation of employees

```

1 cat("No. of possible assignments ", factorial(10)/
  factorial(7))

```

R code Exa 4.17 Permutation of operations

```
1 cat("No. of orderings", factorial(4))
```

R code Exa 4.18 Permutation of divisions

```
1 cat("prob D2 hiighest if all have equal preferences  
  is ", factorial(3)/factorial(4))  
2 cat("prob D2 is I and D3 II is", factorial(2)/  
  factorial(4))
```

R code Exa 4.19.a selection of employees

```
1 cat("No. of possible selections ", choose(10,3))
```

R code Exa 4.19.b Probability of female candidate

```
1 cat("No. of ways to select 1 female ", choose(2,1))  
2 a <- choose(2,1)  
3 cat("No. of ways to select other 2 employees ",  
  choose(8,2))  
4 b <- choose(8,2)  
5 cat("total ways of selecting 3 employees",choose  
  (10,3))  
6 c <- choose(10,3)  
7 d <- (a*b)/c  
8 cat("prob is", d)
```

R code Exa 4.20 Combination of applicants

```
1 cat("No. of ways to select 2 applicants", choose
    (5,2))
2 a <- choose(5,2)
3 cat("No. of ways to select one of the best ", choose
    (2,1))
4 b <- choose(2,1)
5 cat("No. of ways to select other employee", choose
    (3,2))
6 c <- choose(3,2)
7 cat("prob is", (b*c)/a)
```

R code Exa 4.21.a Partitioning of employees

```
1 cat("No. of ways of assigning job", factorial(10)/(
    factorial(3)*factorial(4)*factorial(3)))
```

R code Exa 4.21.b Partitioning of employees specifically

```
1 cat("No. of ways three employees of a certain ethnic
    group get assigned to job I", factorial(7)/(
    factorial(4)*factorial(3)))
2 a <- factorial(7)/(factorial(4)*factorial(3))
3 cat("No. of ways of assigning job", factorial(10)/(
    factorial(3)*factorial(4)*factorial(3)))
4 b <- factorial(10)/(factorial(3)*factorial(4)*
    factorial(3))
5 cat("prob is", a/b)
```

R code Exa 4.25 conditional probability of motors

```
1 cat("probability to select 2 non-defective motors ",  
    4*3/20)  
2 a <- 12/20  
3 cat("probability to select 2 motors in which I is  
    non defective ", 4*4/20)  
4 b <- 16/20  
5 cat("probability II motor is non defective if I one  
    is non-defective ", a/b)
```

R code Exa 4.28 independent event of job

```
1 cat("Prob worker 1 or 2 is selected ", (1/4)+(1/4))  
2 A <- (1/4)+(1/4)  
3 cat("Prob worker 1 or 3 is selected" , (1/4)+(1/4))  
4 B <- (1/4)+(1/4)  
5 cat("Prob worker 1 is selected ",(1/4))  
6 C <- 1/4  
7 cat("Prob(AB) is worker 1 is selected ", 1/4)  
8 AB <- 1/4  
9 A*B  
10 cat("Since A*B is equal to AB, A and B are  
    independent")  
11 cat("Prob(AC) is worker 1 is selected", 1/4)  
12 A*C  
13 cat("Since A*C is not equal to AC, A and C are not  
    independent")
```

R code Exa 4.29 Complementary Events

```
1 cat("ways in which no line chosen more than once ",  
    factorial(10)/factorial(5))  
2 a <- factorial(10)/factorial(5)  
3 cat("total no. of ways to select a line ", 10 ^ 5)  
4 b <- 10^5  
5 cat("prob no line chosen more than once " ,a/b)  
6 c <- a/b  
7 cat("prob atleast one line chosen more than once " ,  
    1-c)
```

R code Exa 4.31 additive rule

```
1 cat("Prob students enrolled in calculus class", 50/  
    100)  
2 a<- 50/100  
3 cat("Prob students enrolled in class" , 45/100)  
4 b <- 45/100  
5 cat("prob students enrolled in both courses" ,10/  
    100)  
6 c <-10/100  
7 cat("prob students enrolled in atleast one course", a  
    +b-c)  
8 cat("prob of students enrolled in none of the  
    courses", 1-(a+b-c))
```

R code Exa 4.32 multiplicative rule for defects

```
1 defect_free_prob <-0.75  
2 cat("prob of defected item ", 1- defect_free_prob)  
3 defected <- 1- defect_free_prob  
4 shaft <- 0.20
```

```

5 bushing <- 0.10
6 both_defect_prob <- shaft + bushing -defected
7 cat("prob of only shaft defect is ", shaft - both_
    defect_prob)

```

R code Exa 4.33 multiplicative rule for relays

```

1 cat("prob both relays r1 nad r2 open ", 0.2 * 0.2)
2 E1 <- 0.2 * 0.2
3 cat("Prob r1 open r2 closed" , 0.2*0.8)
4 E2 <- 0.2*0.8
5 cat("prob r1 closed r2 open ",0.8*0.2)
6 E3<-0.8*0.2
7 cat("Prob r1 r2 both closed ", 0.8*0.8)
8 E4 <-0.8*0.8
9 cat("Prob current will flow", E2+E3+E4)

```

R code Exa 4.35 example of ballpoint pen

```

1 A1<- 80
2 A2<- 120
3 DTA1<-8
4 DTA2<-2
5 DFA1<-13
6 DFA2<-27
7 NDA1<-59
8 NDA2<-91
9 TOTAL<-200
10 cat("overall defective trash rate", (DTA1+DTA2)/
    TOTAL)
11 cat("prob that pen is defective and produced by
    assembly line1 ", DTA1/TOTAL)

```

```
12 cat("Prob pen is defective if produced by line1 ", (
    DTA1/TOTAL)/(A1/TOTAL))
```

R code Exa 4.37 Bayes Rule

```
1 supplier1<-0.40
2 supplier2<-0.60
3 defective_supplier1<-0.10
4 defective_supplier2<-0.05
5 cat("prob tire comes from supplier1 if it is
    defective", (supplier1*defective_supplier1)/((
    supplier1*defective_supplier1)+
6
```

R code Exa 4.39 flight accident

```
1 x= c( 36 ,11 ,51 ,2 ,4 ,13 ,12 ,66 ,9 ,18 ,22 ,14
    ,54 ,10 ,78 ,21 ,14 ,56 ,9 ,100)
2 m <- matrix(x,byrow = TRUE,nrow =4 )
3 rownames(m) <- c(" Business", " Instructional ", "
    Personal", " All")
4 colnames(m) <- c(" Fatal", " Minor", " Serious", " None", "
    All")
5
6 #a) An accident on a business flight
7 m [" Business", " All"] /100
```



```

8
9 #b) An accident that resulted in a fatal injury
10 m["All", "Fatal"] / 100
11
12 #c) An accident that resulted in a minor injury
    given that it was on a business flight
13 m["Business", "Minor"] / 100
14
15 #d) An accident on a business flight that resulted
    in a minor injury
16 m["Business", "Minor"] / 100 * m["Business", "All"] / 100
17
18 #e) An accident on a business flight given that it
    was fatal
19 (m["Business", "All"] * m["Business", "Fatal"] / m["
    All", "Fatal"]) / 100

```

R code Exa 4.40 Bays rule for tower malfunction

```

1 cat("prob of exactly 1 plugging malfunction", (0.096
    * 0.904 * 0.904) + (0.904 * 0.096 * 0.904) + (0.904 * 0.904 *
    0.096))
2 cat("prob of atleast 1 plugging malfunction",
    1 - (0.904 * 0.904 * 0.904))
3 cat("prob of exactly two plugging malfunctions when
    atleast 1 is due to plugging",
4     ((0.096 * 0.096 * 0.904) + (0.096 * 0.904 * 0.096) + (0.904 *
    0.096 * 0.096)) / (1 - (0.904 * 0.904 * 0.904)))

```

R code Exa 4.41 Odd ratio

```

1 x = c
    (139, 10898, 11037, 239, 10795, 11034, 378, 21693, 22071)

```

```

2
3 m <- matrix(x,byrow = TRUE,nrow = 3)
4 rownames(m) <- c("Aspirin","Placebo","Total")
5 colnames(m) <- c("MI","NoMI","Total")
6
7 cat("For the aspirin group, the odds in favor of M.I
  . are ",m["Aspirin","MI"]/m["Aspirin","NoMI"])
8
9 cat("For the placebo group, the odds in favor of M.I
  . are",m["Placebo","MI"]/m["Placebo","NoMI"])
10
11 cat("Odds ratio= ",(m["Aspirin","MI"]/m["Aspirin","
  NoMI"])/(m["Placebo","MI"]/m["Placebo","NoMI"]))
  )

```

R code Exa 4.42 Odd ratio

```

1 #Employment Status by Gender
2
3 x =c(64046,3141,55433,2556)
4 m <- matrix(x,byrow = TRUE,nrow = 2)
5 rownames(m) <- c("Emp","Unemp")
6 colnames(m) <- c("M","F")
7
8 cat("Odds ratio:",( m[1,1]*m[2,2])/(m[1,2]*m[2,1]))
9
10 #Employment Status by Education
11 y =c( 36249, 1962, 39250, 1165)
12 m <- matrix(y,byrow = TRUE,nrow = 2)
13 rownames(m) <- c("Emp","Unemp")
14 colnames(m) <- c("HS","College")
15
16 cat("Odds ratio:",( m[1,1]*m[2,2])/(m[1,2]*m[2,1]))
17 z= ( m[1,1]*m[2,2])/(m[1,2]*m[2,1])
18

```

```
19 cat(" risk of unemployment for those with a high  
    school education is",1/z , " higher than the risk  
    of unemployment for those with college education"  
    )
```

Chapter 5

Discrete Probability Distributions

R code Exa 5.1 Relay example

```
1 a <- function(x){  
2  
3 (0.8^x)*(0.2^(2-x))  
4 }  
5  
6 # The Distribution..  
7 cat("The probability distribution for x=0 is",a(0))  
8  
9 cat("The probability distribution value for x=1 is"  
10     ,2*a(1))  
11 cat("The probability distribution for x=2 is",a(2))
```

R code Exa 5.2 multiplicative rule of probability

```
1 a <- function(x){
```

```

2
3      ((factorial(6)/factorial(6-x))*(factorial(4)/
        factorial(4-(2-x))))/(factorial(10)/factorial
        (8))
4
5  }
6
7  # The Distribution..
8  cat("The probability distribution for x=0 is",a(0))
9
10 cat("The probability distribution value for x=1 is"
      ,2*a(1))
11
12 cat("The probability distribution for x=2 is",a(2))

```

R code Exa 5.3 distribution function

```

1  a <- function(x){
2
3      (0.8^x)*(0.2^(2-x))
4  }
5
6  # The Distribution..
7  cat("The distribution function for b<0 is",0)
8
9  cat("The distribution function for 0<b<1 is",a(0))
10
11 cat("The distribution function for 1<b<2 is",2*a(1)
      + a(0))
12
13 cat("The distribution function for b>=2 is",a(2) + 2
      *a(1) + a(0) )
14
15 b= c(0,1,2,3)
16 fb= c(0,0.04,0.36,1)

```

```
17 plot(b,fb)
```

R code Exa 5.4 mean and sd

```
1 mp =c(3,9,16,21,30,40,55,75,92)
2 year1 =c(7.6,12.8,5.3,10.8,17.3,15.1,18.6,11.3,1.2)
3 year2 =c(6.4,11.6,5.2,9.0,12.5,12.2,22.5,16.0,4.6)
4
5 mean1 =weighted.mean(mp,year1/100)
6 mean2 =weighted.mean(mp,year2/100)
7
8 sd1=sqrt(sum(((mp- mean1)^2)*year1/100))
9 sd2=sqrt(sum(((mp- mean2)^2)*year2/100))
10
11 cat("mean for year 1990s: ",mean1)
12 cat("mean for year 2050s: ",mean2)
13 cat("SD for year 1990s: ",sd1)
14 cat("SD for year 2050s: ",sd2)
```

R code Exa 5.5 expected daily demand and variance

```
1 px <- c(0.1,0.5,0.4)
2 x <- 0:2
3 E <- weighted.mean(x,px)
4 cat("expected daily demand for the tool is ",E)
5
6 V <- sum(((x - E)^2)*px)
7 cat("variance is:", V)
```

R code Exa 5.6 mean and variance of the daily costs

```

1 px <- c(0.1,0.5,0.4)
2 x <- 0:2
3 E <- weighted.mean(x,px)
4
5 #  $E(100X) = 100 E(X)$ 
6 cat("Daily cost of using tool" , 100*E)
7
8
9 V <- sum(((x - E)^2)*px)
10 cat("variance is:", V)
11
12 #  $V(100X) = (100^2)V(X)$ 
13 cat("Variance of daily cost is ", 100*100*V)

```

R code Exa 5.7 mean score and the standard deviation

```

1 probab_marks <- c(0.1,0.2,0.4,0.2,0.1)
2 marks <- 0:4
3 E <- weighted.mean(marks,probab_marks)
4
5 cat("Mean score is:" , E)
6
7 V <- sum(((marks-E)^2)*probab_marks)
8 cat("Standard deviation is ", sqrt(V))

```

R code Exa 5.8.a Tchebyshefs Theorem

```

1 mean <- 120
2 sd <- 10
3 lower_limit <- 100
4 k <- (mean - lower_limit)/sd
5

```

```
6 cat(1-1/k*k, "fraction of days will have prod.  
  between 100 and 140" )
```

R code Exa 5.8.b shortest interval

```
1 mean <- 120  
2 sd <-10  
3 fraction <- 0.90  
4 k <-( 1/(1-fraction))^0.5  
5 cat("shortest interval is from ", mean-(k*sd), "-",  
  mean+(k*sd))
```

R code Exa 5.11 binomial distribution of fuses

```
1  
2 # given probability of defective fuse =0.10  
3  
4 # a)  
5 cat("Probability exactly one fuse in the sample of  
  four is defective", dbinom(1,4,prob = 0.10))  
6  
7 # b)  
8 # Probability atleast one is defective  $P(X \geq 1) = 1 -$   
   $P(X=0)$   
9 none_defective <- dbinom(0,4,prob = 0.10)  
10 cat("Probability that atleast one bulb is defective :  
  ", 1- none_defective)  
11  
12 # c)  
13 n=4  
14 p=0.1  
15 E =n*p  
16 V =n*p*(1-p)
```



```

17
18 E_Y_sq = V+ E^2
19 cat("E(C) =", 3*E_Y_sq)
20 cat(" we could expect to pay an average of $",3*E_Y_
    sq*10, "in repair costs for each shipment of four
    fuses. ")

```

R code Exa 5.12 binomial distribution of battery

```

1 # Probability battery exceeding lifetime of 4 hours
  is 0.135
2
3 # a
4 cat("probability that only one battery lasts 4
    hours or more is ", dbinom(1,3,prob = 0.135))
5
6 #b
7 # probability that at least one battery lasts 4
    hours or more = P(Y>=1)= 1- P(Y=0)
8 no_battery <- dbinom(0,3,prob = 0.135)
9 cat(" probability that at least one battery lasts 4
    hours or more is " , 1- no_battery)

```

R code Exa 5.13 binomial distribution of chemicals

```

1 #a
2 cat("The probability of exactly 3 out of 10 plants
    calling in orders is ",
3     dbinom(3,10,prob = 0.2))
4
5 #b
6 # The probability of at most 3 out of 10 plants
    calling in orders = P(Y<=3) = p(0)+p(1)+p(2)+p(3)

```

```

7 cat("The probability of at most 3 out of 10 plants
   calling in orders is " , pbinom(3,10,prob = 0.2)
   )
8
9 #c
10 # The probability of at least 3 out of 10 plants
   calling in orders = 1- P(Y<3)
11 cat(" The probability of at least 3 out of 10 plants
   calling in orders is" , 1- pbinom(2,10,prob =
   0.2))

```

R code Exa 5.15 binomial distribution of contracts

```

1 # a
2 # The probability that the firm will get none of
   those contracts = P(X=0)
3
4 cat("the probability that the firm will get none of
   those contracts" ,dbinom(0,8,prob = 0.40))
5
6 #b
7 # The probability that the firm will get five out of
   eight contracts = P(X=5)
8
9 cat(" the probability that the firm will get five
   out of eight contracts", dbinom(5,8,prob = 0.40))
10
11 #c
12 # The probability that the firm will get all eight
   contracts
13
14 cat(" the probability that the firm will get all
   eight contracts" , dbinom(8,8,prob = 0.40))

```

R code Exa 5.17 geomatric distribution of interviews

```
1 # the probability that the first applicant having
  advanced training is found on the fifth interview
  = P(Y=5)
2 # using geomatric distribution
3 library(stats)
4 p=0.30
5 cat(" the probability that the first applicant
  having advanced training is found on the fifth
  interview is", dgeom(4,0.30) )
6 cat(" Total cost of interviewing is ", 300/p)
7
8 # V(C)=(300^2)V(Y)
9
10 cat("V(C) is" , ((300^2)*(1-p))/(p^2))
```

R code Exa 5.18 expected time

```
1 #Using negative binomial distribution
2
3 r=3
4 p=0.2
5 E = 10*(r/p) + r*20
6 V= 10*10*(r*(1-p)/(p^2))
7
8 cat(", the total time to use up the kits has an
  expected value of ",E," minutes and a standard
  deviation of",sqrt(V),"minutes")
```

R code Exa 5.19 geomatric distribution of interviews

```
1 # Prob of the applicants for a certain position
  have
2 #advanced training in computer programming of the
  applicants for a certain position
3 #have advanced training in computer programming of
  the applicants for a certain position have
4 #advanced training in computer programming = 0.30
5
6 cat(" The probability that the third qualified
  applicant is found on the fifth interview" ,
7     dnbinom(2,3,0.30))
```

R code Exa 5.20 industrial accidents

```
1 # average no. of accidents in a week is 3
2
3 # using poisson distriution
4
5 #a
6 cat("Prob of no accidents in a week p(0)= ", dpois
  (0,3))
7
8 #b
9 cat("Prob of two accidents in a given week p(2)= ",
  dpois(2,3))
10
11 #c Prob atmost 4 accidents occur in a given week is
12 # p(0)+p(1)+p(2)+p(3)
13
14 cat("Prob atmost 4 accidents occur in a given week
  is ", ppois(4,3))
15
16 #d Average no. of accidents on a given day = 3/7
```

```
17 d <- 3/7
18 cat("Prob of two accidents on any given day is ",
      dpois(2,d))
```

R code Exa 5.22 The Hypergeometric Distribution

```
1 # using hypergeometric distribution
2
3 # the probability that the female is selected for
  one of the jobs = $P(Y=1) = p(1)$ 
4
5 cat("the probability that the female is selected for
  one of the jobs is", dhyper(1,1,5,2))
```

R code Exa 5.23.a The Hypergeometric Distribution of boxes

```
1 # using hypergeometric distribution
2
3 #a
4 # the probability that all five boxes will fit
  properly =  $P(Y=0) = p(0)$ 
5
6 cat("The probability that all five boxes will fit
  properly is", dhyper(0,2,18,5))
```

Chapter 6

Continuous Probability Distributions

R code Exa 6.1 continuous random variable

```
1 a <- function(x){  
2   0.5*exp(-x*0.5)  
3 }  
4  
5  
6 cat("the probability that battery will last longer  
   than 400 hours",integrate(a,4,'infinite')$value)  
7 cat("probability that the lifetime exceeds 9 is",  
   integrate(a,9,'infinity')$value)
```

R code Exa 6.2 probability density function

```
1 a <- function(x){  
2   0.5*exp(-x*0.5)  
3 }  
4
```

```

5 #a
6 cat(" the probability that the lifetime of a
    particular battery of this type is less than 200
    hours",
7     integrate(a,0,2)$value)
8 # ans given in book is wrong ,P(X<2)= 0.6321
9
10 #b
11 cat(" the probability that a battery of this type
    lasts more than 30 hours given that it has
    already been in use for more than 200 hours. ",
12     (integrate(a,3,'infinity')$value)/(integrate(a
    ,2,'infinity')$value))

```

R code Exa 6.4 distribution function

```

1 a = function(x){
2   if(x<0){
3     0
4   } else if(x>=0 && x<=1){
5     x
6   } else if(x>=1 && x<=2){
7     0.5
8   } else {
9     0
10  }
11 }
12
13 cat("the probability that demand will exceed 150
    gallons on a given week",integrate(Vectorize(a)
    ,1.5,2)$value)

```

R code Exa 6.5 Expected value and variance

```

1 a = function(x){
2   3*x*x*x
3 }
4
5 E= integrate(a,0,1)$value
6
7 cat(" on the average , the lathe is in use ", E*100,"
    percent of the time")
8
9 b =function(x){
10   3*x*x*x*x
11 }
12
13 cat(" Variance is ",integrate(b,0,1)$value - E^2 )

```

R code Exa 6.6 expected weekly demand

```

1 a = function(x){
2   if(x<0){
3     0
4   } else if(x>=0 && x<=1){
5     x*x
6   } else if(x>=1 && x<=2){
7     x/2
8   } else {
9     0
10  }
11 }
12
13 b =integrate(Vectorize(a),0,1)$value
14 c= integrate(Vectorize(a),1,2)$value
15 cat("The expected weekly demand for kerosene is ",b+
    c," gallons")

```

R code Exa 6.7 Expected interval

```
1 E =445
2 V =236
3 sd =sqrt(V)
4 p= 0.75
5 k =sqrt((1/(1-p)))
6
7 cat(" This interval isgiven by ",E- k*sd," - ",E+ k*
    sd)
```

R code Exa 6.8.a probability of uniform distribution

```
1
2 # Unniform distriution problem
3
4 #a
5 # the probability that the delivery time is two or
  more days = P(X>=2)
6
7 cat(" the probability that the delivery time is two
  or more days is " , 1-punif(2,1,5))
```

R code Exa 6.9.a Probability of exponential distribution

```
1
2 # Using exponential distriution
3
4 #a
```

```

5 # The probability that any given plant processes
   more than 5 tons of raw sugar on a given day = P(
   Y>=5)
6
7 mean <- 4
8 Theeta <- 1/mean
9 a <- 1 - pexp(5, rate = Theeta)
10 cat("The probability that any given plant processes
      more than 5 tons of raw sugar on a given day is",
      a)

```

R code Exa 6.9.b cumulative probability

```

1
2 # Using binomial distribution
3 #b
4
5 mean <- 4
6 Theeta <- 1/mean
7 a <- 1 - pexp(5, rate = Theeta)
8
9 cat(" the probability that exactly two of the three
      plants process more than 5 tons of raw sugar on a
      given day"
10      , dbinom(2,3,prob = a) )

```

R code Exa 6.10 The Gamma Distribution of components

```

1
2 # using exponential distribution , alpha=1
3 # As Y= X1 + X2 , alpha =2  beta =400
4
5 #b

```

```

6 cat(" Expected value E(Y) =", 2 * 400)
7
8 #c
9 # using Gamma distribution
10 cat(" The probability that the system will survive
      for more than 1000 hours is equal to P(Y>1000) =
      ",
11      1 - pgamma(1000, shape = 2 , scale = 400))

```

R code Exa 6.11 The mean and variance for the length of maintenance times

```

1 alpha <- 3
2 beta <- 2
3 sd <- ((alpha * beta * beta)^2)
4 cat("P(|Y-6| >=14 ) = " , (3.46/14)^2)
5
6
7 cat(" Because P(Y>20 min) is so small , we must
      conclude that our new maintenance man is somewhat
      slower than his predecessor. ")

```

R code Exa 6.12 z value

```

1 cat("P(Z< =1.53) =",pnorm(1.53,lower.tail = TRUE))

```

R code Exa 6.13 Normal distribution

```

1 # using normal distribution
2

```

```

3 #a
4 cat("P(Z <= 1) =", pnorm(1,lower.tail = TRUE))
5
6 #b
7 cat("P(Z < -1.5) =" , pnorm(-1.5,lower.tail = TRUE))
8
9 #c
10 cat("P(Z > 1) =", pnorm(1,lower.tail = FALSE))
11
12 #d
13 cat("P(-1.5 <= Z < =0.5) =", pnorm(0.5)-pnorm(-1.5))
14
15 #e
16 cat("The value of z such that P(Z<=z)= 0.99 is",
      qnorm(0.99,lower.tail = T))

```

R code Exa 6.14 standard normal distribution of bottles

```

1
2 # area using normal distribution
3
4
5 # given population mean = 16 and sd = 1
6
7 cat(" the probability that the machine will dispense
      more than 17 ounces of liquid into any one
      bottle. ",pnorm(17,16,1,lower.tail = F))

```

R code Exa 6.15 Normal distribution

```

1
2 # using normal distribution
3

```

```

4 z <- qnorm(0.05, lower.tail = F)
5 cat("The value of z such that  $P(Z \geq z) = 0.05$  is" , z)
6 sd <- 1.2
7 cat("Therefore, the proper setting for the dial so
      that 17-ounce bottles will overflow only 5% of
      the time is when mean =",
8      17 - sd*z)

```

R code Exa 6.16 Normal distribution of maths score

```

1
2 # using normal distribution
3
4 #a
5 # SAT mathematics scores mean = 480 and sd = 100
6 a <- pnorm(550, 480, 100, lower.tail = T)
7 cat(" percent of students would score less than 550
      in a typical year is  $P(X < 550)$ ", a * 100)
8
9 #b
10 # ACT mathematics scores mean = 18 and sd = 6
11 b <- (550 - 480) / 100
12 cat(" The engineering school set as a comparable
      standard on the ACT math test would be" , 18 + 6 *
      b)
13
14 #c
15 cat(" the probability that a randomly selected
      student will score over 700 on the SAT math test
      =  $P(X > 700)$ " ,
16      pnorm(700, 480, 100, lower.tail = F))

```

R code Exa 6.17.a percentile score

```

1
2 # mean and sd of the batting league is 0.358 and
   0.027 respectively
3
4 #a
5 # z score for Ted Williams is (0.406-0.358)/0.027
6 z1 <- (0.406-0.358)/0.027
7 cat("z-score for Ted Williams is ", z1)
8 cat("Percentile score for Ted Williams is", pnorm(z1
   ,lower.tail = T))
9
10 # z score for George Brett is (0.390-0.358)/0.027
11 z2 <- (0.390-0.358)/0.027
12 cat("z score for George Brett is ", z2)
13 cat("Percentile score for George Brett is", pnorm(z2,
   lower.tail = T))
14
15
16 cat("The percentile score for Ted Williams is 0.96
   while that for George Brett is 0.88. Both the
   performances are outstanding; however, Ted
   Williams did slightly better than George Brett. "
   )

```

R code Exa 6.17.b probability

```

1
2 #b
3
4 cat(" The chance of the League leader hitting over
   0.400 in a given year can be ",
5     pnorm(0.400,0.358,0.027,lower.tail = F))

```

R code Exa 6.18 octane rating

```
1 octane =c(88.5 ,95.6 ,88.3 ,94.2 ,89.2, 93.3 ,89.8
  ,91.8 ,90.4, 92.2 ,87.7, 93.3 ,87.6 ,92.7, 88.3
  ,91.8 ,89.6, 91.6, 89.3, 92.2, 83.4, 94.7 ,84.3
  ,93.2 ,85.3, 92.3, 87.4, 90.4, 89.7, 91.2, 86.7,
  91.1 ,86.7, 91.0, 87.9 ,90.4, 88.9 ,91.1, 90.3
  ,91.0, 87.5, 91.0, 88.2, 90.3 ,88.6, 90.1 ,91.2,
  92.6, 91.6 ,92.2 ,91.5, 94.2, 90.8, 93.4 ,90.9
  ,93.0, 89.3 ,89.8, 90.5, 90.0, 88.6, 87.8, 88.3
  ,88.5 ,89.0 ,88.7 ,94.4, 90.6 ,90.7 ,100.3 ,93.7,
  89.9, 98.8 ,90.1, 96.1 ,89.9, 92.7 ,91.1 ,92.7)
2 summary(octane)
3 p_value =pnorm(96,mean=mean(octane),sd = sd(octane),
  lower.tail = FALSE)
4 cat("P(octane rating) > 96 =",p_value)
5
6 hist(octane)
```

R code Exa 6.20 QQ plot

```
1 percentile =c(5,10,20,30,40,50,60,70,80,90,95,99.9)
2 z_score =(qnorm(percentile /100))
3
4 florida =c
  (180.448,198.383,220.102,235.762,249.11,261.651,278.539,303.200,3
5
6 qqnorm(florida,ylim = c(150,500),xlim =c(-3,3))
7 qqline(florida)
8
9 nebraska =c
  (165.275,178.736,195.036,206.79,216.832,226.219,235.606,245.649,2
10
```

```

11 qqnorm(nebraska,ylim = c(100,350),xlim=c(-3,3))
12 qqline(nebraska)

```

R code Exa 6.22 CO z score

```

1 CO =c( 1.7, 1.8, 2.1, 2.4, 2.4, 3.4, 3.5, 4.1, 4.2,
        4.4, 4.9, 5.1, 8.3, 9.3, 9.5)
2 qqnorm(CO)
3
4 i =1:15
5 z =i/(15+1)
6 z_score =qnorm(z)

```

R code Exa 6.23 lognormal distribution

```

1
2 # Given X = beryllium contamination follows
  lognormal distribution
3 # Parameters of X ,mu = -2.291 , sd = 1.276
4
5 # Using lognormal distribution
6 #a
7 cat(" the probability that wipe sample will have
    beryllium contamination exceeding 0.50 gm/100 cm2
    = P(X>0.50) "
8     , 1 - plnorm(0.50,meanlog = -2.291,sdlog =
        1.276))
9
10 #b
11 cat("The 95th percentile for contamination
    distribution is", qlnorm(0.95,meanlog = -2.291,
        sdlog = 1.276))

```



```

12 cat("The result exceeds the DOE safety limit of 0.20
    and thus, we can conclude that the beryllium
    contamination at this smelter is at an unhealthy
    level for workers")
13
14
15 #c
16 cat("E(X)= ", exp(-2.291 + (1.276^2)/2))
17 cat("V(X)= ", (exp((2*(-2.291)) + (1.276^2)))*((exp
    (1.276^2)) - 1 ))

```

R code Exa 6.24 beta distribution

```

1
2 # Let X denote the proportion of the total supply
    sold in a given week
3 # Given alpha = 4, beta = 2
4
5 # Using beta distribution
6
7 #a
8 cat(" The expected proportion of supply sold in a
    given week is ", 4/(4+2))
9
10 #b
11 cat("P(X>0.90) = 1 - P(X<=0.90) =", 1 - pbeta
    (0.90,4,2))
12 cat("It is not very likely that 90% of the stock
    will be sold in a given week. ")

```

R code Exa 6.25 Relative frequency histograms and densities

```

1 lifetime_sqroot =c(0.637, 1.531, 0.733, 2.256, 2.364
  ,1.601, 0.152 ,1.826 ,1.868, 1.126, 0.828
  ,1.184, 0.484 ,1.207, 0.719, 0.715 ,0.474 ,1.525,
  1.709, 1.305, 2.186, 1.228, 2.006 ,1.032, 1.802
  ,1.668 ,1.230, 0.577, 1.274, 1.623 ,1.313 ,0.542,
  1.823, 0.880 ,1.526, 2.535, 1.793 ,2.741, 0.578,
  1.360 ,2.868, 1.493 ,1.709, 0.872, 1.032, 0.914
  ,1.952 ,0.984 ,2.119, 0.431)
2
3 #Exponential density plot
4
5 hist(lifetime_sqroot^2,breaks = 9,probability = TRUE
  ,main = "Histogram of Lifetime",xlab = "Lifetime"
  ,ylab = "Probability")
6
7 # Weibull distribution plot differs from the one
  given in textbook
8 #Weibull distribution
9
10 hist(lifetime_sqroot,breaks = 12,probability = TRUE,
  main = "Histogram of Sq root of Lifetime",xlab =
  "sq root",ylab = "Probability")

```

R code Exa 6.26 Weibull Distribution

```

1
2 # USIng Weibull Distribution
3
4 Theeta =50
5 gama = 2
6
7 #a
8 cat("P(X>10) = ", 1 - pweibull(10,shape = gama,scale
  = sqrt(Theeta)))
9

```

```

10 #b
11 cat(" Expected lifetime of thermisters is E(X) =", (
    Theeta^(1/gama))*(gamma(1+(1/gama))))

```

R code Exa 6.27 Weibull Distribution

```

1 cables =c
    (5300,5410,5500,5700,5710,5750,5800,5810,5870,5900,5900,5990,6050

2

3 L =c
    (-3.481,-2.772,-2.351,-2.046,-1.806,-1.606,-1.434,-1.281,-1.144,-

4 fit  =lm(L~log(cables))
5 plot(log(cables),L,main = "Plot of LF( x ) versus ln
    ( x ) for the cable strength data")
6 abline(reg = fit)

```

Chapter 7

Multivariate Probability Distributions

R code Exa 7.2 joint probability distribution of X_1 and X_2

```
1
2 # Joint Probability Distribution
3 # a
4 library(MASS)
5 # formula for Joint Probability Distribution
6
7 a <- function(x,y)
8 {
9   if(x+y <= 2)
10   {
11     if(x==1 || y==1)
12     {
13       ans = ( 2 * (1/3) * (1/3))
14
15     }
16     else{
17       ans = ((1/3) * (1/3))
18     }
19   }
```

```

20
21     else{
22         ans = (0)
23     }
24 }
25
26 for(i in 0:2){
27     for(j in 0:2)
28     {
29         cat("p(", i, ", ", j, ") =")
30         print(fractions(a(i,j)))
31     }
32 }
33 }
34
35 # b
36 # the probability that one of the customers visits
    counter B given that one of the customers is
    known to have
37 # visited counter A.
38 cat("P(X2 =1 | X1 =1) =", a(1,1)/(a(1,0) + a(1,1) +
    a(1,2)))

```

R code Exa 7.3 Joint probability distribution of imprivities

```

1
2 # Joint Density Function of Continuous Random
    Variable..
3
4 a <- function(x1,x2){
5     2 * (1-x1)
6
7 }
8
9

```

```

10 # to find the probability of a region A: {(x1,x2) |
    0<x1<0.5, 0.4<x2<0.7}
11 cat("The probability is", " ")
12
13 integrate(function(x2){
14
15     sapply(x2,function(x2){
16
17         integrate(function(x1)a(x1,x2),0,0.5)$value
18     })
19 },0.4,0.7)

```

R code Exa 7.4 marginal probability density functions

```

1
2 a <- function(x1,x2){
3     2 * (1-x1)
4 }
5 cat("f1(x1) = 2(1-x1)")
6 cat("Since , f2(x2) =", integrate(a,0,1)$value," ,
    our conjecture is verified")

```

R code Exa 7.5 The marginal probability density functions

```

1
2 a <- function(x1){
3     3 * x1
4 }
5
6 b <- function(x2){
7     1.5 * (1-(x2*x2))
8 }
9

```

```

10 cat("the probability that X2 will be between 0.2 and
      0.4 for a given week = ", integrate(b,0.2,0.4)$
      value)

```

R code Exa 7.6 conditional probability

```

1
2 # using conditional probability distribution
3
4 f <- function(x,y){
5   3*x
6 }
7
8 f1 <- function(x){
9   3*x*x
10  }
11
12
13 d <- function(y){
14
15   f(0.5,y)/f1(0.5)
16
17   }
18
19
20 cat("The value of conditional Probability  $P(0 < X_2$ 
       $< 0.2 | X_1 = 0.5)$  is ", integrate(Vectorize(d), 0,
      0.2)$value)

```

R code Exa 7.7 conditional probability density function

```

1
2 # using conditional probability distribution

```

```

3
4 a <- function(x,y){
5   0.5
6 }
7
8 b <- function(y){
9   0.5 * y
10 }
11
12 c <- function(x){
13   a(x,1)/b(1)
14 }
15
16 cat(" The probability of interest is ", integrate(
17   Vectorize(c),0,0.5)$value)
18 # if the machine had contained 2 gallons at the
19   start of the day
20 d <- function(x){
21   a(x,2)/b(2)
22 }
23 cat(" The probability of interest is ", integrate(
24   Vectorize(d),0,0.5)$value)

```

R code Exa 7.10 covariance between two random variables

```

1
2 # to find the covariance between two variables X and
3   Y
4 rm("c")
5
6 a <- matrix(c(0.04,0,0,0.16,0.10,0,0.20,0.30,0.20),
7   nrow = 3, ncol = 3)

```



```

7
8
9 # individual column sum
10 ax <- apply(a,2,sum)
11
12
13 # indivdual row sum
14 ay <- apply(a,1,sum)
15
16 # E[X]..
17 ex <- sum(0:2*ax)
18
19 # E[Y]..
20 ey <- sum(0:2*ay)
21
22
23 # E(XY) ..
24 exy <- 0
25
26 for(i in 0:2){
27
28     for(j in 0:2){
29
30         exy <- exy + i*j*a[i+1,j+1]
31
32     }
33 }
34 Cov <- exy - ex*ey
35 cat("CoV(X1,X2 =)", Cov)
36
37 df<- function(l,m)
38 {
39     (1-m) * (1-m)
40 }
41 cat("V(X1) = " , sum(df((0:2),ey)*ay))
42 cat("V(X2) =" , sum(df((0:2),ex)*ax))
43
44 V1 <- sum(df((0:2),ey)*ay)

```

```

45 V2 <- sum(df((0:2),ex)*ax)
46
47 Ro <- Cov/(sqrt(V1 * V2))
48
49 cat(" Correlation = ", Ro)

```

R code Exa 7.12 mean and variance

```

1 a <- function(x)
2 {
3   x*(3*x*x)
4 }
5 EX1 =integrate(a,0,1)$value
6
7 b <- function(y)
8 {
9   y*1.5*(1- y^2)
10 }
11 EX2 =integrate(b,0,1)$value
12
13 c <- function(x)
14 {
15   x*x*(3*x*x)
16 }
17 EX1sq =integrate(c,0,1)$value
18
19 d <-function(y)
20 {
21   y*y*1.5*(1- y^2)
22 }
23 EX2sq =integrate(d,0,1)$value
24
25 VX1 = EX1sq - EX1^2
26 VX2 = EX2sq - EX2^2
27

```

```

28 e <- function(x)
29 {
30   1.5*(x^4)
31 }
32 EX1X2 =integrate(e,0,1)$value
33
34 Cov = EX1X2 - EX1*EX2
35
36 cat("E(Y) = ",EX1 - EX2)
37 cat("V(Y) = ",VX1 + VX2 + 2*1*(-1)*Cov)

```

R code Exa 7.13 mean and variance of the total weekly amount

```

1 # The dollar amount spent per week is given by  $Y = 3X_1 + 5X_2$ 
2 # given ,  $E(X_1) = 40$  ,  $V(X_1) = 4$ 
3 #  $E(X_2) = 65$  ,  $V(X_2) = 8$ 
4
5 cat("E(Y) =  $3E(X_1) + 5(E(X_2))$  :", 3*40 + 5*65)
6 cat("V(Y) =  $(3^2)*V(X_1) + (5^2)*V(X_2)$  :", (3^2)*4 +
      (5^2)*8)

```

R code Exa 7.15 multinomial probability distribution

```

1
2 # to find the probability that three bulbs have no
   defects , one has a type A defect , and two have
   type B defects out of 6
3 # bulbs chosen from a lot
4
5 # Using multinomial distribution
6
7 p_bulb <- c(0.70,0.20,0.10)

```

```

8
9 # No. of bulbs chosen without defect , with defect A,
   with defect B are:
10 x <- c(3,1,2)
11
12 cat(" the probability that three bulbs have no
      defects , one has a type A defect , and two have
      type B defects is",
13      dmultinom(x,6,p_bulb))

```

R code Exa 7.19 conditional expectation

```

1
2 a <- function(x1,x2){
3   0.5
4 }
5
6 b <- function(x2){
7   0.5 * x2
8 }
9
10 c <- function(x1){
11   x1 * a(x1,1)/b(1)
12 }
13
14 cat(" the conditional expectation of amount of
      sales X1 given that X2 = 1 is ", integrate(
      Vectorize(c),0,1)$value)

```

R code Exa 7.20 conditional distribution of X1 given X2

```

1
2 # E(Y) = E(E(Y|p))

```

```

3 # using binomial distribution
4 #  $E(Y|p) = E(n \cdot p) = n \cdot E(p)$ 
5 # given , n=10
6
7 a <- function(p)
8 {
9     10 * (4*p)
10 }
11
12 cat(" the expected value of Y for any given day",
      integrate(a,0,0.25)$value)

```

Chapter 8

Statistics Sampling Distributions and Control Charts

R code Exa 8.2 central limit theorem

```
1
2 # given , mean =10, sd=10
3
4 # area for the region  $P(a \leq X \leq b) = 0.95$ 
5 # area for the region  $P(X \leq b)$ 
6 x <- (1-0.95)/2 +0.95
7
8 ll <- qnorm(x,lower.tail = FALSE)
9 ul <- qnorm(x,lower.tail = T)
10 cat("Therefore , Z lies between", ll," to ",ul)
11
12 y =function(n){
13   a <- ( 10 + (ll*(10/sqrt(n))))
14   b <- ( 10 + (ul*(10/sqrt(n))))
15   output <- list(a,b)
16   return(output)
17 }
```

```

18
19 output1 <- y(25)
20 cat("interval when n=25 is ")
21 cat("lower limit")
22 output1[1]
23 cat("Upper limit")
24 output1[2]
25
26
27 output2 <- y(50)
28 cat("interval when n=50 is ")
29 cat("lower limit")
30 output2[1]
31 cat("Upper limit")
32 output2[2]
33
34
35 output3 <- y(100)
36 cat("interval when n=100 is ")
37 cat("lower limit")
38 output3[1]
39 cat("Upper limit")
40 output3[2]

```

R code Exa 8.3 central limit theorem for average fracture strength

```

1 # given , sigma =2 , n=100
2
3 #a
4 # according to central limit theorem
5 sigma =2
6 n = 100
7 sd = sigma/sqrt(n)
8 cat("the probability that the average fracture
    strength of glass exceeds 14.5 is",pnorm(14.5,14,

```

```

      sd,lower.tail = F),
9      "which is very small")
10
11 #b
12 x <- (1-0.95)/2 +0.95
13
14 ll <- qnorm(x,lower.tail = F)
15 ul <- qnorm(x,lower.tail = T)
16 cat("The limit is a = ", 14+ ll*sd,"to b = ",14 + ul
      *sd)

```

R code Exa 8.4 Probability of sample mean

```

1 #given sigma=1, n=25
2 sigma =1
3
4 #a
5 n = 25
6 sd <= sigma/sqrt(n)
7
8 cat(" the probability that the sample mean will be
      within 0.3 ounces of the true population mean is"
9      ,
      pnorm(.3/sd) - pnorm(-0.3/sd))
10
11 #b
12
13 # given ,  $P(-0.3 < |X - \mu| < 0.3) = 0.95$ 
14 x <- (1-0.95)/2 +0.95
15
16 ll <- qnorm(x,lower.tail = F)
17 ul <- qnorm(x,lower.tail = T)
18
19 n<- (ul/0.3)^2
20 cat("Value of n so that the sample mean will be

```


within 0.3 ounces of the population mean with probability 0.95 is ",n)

R code Exa 8.5 Normal distribution of failure strengths

```
1
2 # Application of t distribution
3 #a
4 xbar1 <- 2000
5
6 mu <- 3000
7
8 sd <- 989
9
10 n <- 12
11
12 t_value = (xbar1-mu)/(sd/sqrt(n))
13 cat("The required probability is ", pt(t_value, df =
      n-1))
14
15 #b
16 xbar2 <- 2500
17
18 t_value = (xbar2-mu)/(sd/sqrt(n))
19 cat("The required probability is ", pt(t_value, df =
      n-1))
```

R code Exa 8.6 The Sampling Distribution of large samples

```
1 Y <- 12
2 p <- 0.2
3 n <- 100
4 mu <- p
```

```

5 sd <- sqrt(p*(1-p)/n)
6
7
8 z_value = (Y/n - mu)/sd
9 cat("The required probability is ", pnorm(z_value))
10 cat("There is only a small probability of" ,pnorm(z_
    value) , " of accepting any lot that has 20%
    nonconforming wafers. ")

```

R code Exa 8.7 The Sampling Distribution of S²

```

1 # USing chi-square distribution
2 # give , vraiance =0.8,n =10
3 var=0.80
4 n =10
5 ll = 0.05
6 ul = 0.95
7 a= var*qchisq(0.05,df=9)/(n-1)
8 b= var*qchisq(0.95,df=9)/(n-1)
9 cat("value of a and b such that the sample variance
    of the amounts dispensed will be between a and b
    with probability 0.90 is",
10     a,"-",b )

```

R code Exa 8.8 The Sampling Distribution of S²

```

1 # using chi-sq distribution
2
3 # given , variance =100 , n=25
4
5 var =100
6 n=25
7

```

```

8 #a
9 #P(S^2 > 50)
10 a <- (n-1)*50/var
11 cat("approximate probability = ", pchisq(a,df=n-1,
      lower.tail = F))
12
13 #b
14 #P(S^2 > 150)
15 b <- (n-1)*150/var
16 cat("approximate probability = ", pchisq(b,df=n-1,
      lower.tail = F))
17
18 #c
19 cat("E(S^2) = ", var)
20 cat("V(S^2) = ", 2*var*var/(n-1))

```

R code Exa 8.9 General distribution large samples

```

1 rm("c")
2 #Sampling Distribution of the difference of 2 means
3 ..
4 # Mean, variance and no. of samples for both
  machines are as follows
5
6 A <- c(1,200,25)
7 B <- c(1,200,25)
8
9 diff_mean <- A[1]-B[1]
10
11 diff_sd <- sqrt((A[2]/A[3])+(B[2]/B[3]))
12
13 cat(" the probability that the difference in sample
      means for two machines will be at most 10 ml is "
      ,

```

```

14      pnorm(10,diff_mean,diff_sd) - pnorm(-10,diff_
      mean,diff_sd) )

```

R code Exa 8.10 Small samples case equal variances

```

1
2 #Sampling Distribution of the difference of 2 means
3   ..
4 # Mean, variance and no. of samples for both groups
5   are as follows
6
7 A <- c(450,17.795,6)
8 B <- c(250,9.129,4)
9
10 diff_mean <- A[1] - B[1]
11
12 diff_sd <- sqrt(((A[3]-1)*A[2]*A[2] + (B[3]-1)*B[2]*
13   B[2])) / (A[3]-1 + B[3]-1))
14
15 # P(X1-X2 >= 150)
16 x <- (150 - diff_mean)/sqrt(diff_sd*diff_sd*(1/A[3]
17   + 1/B[3]))
18
19 #degree of freedom = 6+4-2
20 cat(" probability that the sample mean tensile proof
21   stress for group 1 is at least 150 MPa larger
22   than that for group 2 is ",pt(x,df=8,lower.tail =
23   F))

```

R code Exa 8.11 The sampling distribution of \bar{X} Dbar

```

1 A <- c(10.18,12.19,12)

```

```

2 B <- c(-14.71,4.40,12)
3 mu <- 0
4 #a
5 #P(d>=10.18)
6 a <- (A[1] - mu)/(A[2]/sqrt(A[3]))
7
8 cat("probability of observing a mean difference of a
    At least 10.18 kN for EOF measurements is",pt(a,
    df=A[3]-1,lower.tail = F))
9
10 #b
11 #P(d<= -14.71)
12 b <- (B[1] - mu)/(B[2]/sqrt(B[3]))
13 cat("probability of observing a mean difference of a
    At most -14.71 kN for EOF measurements is", pt(b
    ,df=B[3]-1,lower.tail = T))

```

R code Exa 8.12 sampling distribution of p1 minus p2

```

1 p1= 0.04
2 p2= 0.025
3 n1 = 200
4 n2 = 200
5 diff_mean = p1-p2
6 diff_sd = sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)
7
8 cat("P(|p1 - p2|) =", pnorm(0.02,diff_mean,diff_sd,
    lower.tail = FALSE) )
9
10 # There is a fairly high chance (38.88%) of
    observing a difference of at most 2 percentage
    points between the sample proportion defectives
11
12 #Answer given in book is wrong. Answer will be twice
    of what given in book.

```

R code Exa 8.13.a sampling distribution of $S1sq$ divided $S2sq$

```
1 n1 = 10
2 n2 = 10
3
4 # 2 cases:  $s1*s1 > 2*s2*s2$  and  $s2*s2 > 2*s1*s1$ 
5 # Let  $s1^2/s2^2 = X$ 
6
7 # The probability of observing one sample variance
  at least 2 times larger than the other is
8 # $P(X < 0.5) + P(X > 2)$ 
9
10 # Using F distribution
11
12 cat("P(F(9,9) < 0.5) + P(F(9,9) > 2) = ", pf(0.5,9,9) +
    pf(2,9,9,lower.tail = F))
13
14 cat(" There is approximately ", (pf(0.5,9,9) + pf
    (2,9,9,lower.tail = F))*100, " chance that one
    sample variance will be at least 2 times larger
    than the other
15 , even if the population variances are equal.")
```

R code Exa 8.13.b probability of sampling distribution

```
1 n1 = 10
2 n2 = 10
3
4 # 2 cases:  $s1*s1 > 4*s2*s2$  and  $s2*s2 > 4*s1*s1$ 
5 # Let  $s1^2/s2^2 = X$ 
6
```

```

7 # The probability of observing one sample variance
  at least 2 times larger than the other is
8 #P(X<0.25) + P(X>4)
9
10 # Using F distribution
11
12 cat("P(F(9,9) < 0.25) + P(F(9,9) > 4) = ", pf(0.25,9,9)
    + pf(4,9,9,lower.tail = F))
13
14 cat(" There is approximately ", (pf(0.25,9,9) + pf
    (4,9,9,lower.tail = F))*100, " chance that one
    sample variance will be at least 2 times larger
    than the other
15     , even if the population variances are equal.")

```

R code Exa 8.14 Xbar and R charts

```

1 m1 =c
    (16.1,16.2,16.0,16.1,16.5,16.8,16.1,15.9,15.7,16.2,16.4,16.5,16.7
2 m2 =c
    (16.2,16.4,16.1,16.2,16.1,15.9,16.9,16.2,16.7,16.9,16.9,16.9,16.2
3 m3 =c
    (15.9,15.8,15.7,15.9,16.4,16.1,16.2,16.8,16.1,16.1,17.1,17.2,16.4
4 m4 =c
    (16.0,16.1,16.3,16.4,16.4,16.3,16.5,16.1,16.4,17.0,16.2,16.1,15.8
5 m5 =c
    (16.1,16.2,16.1,16.6,16.2,16.4,16.5,16.4,16.8,16.4,16.1,16.4,16.6
6
7 obseravation <- c(m1,m2,m3,m4,m5)
8

```

```

9  samples <- c(rep(1:20,5))
10
11
12  dat <- data.frame(observavation,samples)
13
14
15  print("The xbar and S chart for the above data is:")
16
17  #install the package qicharts for xbar chart
18
19  library(qicharts)
20  # Run the below two code individually..
21  #xbar chart
22  qic(observavation,
23      x= samples,
24      data = dat,
25      chart = 'xbar',
26      xlab = 'Sample Number')
27
28  #install the package qcc for R chart
29
30  # R chart
31  library(qcc)
32  dat1=data.frame(m1,m2,m3,m4,m5)
33  qcc(dat1,type = "R")

```

R code Exa 8.15 X bar and S charts

```

1  m1 =c
    (16.1,16.2,16.0,16.1,16.5,16.8,16.1,15.9,15.7,16.2,16.4,16.5,16.7
2  m2 =c
    (16.2,16.4,16.1,16.2,16.1,15.9,16.9,16.2,16.7,16.9,16.9,16.9,16.2
3  m3 =c

```



```

(15.9,15.8,15.7,15.9,16.4,16.1,16.2,16.8,16.1,16.1,17.1,17.2,16.4
4  m4 =c
    (16.0,16.1,16.3,16.4,16.4,16.3,16.5,16.1,16.4,17.0,16.2,16.1,15.8
5  m5 =c
    (16.1,16.2,16.1,16.6,16.2,16.4,16.5,16.4,16.8,16.4,16.1,16.4,16.6

6
7  obseravation <- c(m1,m2,m3,m4,m5)
8
9  samples <- c(rep(1:20,5))
10
11
12  dat <- data.frame(obseravation,samples)
13
14
15  print("The xbar and S chart for the above data is:")
16
17
18  #Use the package qicharts
19
20  library(qicharts)
21  # Run the below two code individually..
22  #xbar chart
23  qic(obseravation,
24      x= samples,
25      data = dat,
26      chart = 'xbar',
27      xlab = 'Sample Number')
28
29  # S chart
30  qic(obseravation,
31      x = samples,
32      chart = 's',
33      xlab = 'Sample Number',
34      data = dat)
35

```

36 #UCL and LCL values have been rounded off.

R code Exa 8.16 p chart

```
1 rm("c")
2 x =c
   (3,1,4,2,0,2,3,3,5,4,1,1,1,2,0,3,2,2,4,1,3,0,2,3)
3
4 sample =1:24
5 dat <- data.frame(sample,x)
6
7 p <- mean(dat$x/50)
8
9
10 u <- p + 3*sqrt(p*(1-p)/50)
11
12 l <- p- 3*sqrt(p*(1-p)/50)
13
14 cat("The LCL and UCL are",0,"and", u,"respectively")
15
16 #Since l is neg. , we take lower limit to be 0.
17
18 #install the package qcc
19
20 library(qcc)
21 qcc(dat$x, sizes =50,type="p")
```

R code Exa 8.17 c chart

```
1 defective =c
   (6,3,4,0,2,7,3,1,0,0,4,3,2,2,6,5,0,7,2,1)
2 sample =1:20
3 dat <- data.frame(sample,defective)
```

```

4 n=20
5 c =sum(defective)/n
6 u =c + 3*sqrt(c)
7 l =c - 3*sqrt(c)
8
9 cat("The LCL and UCL are",0,"and", u," respectively")
10 #Since l is neg. , we take lower limit to be 0.
11
12 # install the package qcc.
13 library(qcc)
14 qcc(dat$defective, type = "c")

```

R code Exa 8.18 u chart

```

1 defect =c(1,4,1,2,1,4,3,5,3,1,2,1)
2 hours =c
      (58.33,80.22,209.24,164.70,253.70,426.90,380.20,527.70,319.30,340.
3 part =1:12
4 dat = data.frame(part,defect,hours)
5 u_bar= sum(dat$defect)/sum(dat$hours)
6
7 ucl1=u_bar + 3*sqrt(u_bar/dat$hours[1])
8 ucl2=u_bar + 3*sqrt(u_bar/dat$hours[2])
9
10 u1 =dat$defect[1]/dat$hours[1]
11 u2 =dat$defect[2]/dat$hours[2]
12
13
14 #install the package qicharts for u chart
15
16 library(qicharts)
17 # Run the below code ..
18 # u chart
19 qic(defect,hours,

```

```

20     x= part      ,
21     data = dat ,
22     chart = 'u' ,
23     xlab = 'Sample Number')

```

R code Exa 8.19 total proportion out of specification

```

1  m1 =c
    (16.1,16.2,16.0,16.1,16.5,16.8,16.1,15.9,15.7,16.2,16.4,16.5,16.7
2  m2 =c
    (16.2,16.4,16.1,16.2,16.1,15.9,16.9,16.2,16.7,16.9,16.9,16.9,16.2
3  m3 =c
    (15.9,15.8,15.7,15.9,16.4,16.1,16.2,16.8,16.1,16.1,17.1,17.2,16.4
4  m4 =c
    (16.0,16.1,16.3,16.4,16.4,16.3,16.5,16.1,16.4,17.0,16.2,16.1,15.8
5  m5 =c
    (16.1,16.2,16.1,16.6,16.2,16.4,16.5,16.4,16.8,16.4,16.1,16.4,16.6
6
7  obseravation <- c(m1,m2,m3,m4,m5)
8  m =mean(obseravation)
9  sigma1 =0.361
10 sigma2 =0.367
11 USL =17
12 LSL =16
13 zUSL = (USL - m)/sigma1
14 zLSL = (m - LSL)/sigma2
15 zmin = min(zUSL,zLSL)
16 Cpk =zmin/3
17 cat("The area below zLSL is",1- pnorm(zLSL))
18 cat("The area above zUSL is",1- pnorm(zUSL) )

```

```
19 prop= 1- pnorm(zLSL) + 1- pnorm(zUSL)
20 cat("proportion out of specification =",prop)
```

Chapter 9

Estimate

R code Exa 9.4 Large Sample Confidence Interval for a Mean

```
1 #Confidence interval evaluation
2 #For confidence coefficient to be 95%, , it leaves
   (1 - 0.025 = 0.975) area to the left
3
4 n <- 50
5
6 mu <- 2.268
7
8 sig <- 1.932
9
10 a <- qnorm(0.975)*sig/sqrt(n)
11
12
13 cat("The 95% confidence interval for the above
      parameters is",mu-a,mu+a," hours")
```

R code Exa 9.6 Determining Sample Size to Estimate Mean

```

1 # Determining Sample Size to Estimate Mean...
2
3 sigma =2
4 B = 0.1
5
6 cat("Thus, at least ", round((qnorm(0.975)*sigma/B)
  ^2) ,"employees should be sampled to achieve the
  desired results. ")

```

R code Exa 9.7 lower limit for the mean lifetime of batteries

```

1 # givn confidence level =95%, we'll calculate z(1-
  0.95)
2
3 n <- 50
4
5 mu <- 2.266
6
7 sig <- 1.935
8
9 a <- qnorm(0.05, lower.tail = F)
10
11 # To calculate lower limit
12 cat("Lower limit is ", mu - a*sig/sqrt(n) )

```

R code Exa 9.8 Confidence Interval for a Mean Based on t distribution

```

1
2 # Using T distribution
3
4 uts <- c(253,261,258,255,256)
5
6 mu = mean(uts)

```

```

7 sig = sd(uts)
8 n = 5
9 alpha = 1 - 0.95
10
11 a <- qt(alpha/2, df= n-1)
12
13 cat(" Therefore , the interval is ", mu + a*sig/sqrt(n
    ), " - ", mu - a*sig/sqrt(n))

```

R code Exa 9.9 Large Sample Confidence Interval for a Proportion

```

1 # confidence interval =90%
2
3 alpha = 1-0.90
4 p =0.20
5 n =100
6 a =qnorm(1 - alpha/2,lower.tail = F)
7
8 cat(" the true probability p of finding this
    microorganism in a sample is somewhere between ",
    p - a*sqrt(p*(1-p)/n),
9     " - ", p + a*sqrt(p*(1-p)/n))

```

R code Exa 9.10 number of workers

```

1 #Determining sample size
2 #Given , confidence limit =90%
3 #Because no prior knowledge of p is available , use p
  = 0.5
4
5 alpha=1-0.90
6 B =0.05
7 p=0.5

```



```

8 a <- qnorm(alpha/2, lower.tail = F)
9
10 cat("A random sample of at least ", round(((a/B)^2)*p
      *(1-p)), " workers is required in order to
      estimate the true proportion favoring
11 the revised policy to within 0.05. ")

```

R code Exa 9.11 Confidence Interval for a Variance

```

1 readings <- c(9.54, 9.61, 9.32, 9.48, 9.70, 9.26)
2
3 var= (sd(readings))^2
4 n=6
5
6 # given Confidence interval =90%
7 alpha= 1-0.90
8
9 # Using chi-sq distribution
10
11 a <- qchisq(1 - (alpha/2), df= n-1, lower.tail = F)
12 b <- qchisq((alpha/2), df= n-1, lower.tail = F)
13
14 cat("Thus, confidence interval is ", (n-1)*var/b, " -
      ", (n-1)*var/a)

```

R code Exa 9.12 Confidence Interval for a Difference in Means

```

1
2 # Mean , variance and no. of observations for both
  machines are as follows:
3 A <- c(12, 6, 100)
4 B <- c(9, 4, 100)
5

```

```

6 # given confidence interval =90%
7 alpha = 1-0.90
8 diff_mean <- A[1] - B[1]
9
10 diff_sd <- sqrt(A[2]/A[3] + B[2]/B[3])
11
12
13 a = qnorm(alpha/2, lower.tail = F)
14 cat("We are about 90% confident that the difference
    in mean daily downtimes is between, ", diff_mean
    - a*diff_sd," - ", diff_mean + a*diff_sd," min")

```

R code Exa 9.13 Confidence Interval for a Linear Function of Means

```

1 # Mean , variance and no. of observations for 3
  machines are as follows:
2
3 A <- c(12,6,100)
4 B <- c(9,4,100)
5 C <- c(14,5,100)
6
7 #Expected daily cost for downtime on 3 machines is 3
  *mu1 + 5*mu2 + 2*mu3
8
9 mu = 3*A[1] + 5*B[1] + 2*C[1]
10 cat("The estimated daily cost is ",mu )
11
12 var = 9*A[2]/A[3] + 25*B[2]/B[3] + 4*C[2]/C[3]
13 cat("Estimated variance is ", var)
14
15 # Confidence interval =95%
16 alpha=1-0.95
17 z = qnorm(1 - alpha/2)
18
19 cat(" We are 95% confident that the mean daily cost

```

```

20      of downtimes on these machines is between $",
      mu - z*sqrt(var)," and $",mu + z*sqrt(var))

```

R code Exa 9.14 Normal Distributions with Common Variance

```

1  # Mean , variance and no. of observations for 2
   batches are as follows:
2
3  A <- c(0.22,0.0010,4)
4  B <- c(0.17,0.0020,5)
5
6  diff_mean= A[1]-B[1]
7
8  common_var = ((A[3]-1)*A[2] + (B[3]-1)*B[2]) / (A[3]+
   B[3]-2)
9  sigma = sqrt(common_var)
10
11 # Confidence interval =95%
12 alpha=1-0.95
13 t = qt(alpha/2, df=A[3]+B[3]-2,lower.tail = F)
14
15 c = t*sigma*sqrt(1/A[3]+ 1/B[3])
16
17 cat("Thus, we are 95% confident that the difference
   in the mean porosity measurements for two batches
   is between ",
18     diff_mean - c," and ", diff_mean +c)

```

R code Exa 9.15 confidence interval for difference in mean denier

```

1
2 #Normal Distributions with Common Variance
3

```

```

4
5 M1 <- c
    (9.17,12.85,5.16,6.37,6.64,8.42,7.33,8.91,9.45,11.39,10.90,6.34,1

6 A <- c(mean(M1),sd(M1),length(M1))
7
8 M2 <- c
    (18.86,8.86,17.11,17.38,9.38,11.64,11.25,15.00,12.77,18.89,16.88,

9 B <- c(mean(M2),sd(M2),length(M2))
10
11 diff_mean= A[1]-B[1]
12
13
14 common_var = ((A[3]-1)*A[2]*A[2] + (B[3]-1)*B[2]*B
    [2])/ (A[3]+B[3]-2)
15 sigma = sqrt(common_var)
16
17
18 # Confidence interval =95%
19 alpha=1-0.95
20 t = qt( alpha/2, df=A[3]+B[3]-2,lower.tail = F)
21
22
23 c = t*sigma*sqrt(1/A[3]+ 1/B[3])
24
25 cat("Thus, we are 95% confident that the difference
    in the mean denier is between ",
26     diff_mean - c," and ", diff_mean +c)

```

R code Exa 9.16 95 percent confidence level

```

1
2 #Normal Distributions with Common Variance
3 rm("c")

```

```

4 previous = c(13.18, 9.42, 10.55, 10.11, 7.28, 8.53,
              7.52, 8.04, 8.34, 6.91, 10.70, 9.21, 7.84, 9.46,
              6.49)
5
6
7 after = c(5.31, 5.77, 3.36, 5.26, 2.43, 6.08, 3.77,
           3.20, 3.49, 3.39, 2.99, 4.79, 6.99, 4.81, 3.99,
           4.41, 7.12, 3.83, 3.57, 5.41)
8
9 t.test(previous,after)

```

R code Exa 9.17 95 percent confidence interval for normal distribution

```

1 # Mean , variance and no. of observations for 3
  batches are as follows:
2
3 A <- c(0.22,0.0010,4)
4 B <- c(0.17,0.0020,5)
5 C <- c(0.12,0.0018,10)
6
7 mu =( A[1]*A[3] + B[1]*B[3])/(A[3]+B[3])
8
9 # Difference between average f 2 batches and the 3rd
  batch
10 diff_mean = mu - C[1]
11
12 diff_sd = sqrt(((A[3]-1)*A[2] + (B[3]-1)*B[2] + (C
  [3]-1)*C[2])/(A[3]+B[3]+C[3]- 3))
13
14 a1 = A[3]/(A[3]+B[3])
15 a2 = B[3]/(A[3]+B[3])
16 a3 = -1
17
18 #Given , confidence interval =95%
19 alpha= 1-0.95

```

```

20 b = qt(alpha/2,df= A[3]+B[3]+C[3]-3,lower.tail = F)*
    diff_sd*sqrt(a1*a1/A[3] + a2*a2/B[3] + a3*a3/C
    [3])
21
22 cat(" we are 95% confident that the difference
    between the mean of the third batch and the
    average
23 of means of the first two batches is between ",
    diff_mean - b ," - ", diff_mean + b)

```

R code Exa 9.18 Normal Distributions with Unequal Variances

```

1 #Normal Distributions with Unequal Variances
2
3 M1 <- c
    (9.17,12.85,5.16,6.37,6.64,8.42,7.33,8.91,9.45,11.39,10.90,6.34,1
4
5
6 M3 <- c
    (12.17,11.22,11.42,11.73,12.33,12.21,12.21,10.93,12.16,11.61,10.4
7
8 t.test(M1,M3)

```

R code Exa 9.19 Two sample T test for Chemical vs Atmospheric

```

1 #Two sample T test for Chemical vs Atmospheric
2
3 chemical <- c
    (2.30143,2.29890,2.29816,2.30182,2.29869,2.29940,2.29849,2.29889,
4

```

```

5
6 atmp <- c
  (2.31017,2.30986,2.31010,2.31001,2.31010,2.31024,2.31028,2.31163,
7
8 t.test(chemical,atmp)

```

R code Exa 9.20 Large Sample Confidence Interval for a Difference in Proportions

```

1
2 #Large Sample Confidence Interval for a Difference
  in Proportions
3 #Data for motors
4
5 n1 = 250
6 y1 = 25
7 p1 = y1/n1
8 n2 = 200
9 y2 = 30
10 p2 = y2/n2
11
12 diff_prop = p1 - p2
13
14 # givem confidence interval =95%
15 aplha = 1-0.95
16 a = qnorm(aplha/2, lower.tail = F)*sqrt(p1*(1-p1)/n1
  + p2*(1-p2)/n2)
17
18
19 cat("We are 95% confident that the true difference
  in proportion of defective motors produced by two
  shifts is between ",diff_prop -a," - ", diff_
  prop+a)

```

R code Exa 9.21 95 percent confidence interval

```
1 p11=0.7
2 p12=0.9
3 p21=0.8
4 p22=0.9
5
6 #Estimated mean
7 p= (p12 - p11) - (p22 - p21)
8
9 #Estimated variance
10 var= sum(p11*(1-p11),p12*(1-p12),p21*(1-p21),p22*(1-
    p22))/100
11
12 # givem confidence interval =95%
13 aplha = 1-0.95
14 a = qnorm(aplha/2, lower.tail = F)*sqrt(var)
15
16 cat("we are 95% confident that the difference in the
    change in probability for males and females is
    between ",p-a," - ",p+a)
```

R code Exa 9.22 Confidence Interval for a Ratio of Population Variances

```
1 #Confidence Interval for a Ratio of Population
  Variances
2
3 n1 =10
4 var1 =2.31
5 n2 =16
6 var2= 3.68
7 x = var2/var1
```



```

8
9 #given , confidence interval=90%
10 alpha=1-0.90
11 cat(" we are 90% confident that the ratio is between
      ",x/qf(alpha/2,n2 -1,n1 -1,lower.tail = F)," - "
      , x*qf(alpha/2,n1 -1,n2-1,lower.tail = F))

```

R code Exa 9.23 A Prediction Interval

```

1 #Prediction interval
2 mu =16.1
3 s=0.01
4 n=16
5
6 #Given, confidence interval =95%
7 alpha=1-0.95
8 x= qt(alpha/2,df = n-1,lower.tail = F)*s*sqrt(1 + 1/
      n)
9 cat("We are about 95% confident that the next
      observation will lie between ", mu-x ," - ", mu+x
      )

```

R code Exa 9.24 Tolerance Intervals

```

1 n=45
2 mu=498
3 s=4
4 delta=0.90
5 alpha=1-0.95
6 cat("For these data, k=2.021")
7 k=2.021

```

```

8 cat(" We are 95% confident that 90% of the
    population resistances in the population lie
    between "
9      ,mu- k*s,"-",mu+ k*s)

```

R code Exa 9.25 the confidence coefficient

```

1
2 # find the confidence coefficient
3
4 delta=0.90
5 n=50
6 a= 1 - n*(delta^(n-1)) + (n-1)*(delta^n)
7 cat("We are ",a*100," confident that the interval
    (2,150, 2,610) contains at least 90% of the
    lifelength measurements for the population under
    study. ")

```

R code Exa 9.29 a 95 percent confidence interval for theeta

```

1
2 #Method of maximum liklihood
3
4 X <- c( 0.406, 2.343, 0.538, 5.088, 5.587, 2.563,
    0.023, 3.334, 3.491, 1.267)
5 mean = mean(X)*10
6
7 #Given confidence interval=95%
8
9 alpha =1-0.95
10 a <- qchisq(alpha/2,df=2*(length(X)),lower.tail = F)
11 b <- qchisq(alpha/2,df=2*(length(X)))
12

```

```
13 cat(" We are about 95% confident that the true mean
    lifelength is between",2*mean/a, "and", 2*mean/b
    )
```

R code Exa 9.30 confidence interval

```
1 data=c(0.406 ,0.685 ,4.778 ,1.725 ,8.223, 2.343
    ,1.401 ,1.507 ,0.294, 2.230, 0.538, 0.234 ,4.025
    ,3.323, 2.920, 5.088 ,1.458, 1.064, 0.774 ,0.761
    ,5.587 ,0.517, 3.246, 2.330 ,1.064 ,2.563 ,0.511
    ,2.782 ,6.426 ,0.836 ,0.023 ,0.225, 1.514 ,3.214
    ,3.810 ,3.334 ,2.325 ,0.333 ,7.514 ,0.968 ,3.491,
    2.921 ,
2      1.624, 0.334, 4.490, 1.267, 1.702, 2.634
    ,1.849 ,0.186)
3 x =mean(data)
4 t=5
5 n=50
6 alpha=0.05
7 z =qnorm(1- alpha/2)
8 u = exp(-t/x) + (z/sqrt(n))*(t/x)*exp(-t/x)
9 l = exp(-t/x) - (z/sqrt(n))*(t/x)*exp(-t/x)
10
11 cat(" we are about 95% confident that the
    probability is between ",l,"and",u)
```

Chapter 10

Hypothesis Testing

R code Exa 10.8 Testing for mean

```
1 #H0 : mu =2
2 #H1 : mu Not = 2
3
4 n= 100
5 mu =2
6 sample_mean=2.005
7 sd = 0.03
8 alpha =0.05
9
10 stat = (sample_mean-mu)/(sd/sqrt(n))
11
12 compare = qnorm(alpha/2,lower.tail = F)
13
14 if(stat<compare){
15   cat("Hypothesis is accepted");
16 } else{
17   cat("Hypothesis is not accepted")
18 }
19 error <- qnorm(0.975)*sd/sqrt(n)
20 cat(" The 95% confidence interval for mu is ",2 -
      error, "- ",2+ error)
```

R code Exa 10.9 Hypothesis testing at 5 percent significance level

```
1 #H0 : mu<=15
2 #H1 : mu>15
3
4 n= 36
5 mu =15
6 sample_mean=17
7 sd = 3
8 alpha =0.05
9
10 stat = (sample_mean-mu)/(sd/sqrt(n))
11
12 compare = qnorm(alpha,lower.tail = F)
13
14 if(stat<compare){
15   cat("Hypothesis is accepted");
16 } else{
17   cat("Hypothesis is rejected")
18 }
```

R code Exa 10.10 Observed Significance Level or p value

```
1 n= 36
2 mu =15
3 sample_mean=17
4 sd = 3
5 alpha =0.05
6
7 stat = (sample_mean-mu)/(sd/sqrt(n))
8 p_value= pnorm(stat,lower.tail = F)
```

```

9 cat("Thus the p-value for this test is ",p_value,"
    and we would reject H0 for any significance level
10    greater than or equal to this p-value")

```

R code Exa 10.11 p value for the situation

```

1 n= 100
2 mu =2
3 sample_mean=2.005
4 sd = 0.03
5 alpha =0.05
6
7 stat = (sample_mean-mu)/(sd/sqrt(n))
8 p_value= pnorm(stat,lower.tail = F)
9 cat("The probability of getting a sample of holes
    created with this drill with a mean diameter
10    that is at least 1.67 standard deviations away
    from the depth setting of 2 inches is ",2*p_
    value)

```

R code Exa 10.12 hypothesis about the population mean

```

1 #H0 : mu<2
2 #H1 : mu>2
3
4 data=c(0.406 ,0.685 ,4.778 ,1.725 ,8.223, 2.343
    ,1.401 ,1.507 ,0.294, 2.230, 0.538, 0.234 ,4.025
    ,3.323, 2.920, 5.088 ,1.458, 1.064, 0.774 ,0.761
    ,5.587 ,0.517, 3.246, 2.330 ,1.064 ,2.563 ,0.511
    ,2.782 ,6.426 ,0.836 ,0.023 ,0.225, 1.514 ,3.214
    ,3.810 ,3.334 ,2.325 ,0.333 ,7.514 ,0.968 ,3.491,
    2.921 ,

```

```

5          1.624, 0.334, 4.490, 1.267, 1.702, 2.634
          ,1.849 ,0.186)
6
7 # install the package DescTools
8
9 library(DescTools)
10 p_value = ZTest(data, alternative = "greater", mu=2, sd_
      pop = sd(data), conf.level = 0.95)$p.value
11 alpha = 0.05
12
13 if(p_value > alpha){
14   cat("Null hypothesis is accepted")
15 }else{
16   cat("Null hypothesis is rejected")
17 }

```

R code Exa 10.13 the probability of a type II error

```

1 #H0 : mu<=100
2 #H1 : mu=103
3
4 n= 30
5 mu =100
6 sd = 4
7 alpha =0.01
8
9 z = qnorm(alpha, lower.tail = F)
10 sample_mean = mu + z*sd/sqrt(n)
11
12 true_mean =103
13 s = sd/sqrt(n)
14
15 #P(X<= sample_mean)
16 p_value = pnorm(sample_mean, true_mean, s)
17 cat("Therefore, the probability of type II error

```

when true avg pressure is 103 =",p_value)

R code Exa 10.14 Determining Sample Size

```
1 mu0 =100
2 mu1= 103
3 sd=4
4 alpha = beta =0.01
5 n = ((qnorm(alpha)+qnorm(beta))^2)*sd*sd/((mu1 - mu0
    )^2)
6 cat("By taking ",ceiling(n),"measurements , we can
    reduce to 0.01 while also holding at 0.01. ")
```

R code Exa 10.15 Testing a Mean Normal Distribution Case

```
1 #H0 : mu=1200
2 #H1 : mu=not=1200
3
4 n= 10
5 mu =1200
6 sample_mean=1290
7 sd = 110
8 alpha =0.05
9
10 t = (sample_mean -mu)/(sd/sqrt(n))
11
12 #using rejection region approach
13 compare = qt(1 - alpha/2,df=n-1,lower.tail = T)
14
15 if(t<compare){
16   cat("Hypothesis is accepted");
17 } else{
18   cat("Hypothesis is rejected")
```


19 }

R code Exa 10.16 Hypothesis about contradicting the manufacturers claim

```
1 #H0 : mu=3000
2 #H1 : mu<3000
3
4 n= 8
5 mu =3000
6 sample_mean=2959
7 sd = 39.4
8 alpha =0.05
9
10 t = (sample_mean -mu)/(sd/sqrt(n))
11
12 compare = qt(alpha,df=n-1)
13
14 if(t>compare){
15   cat("Hypothesis is accepted");
16 } else{
17   cat("Hypothesis is rejected")
18 }
```

R code Exa 10.18 Testing for proportion Large sample case

```
1 #H0 : p<=0.10
2 #H1 : p>0.10
3
4 n=100
5 p_bar=0.15
6 p0=0.10
7 alpha=0.01
8
```

```

9 z=(p_bar-p0)/sqrt(p0*(1-p0)/n)
10 compare= qnorm(alpha,lower.tail = F)
11
12 if(z<compare){
13   cat("Hypothesis is accepted");
14 } else{
15   cat("Hypothesis is rejected")
16 }

```

R code Exa 10.19 Testing for variance Normal distribution case

```

1 #Testing for variance: Normal distribution case
2
3 n=10
4 var=0.0002
5 sample_var=0.0003
6 alpha=0.05
7
8 K= (n-1)*sample_var/var
9 compare= qchisq(alpha,df=n-1,lower.tail = F)
10
11
12 if(K<compare){
13   cat("Hypothesis is accepted");
14 } else{
15   cat("Hypothesis is rejected")
16 }

```

R code Exa 10.20 Testing the difference between two means

```

1 # H0: mu1-mu2=0
2 # H1: mu1-m2 not=0
3

```

```

4 M =c(42,18,50)
5 W =c(38,14,50)
6
7 diff_mean = M[1]-W[1]
8 D0=0
9 alpha=0.05
10
11 z=(diff_mean-D0)/sqrt(M[2]/M[3] + W[2]/W[3])
12
13 #Using rejection region approach
14 compare= qnorm(alpha/2,lower.tail = F)
15
16 if(z<compare){
17   cat("Hypothesis is accepted");
18 } else{
19   cat("Hypothesis is rejected")
20 }

```

R code Exa 10.21 Checking the condition of equal variances

```

1 #H0 : mu1 =mu2
2 #H1 : mu1>mu2
3
4 sm <- c(35.22,24.44,9)
5 nm <- c(31.56,20.03,9)
6
7 diff_mean = sm[1]-nm[1]
8 D0=0
9 alpha=0.05
10
11 s = sqrt(((sm[3]-1)*sm[2] + (nm[3]-1)*nm[2])/(sm
    [3]-1 + nm[3]-1))
12 t = (diff_mean - D0)/(s*sqrt(1/sm[3] + 1/nm[3]))
13
14 compare= qt(alpha,df=sm[3]-1 + nm[3]-1,lower.tail =

```

```

        F)
15
16 if(t<compare){
17   cat("Hypothesis is accepted");
18 } else{
19   cat("Hypothesis is rejected")
20 }

```

R code Exa 10.22 Checking the condition of equal variances

```

1 #H0 : mu2=mu3
2 #H1 : mu2 not= mu3
3 Class2 = c(253, 261, 258, 255, 256 )
4 Class3 = c(274, 275, 271, 277, 256 )
5
6 p_value= t.test(Class2,Class3)$p.value
7 if(p_value>alpha){
8   cat("Hypothesis is accepted");
9 } else{
10   cat("Hypothesis is rejected")
11 }
12
13 # Mean calculated for class III is incorrect and
    therefore gives the wrong answer.

```

R code Exa 10.23 Testing the Difference between 2 Means Unequal Variances Case

```

1 #H0 : mu1 - mu2 =D0
2 #H1 : mu1 - mu2 not=D0
3
4 C1 = c( 5.83, 5.66, 4.75, 3.00, 3.37, 3.63, 4.00,
        4.63, 4.25, 4.13 )

```

```

5 C2 = c( 3.38, 2.81, 7.00, 1.50, 5.88, 5.25, 4.08,
          7.63, 4.50, 4.88 )
6
7 p_value= t.test(C1,C2)$p.value
8 alpha= 0.05
9 if(p_value>alpha)
10 {
11   cat(" Hypothesis  accepted")
12 }else{
13   cat(" Hypothesis  is  rejected")
14 }

```

R code Exa 10.24 Testing the Difference between Means for Paired Samples

```

1 #H0 : $\mu_D \geq 0$ 
2 #H1 : $\mu_D < 0$ 
3
4 A = c
      (38.25,31.68,26.24,41.29,44.81,46.37,35.42,38.41,42.68,46.71,29.2
5
6 B = c
      (38.25,31.71,26.25,41.33,44.80,46.39,35.46,38.42,42.70,46.76,29.1
7
8 p_value= t.test(A,B,paired = TRUE,alternative = "
          less")$p.value
9
10 alpha= 0.05
11 if(p_value>alpha)
12 {
13   cat(" Null Hypothesis  accepted")
14 }else{
15   cat(" Hypothesis  is  rejected")
16 }

```

R code Exa 10.25 Testing the Difference between Means for Paired Samples

```
1 #H0 : muD=0
2 #H1 : muD not=0
3
4 E = c( 2727.6, 2902.6, 2463.1 ,3744.5 ,3855.3,
        3807.3, 3610.1 ,3596.3 ,3457.0 ,3507.1, 3184.2,
        3104.7 )
5 A = c( 2741.0 ,2885.0 ,2476.0 ,3745.0 ,3862.0
        ,3812.0 ,3609.0 ,3568.0 ,3465.0 ,3541.0 ,3213.0,
        3092.0)
6
7 alpha= 0.05
8 p_value=t.test(E,A,paired = TRUE)$p.value
9 if(p_value>alpha)
10 {
11   cat(" Null Hypothesis  accepted")
12 }else{
13   cat(" Hypothesis  is  rejected")
14 }
```

R code Exa 10.26 Testing the ratio of variances Normal distributions case

```
1 #H0 : sigma1^2 = sigma2^2
2 #H1 : sigma1^2 < sigma2^2
3
4
5 n1= 10
6 n2= 20
7 var1=0.003
```

```

8 var2=0.001
9 alpha=0.05
10
11 F = var1/var2
12
13 #Left-tailed test
14 compare= qf(1-alpha,n1 -1,n2 -1,lower.tail = T)
15
16 if(F<compare){
17   cat("Hypothesis is accepted");
18 } else{
19   cat("Hypothesis is rejected")
20 }
21
22
23 #Alternative solution
24
25 p_value = 1-pf(F,n1 -1,n2 -1,lower.tail = T)
26 if(p_value>alpha)
27 {
28   cat("Null Hypothesis accepted")
29 }else{
30   cat("Hypothesis is rejected")
31 }
32
33 #Note: t.test function cannot be used as numeric
    vector of data values is not given.

```

R code Exa 10.27 Testing Parameters of the Multinomial Distribution ChiSquare Test

```

1 #H0 : p1=4/7,p2=2/7,p3=1/7
2 #H1 : The proportions differ from those indicated in
    the null hypothesis.
3

```

```

4 X=c(20,16,14)
5 p=c(4/7,2/7,1/7)
6 n=50
7 EX=n*p
8 alpha=0.05
9
10 stat= sum(((X-EX)^2)/EX)
11 compare= qchisq(1-alpha,df=2,lower.tail = T)
12
13
14 if(stat<compare){
15   cat("Hypothesis is accepted");
16 } else{
17   cat("Hypothesis is rejected")
18 }

```

R code Exa 10.28 Testing Equality among Binomial Parameters ChiSquare Test

```

1
2 #H0 : equal kill rates for the four chemicals
3 #H1 : at least two mixtures have different kill
   rates.
4
5 dead = c(124,147,141,142)
6 not_dead = c(76,53,59,48)
7
8 observed =as.data.frame(rbind(dead,not_dead))
9 names(observed) <- c('Mix1','Mix2','Mix3','Mix4')
10
11 stat= chisq.test(observed)$statistic
12
13 alpha=0.05
14 compare=qchisq(1-alpha,df=3,lower.tail = F)
15

```



```

16 if(stat<compare){
17   cat("Hypothesis is accepted");
18 } else{
19   cat("Hypothesis is rejected")
20 }
21
22 # *chi-sq value =10.72 given in book is wrong

```

R code Exa 10.29 Test of Independence ChiSq test

```

1 #H0: The defective/nondefective classification is
   independent of machinist classification
2 #H1: The defective/nondefective classification
   depends on machinist classification
3
4 def= c(10,8,14)
5 not_def =c(52,60,56)
6
7 observed =as.data.frame(rbind(def,not_def))
8 names(observed) <- c('Machinist A','Machinist B','
   Machinist C')
9
10 stat= chisq.test(observed)$statistic
11
12 alpha=0.01
13 dof=(3-1)*(2-1)
14
15 #Using rejection region approach
16 p_value=1 -pchisq(stat,df=2,lower.tail = F)
17
18 if(p_value>alpha){
19   cat("Hypothesis is accepted");
20 } else{
21   cat("Hypothesis is rejected")
22 }

```

R code Exa 10.30 ChiSq test

```
1 #H0 : Y follows a Poisson distribution
2 #H1 : Y does not follow a Poisson distribution
3
4 x= rep(0:2, times=c(32,12,6))
5 table(x)
6 mean(x)
7 probs = dpois(0:1, lambda=mean(x))
8 comp= 1- sum(probs)
9
10 stat = chisq.test(x=c(32,12,6), p= c(probs,comp),
11     simulate.p.value = TRUE)$statistic
12
13 alpha= 0.05
14 #degree of freedom = (3-1)- 1, as 1 parameter is
15     estimated
16 compare = qchisq(1- alpha,df=1)
17
18 if(stat<compare){
19     cat("Hypothesis is accepted");
20 } else{
21     cat("Hypothesis is rejected")
22 }
23
24 #Alternative soln..
25 p_value = 1-pchisq(stat,df=1,lower.tail = T)
26
27 if(p_value>0.05)
28 {
29     cat("Hypothesis is accepted")
30 } else{
```

```

31   cat("Hypothesis is rejected")
32 }
33
34 #Both solutions generate same results

```

R code Exa 10.31 Kolmogorov Smirnov test

```

1  #H0 : F(y) is exponential with theeta=2
2  #H1 : F(y) is not exponential with theeta=2
3
4  y = c
      (0.023,0.406,0.538,1.267,2.343,2.563,3.334,3.491,5.088,5.587)

5  Fy= 1- exp(-y/2)
6  n=10
7  i = 1:10
8
9  D_plus = i/n - Fy
10 D_minus = Fy - (i-1)/n
11 D = max(max(D_plus),max(D_minus))
12
13 # the critical value for a two-sided test with n =
      10 and alpha= 0.05 is 0.409.
14 D0=0.409
15 if(D0>D){
16   cat("Hypothesis is accepeted")
17 } else{
18   cat("Hypothesis is rejected")
19 }

```

R code Exa 10.32 Kolmogrov Smirnov Normality Test

```

1 data =c(0.3780, 0.5090, 0.6230 ,0.6860 ,0.7350 ,
          0.7520, 0.7580, 0.8690, 0.8890, 0.8890, 0.8990,
          0.9370, 0.9820 ,1.0220 ,1.0370 , 1.0880, 1.1230
          ,1.2060, 1.3340 ,1.4230
2 )
3
4 ks.test(data,"pnorm",mean(data),sd(data))
5
6 # the answers are different from those given in the
  book.

```

R code Exa 10.33 Kolmogrov Smirnov Normality Test

```

1 data =c
  (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3
2
3 #Exponential Distribution
4 ks.test(data,"pnorm",mean(data),sd(data))
5
6 #Lognormal Distribution
7 ks.test(log(data),"pnorm",mean(data),sd(data))
8
9 # the answers are different from those given in the
  book.

```

Chapter 11

Inference for Regression Parameters

R code Exa 11.2 SSE for the least squares line

```
1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3 n=length(x)
4
5 pol <- data.frame(x,y)
6
7 line_eq <- lm(y~x,data=pol)
8
9 cat("s=",summary(line_eq)$sigma)
```

R code Exa 11.3 95 percent confidence interval for the slope β_1

```
1 # to find the confidence interval of the given data
  ..
2
3 x = c(95,82,90,81,99,100,93,95,93,87)
```

```

4 y = c(214,152,156,129,254,266,210,204,213,150)
5
6 pol <- data.frame(x,y)
7
8 line_eq <- lm(y~x,data=pol)
9
10
11 cat("The 95% confidence interval for beta0(i.e
      intercept) in the regression line is",confint(
      line_eq,'x',level=0.95))

```

R code Exa 11.4 Testing the Slope of a Straight Line Model T test

```

1 #H0 : beta1 =0
2 #H1 : beta1 not= 0
3
4 #Given , alpha=0.05
5 x = c(95,82,90,81,99,100,93,95,93,87)
6 y = c(214,152,156,129,254,266,210,204,213,150)
7
8 pol <- data.frame(x,y)
9
10 line_eq <- summary(lm(y~x,data=pol))
11 t_value =(coefficients(line_eq)[2,1] - 0)/
      coefficients(line_eq)[2,2]
12
13 cat("Since the T value is",t_value," greater than ",
      qt(1- 0.025,df=10-2,lower.tail = F)," suggesting
      strong evidence that beta1 < 1.0")

```

R code Exa 11.5 fitting a line

```

1 #H0 :(OT contributes no information for the
  prediction of WR using a simple linear model
2 #H1 :(OT contributes information for the prediction
  of WR using a simple linear model
3
4 OT =c
  (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
5 WR =c
  (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
6 pol <- lm(WR~OT)
7 summary(pol)
8
9 # p_value <0.0001
10 cat(", we would reject the null hypothesis and
  conclude that the overtolerance provides
  information for the prediction of wall reduction"
  )

```

R code Exa 11.6 association between the test strength

```

1 #H0 : beta1 =1
2 #H1 : beta1 not=1
3
4 NAS <- c
  (25.1,20.9,25.5,21.3,25.4,20.9,38.5,31.6,63.8,52.6,61.8,51.2)
5 AS<- c
  (35.7,33.6,35.0,33.2,35.7,33.1,52.5,48.7,80.8,75.3,78.3,73.2)
6 pol <- data.frame(NAS,AS)
7
8 line_eq <- summary(lm(AS~NAS,data=pol))
9 t_value <-(coefficients(line_eq)[2,1] -1)/

```

```

    coefficients(line_eq)[2,2]
10
11
12 cat("Since the T value is",t_value," greater than ",
    qt(1- 0.025,df=12-2,lower.tail = F)," suggesting
    strong evidence that beta1 < 1.0")

```

R code Exa 11.7 tool life and the cutting speeds

```

1 speed =c
    (340,570,340,570,340,570,340,570,440,440,440,440,305,635,440,440,
2 life =c
    (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3
3
4 pol <- data.frame(speed,life)
5
6 l1 <- lm(life~speed,data = pol)
7 c =coefficients(l1)
8 cat("Tool Life =",c[1]," + ",c[2]," Cutting Speed")
9
10 p_value =summary(l1)$coefficients[,4][2]
11
12 cat("Because the p-value <0.001, we reject H0 :beta1
    =0 and conclude that cutting speed contributes
    significant amount of information toward the
    prediction of tool life. ")

```

R code Exa 11.8.a confidence interval for the mean peak power load

```

1 # To find the confidence interval of the mean
    response..

```



```

2
3 x = c(95,82,90,81,99,100,93,95,93,87)
4 y = c(214,152,156,129,254,266,210,204,213,150)
5
6 pol <- data.frame(x,y)
7
8 line <- lm(y~x,data = pol)
9
10 # here x0 <- 90 F
11 data <- data.frame(x=90)
12
13 c <- predict(line,data, interval = "confidence")
14
15 cat(" we can be 95% confident that the mean peak
      power load is between",c[2],c[3],"      megawatts
      for days with a maximum temperature of 90F  ")

```

R code Exa 11.8.b Predict the peak power load for a day

```

1 # To find the confidence interval of the mean
  response..
2
3 x = c(95,82,90,81,99,100,93,95,93,87)
4 y = c(214,152,156,129,254,266,210,204,213,150)
5
6 pol <- data.frame(x,y)
7
8 line <- lm(y~x,data = pol)
9
10 # here x0 <- 90 F
11 data <- data.frame(x=90)
12 res <- predict(line,data, interval = "prediction")
13
14 cat(" we are 95% confident that the peak power load
      will be between ",res[2],"—" ,res[3]," megawatts

```

on a particular day when the maximum temperature
is 90 F ")

R code Exa 11.9.a Estimate the mean wall reduction

```
1 OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
2 WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
3
4 pol <- data.frame(OT,WR)
5
6 line <- lm(WR~OT,data = pol)
7 data <- data.frame(OT=0.0020)
8
9 c <- predict(line,data, interval = "confidence")
10 cat(" The 95% confidence interval for estimating the
    mean wall reduction at OT = 0.0020 is ",c[2],c
    [3])
```

R code Exa 11.9.b Predict the amount of wall reduction

```
1 OT =c
    (-0.0010,-0.0010,-0.0005,-0.0005,-0.0005,-0.0005,0.0005,0.0010,0.
2 WR =c
    (1.30,3.00,1.60,3.50,4.25,4.30,3.80,3.80,2.70,4.40,4.90,2.80,3.50
3
4 pol <- data.frame(OT,WR)
5
```

```

6 line <- lm(WR~OT,data = pol)
7 data <- data.frame(OT=0.0020)
8 res <- predict(line,data, interval = "prediction")
9
10 cat(" we are 95% confident that the peak power load
      will be between ",res[2],"—" ,res[3]," megawatts
      on a particular day when the maximum temperature
      is 90 F ")

```

R code Exa 11.10 Polynomial Regression of degree 2

```

1 #Multiple Regression Analysis
2
3 usage = function(size){
4   -1216.14 + 2.39893*size - 0.00045*size*size
5 }
6
7 cat("For a house with 2,000 square feet area, the
      predicted electricity usage is ",usage(2000))

```

R code Exa 11.14 Fitting the model The least squares approach

```

1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3
4 pol <- lm(y~x+I(x^2))
5 coef = coefficients(pol)
6
7 cat("load =", coef[1]," ",coef[2],"temperature  +",
      coef[3],"temp^2")
8
9 cat("The SSE for this best line of fit is equal to "
      , anova(pol)["Residuals","Sum Sq"])

```

R code Exa 11.15 Estimation of error variance s^2

```
1 #Estimation of error variance sigma^2
2
3 x = c(95,82,90,81,99,100,93,95,93,87)
4 y = c(214,152,156,129,254,266,210,204,213,150)
5
6 pol <- lm(y~x+I(x^2))
7
8 s= summary(pol)$sigma
9
10 cat(" the mean square for error , or MSE= ", s*s)
```

R code Exa 11.16 Testing the Utility of a Multiple Regression Model The Global F test

```
1 #H0 : beta1 = beta2 =0
2 #H1 : at least one of the coefficients is nonzero
3
4 x = c(95,82,90,81,99,100,93,95,93,87)
5 y = c(214,152,156,129,254,266,210,204,213,150)
6 n=10
7 k=2
8 dof=n- (k+1)
9 pol <- lm(y~x+I(x^2))
10 F= summary(pol)$fstatistic["value"]
11 F
12 compare = qf(1-0.05,k,dof)
13
14 if(F < compare)
15 {
```

```

16   cat("Null Hypothesis is accepted")
17 } else{
18   cat("Null Hypothesis is rejected")
19 }

```

R code Exa 11.17 least square fit of the modal

```

1  #H0 : beta1 =beta2 =0
2  #H1 :: At least one is nonzero , i.e., the model is
      useful for predicting Y. The rejection region for
      this test at is .
3
4  y =c
      (121,169,172,116,53,177,31,94,72,171,23,177,178,65,146,129,40,167,
5
6  x1=c
      (6490,7244,7943,6478,3138,8747,2020,4090,3230,8786,1986,9653,9429,
7
8  x2=c
      (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,
9
10 fit=lm(y~x1+x2)
11 x=summary(fit)
12 coef =coefficients(x)
13 #a
14 F_value = 112.9
15 compare =qf(1-0.05,2,37,lower.tail = F)
16 if(F_value>compare){
17   cat("Null hypothesis is accepted")
18 } else{
19   cat("Null hypothesis is rejected")
20 }

```

```

21 #b
22 cat("y =",coef[1]," + ",coef[2],"x1 ",coef[3],"x2")

```

R code Exa 11.18 Estimating and testing hypotheses about beta2

```

1 #H0 : beta2 =0 , (No quadratic relationship exists
2 #H1 : beta2 >0 , (The peak power load increases at
   an increasing rate as the daily maximum
   temperature increases.)
3
4 x = c(95,82,90,81,99,100,93,95,93,87)
5 y = c(214,152,156,129,254,266,210,204,213,150)
6 n=10
7 k=2
8 dof=n- (k+1)
9
10 pol<- data.frame(x,y)
11 line <- lm(y~x+I(x^2),data = pol)
12 summary(line)
13 T =coef(summary(line))[3,3]
14 compare= qt(1-0.05,df=dof)
15 if(T < compare)
16 {
17   cat("Null Hypothesis is accepted")
18 } else{
19   cat("Null Hypothesis is rejected")
20 }
21 s=coef(summary(line))[3,2]
22 t =qt(1- 0.05,df=dof)
23 beta2 =coef(summary(line))[3,1]
24
25 cat(" confidence interval for the parameter beta2 as
   follows:",beta2- t*s," - ",beta2+ t*s)

```

R code Exa 11.19 model for mean lost work hours

```

1 #H0 : beta3 =0
2 #H1 : beta3 <0
3
4 y =c
      (121,169,172,116,53,177,31,94,72,171,23,177,178,65,146,129,40,167
5 x1=c
      (6490,7244,7943,6478,3138,8747,2020,4090,3230,8786,1986,9653,9429
6 x2=c
      (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,
7
8 fit=lm(y~x1+x2+I(x1*x2))
9 x=summary(fit)
10 coef =coefficients(x)
11 cat("The regression equation is y= ",coef[1]," + ",
      coef[2],"x1 + ",coef[3],"x2 ",coef[4],"x1*x2")
12
13 anova(fit)
14 dof=36
15 T =coef(summary(fit))[4,3]
16 compare= qt(0.05,df=dof)
17 if(T > compare)
18 {
19   cat("Null Hypothesis is accepted")
20 } else{
21   cat("Null Hypothesis is rejected")
22 }

```

R code Exa 11.20 multiple regression model for estimation and prediction

```
1 x = c(95,82,90,81,99,100,93,95,93,87)
2 y = c(214,152,156,129,254,266,210,204,213,150)
3
4 pol<- data.frame(x,y)
5 line <- lm(y~x+I(x^2),data = pol)
6 coef = coefficients(line)
7 y_cap = function(xp){
8   coef[1] + coef[2]*xp + coef[3]*xp*xp
9 }
10
11 cat(" the electrical usage for a particular day on
      which the high temperature is 90F, y_cap=", y_cap
      (90))
12 data <- data.frame(x=90)
13
14 d <- predict(line,data, interval = "prediction")
15
16 cat("The 95% prediction interval for y0 when x0=90
      is",d[2],"—" ,d[3])
17
18 f <- predict(line,data,interval = "confidence")
19 cat("The 95% prediction interval for y0 when x0=90
      is",f[2],"—" ,f[3])
```

R code Exa 11.21 least squares equation to predict tool life

```
1 speed =c
      (340,570,340,570,340,570,340,570,440,440,440,440,305,635,440,440,
2 life =c
      (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3
3 feed =c
```



```

(0.00630,0.00630,0.01410,0.01416,0.00630,0.00630,0.01416,0.01416,
4         0.00905,0.00905,0.00905,0.00905,0.00472,0.01732,
          0.00905, 0.00905)
5 depth =c
          (0.02100,0.02100,0.02100,0.02100,0.02100,0.04000,0.04000,0.04000,
           0.02900, 0.02900, 0.02900, 0.02900, 0.02900,
           0.02900, 0.02900, 0.02900,
6           0.01350,
           0.04550,0.02900,0.02900,0.02900,0.02900,
           0.01350,0.04550 )
7
8 dat= data.frame(life,speed,feed,depth)
9 fit =lm(life~speed+feed+depth)
10 #a
11 cor(dat)
12
13 #b
14 x=summary(fit)
15 coef= coefficients(x)
16 cat("the leastsquares equation : Tool life = ",coef
      [1],coef[2],"Speed  ",coef[3],"Feed rate  ",coef
      [4],"Depth of cut")
17
18
19 #c
20 y =data.frame(depth=0.03,speed=450,feed =0.01)
21
22 val= predict(fit,y,interval = "confidence")
23 cat(" A tool that is used to cut depths of 0.03 inch
      at a speed of 450 fpm with a feed rate of 0.01
      ipr is expected to last on the average ",val[1])
24 cat(" we are 95% confident that the mean life of
      such a tool used to cut depths of 0.03 inch at a
      speed of 450 fpm with a feed rate of 0.01 ipr
      will be between",val[2],"—",val[3])

```

R code Exa 11.22 A Test for a Portion of a Model

```
1 # A Test for a Portion of a Model
2
3
4 y=c
   (48.5,55.0,68.0,137.0,309.4,17.5,19.6,24.5,34.8,32.0,28.0,49.9,59.0)
5 area =c
   (1.1,1.01,1.45,2.4,3.3,0.4,1.28,0.74,0.78,0.97,0.84,1.08,0.99,1.0)
6 bedroom =c
   (3,3,3,3,4,1,3,3,2,3,3,2,2,3,3,3,3,3,3,3,3,3,3,3,4,3,2,2,3,3,3,
7 bathroom =c
   (1,2,2,3,3,1,1,1,1,1,1,2,1,2,2,1,2,2,2,2,2,1,2,2,
   rep(2,59),3,2,2,2,2,2,3,2,3,3)
8 style =c(0,0,0,1,rep(0,31)
   ,1,1,1,1,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,1,1,0,1,1,1,0,1,0,1,1,0,1,0,1,
9
10 # Multiple regression model for selling price of
   houses
11 fit =lm(y~area+bathroom+style)
12 summary(fit)
13
14 # Full model for selling prices
15 # Multiple regression model for selling price of
   houses
16 fit =lm(y~area+bedroom+bathroom+style)
17 summary(fit)
```

R code Exa 11.23.b Testing a Portion of a Model F test

```
1 speed =c
    (340,570,340,570,340,570,340,570,440,440,440,440,305,635,440,440,
2 life =c
    (70,29,60,28,64,32,44,24,35,31,38,35,52,23,40,28,46,33,46,27,37,3
3 feed =c
    (0.00630,0.00630,0.01410,0.01416,0.00630,0.00630,0.01416,0.01416,
4
    0.00905,0.00905,0.00905,0.00905,0.00472,0.01732,
    0.00905, 0.00905)
5 depth =c
    (0.02100,0.02100,0.02100,0.02100,0.02100,0.04000,0.04000,0.04000,
    0.02900, 0.02900, 0.02900, 0.02900, 0.02900,
    0.02900, 0.02900, 0.02900,
6
    0.01350,
    0.04550,0.02900,0.02900,0.02900,0.02900,
    0.01350,0.04550 )
7
8 dat= data.frame(life,speed,feed,depth)
9 fit =lm(life~speed+feed+depth+I(speed*feed)+I(feed*
    depth)+I(speed*depth)+I(speed*depth*feed))
10 x=summary(fit)
11 coef =coefficients(x)
12 cat("the leastsquares equation : Tool life = ",coef
    [1],coef[2]," Speed  ",coef[3]," Feed  ",coef[4],"
    Depth  + ",coef[5]," speed*feed + ",coef[6],
13 " feed*depth + ",coef[7]," speed*depth ",coef[8],"
    speed*feed*depth")
```

R code Exa 11.25 Representation of Mean Profit in the Additive Model

```
1 # Correspondence between Means and Model Parameters
```

```

2
3 y =c
    (0.065,0.073,0.068,0.036,0.078,0.082,0.050,0.043,0.048,0.046,0.06

4 x1 =c(0,0,0,1,0,0,1,1,0,0,1,1)
5 x2 =c(0,0,0,0,1,1,1,1,0,0,0,0)
6 x3 =c(0,0,0,0,0,0,0,0,1,1,1,1)
7
8 # a
9 # Main effects model
10 fit =glm(y~x1+x2+x3)
11 coef=coefficients(summary.glm(fit))
12 cat(" The least-squares prediction equation is yv= "
    ,coef[1],coef[2],"x1 + ",coef[3],"x2 ",coef[4],"
    x3")
13
14 # b
15 # Complete model including interactions
16 fit1 =glm(y~x1+x2+x3+(x1*x2)+(x1*x3))
17 coef1=coefficients(summary.glm(fit1))
18 cat(" . The least-squares prediction equation is ",
    coef1[1],coef1[2],"x1 +",coef1[3],"x2 ",coef1[4],
    "x3 ",coef1[5],"x1*x2 + ",coef1[6],"x1*x3")
19
20 # c
21 # H0:The interaction terms do not contribute to the
    model.(beta4 =beta5 =0)
22 # H1: At least one of interaction parameters is
    nonzero.
23
24 F_value= 64.04
25 compare =qf(1- 0.05,2,6)
26 if(F_value <compare){
27     cat(" Null hypothesis is accepted")
28 } else{
29     cat(" Null hypothesis is rejected")
30 }

```

R code Exa 11.26 Response surface method

```

1 # Response surface method
2
3
4 x1 =c
      (80,80,80,80,80,80,80,80,80,80,90,90,90,90,90,90,90,90,90,90,100,100,100)
5 x2 =c
      (50,50,50,55,55,55,60,60,60,50,50,50,55,55,55,60,60,60,50,50,50,55)
6 y = c
      (50.8,50.7,49.4,93.7,90.9,90.9,74.5,73.0,71.2,63.4,61.6,63.4,93.8,
7
8 dat <- data.frame(x1,x2,y)
9 model <- lm(y~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2), data =
      dat)
10 c =coefficients(summary(model))
11 cat(" The least-squares model is as follows:")
12 cat(c[1][1],"+",c[2][1],"x1 +",c[3][1],"x2 + ",c
      [4][1],"x1^2 + ",c[5][1],"x2^2 + ",c[6][1],"x1x2"
13
14 data <- data.frame(x1=86.25 ,x2= 55.58 )
15
16 d <- predict(model,data, interval = "confidence")
17
18 cat("The 95% confidence interval for y when x1=86.25
      and x2=55.58 is",d[2],"-" ,d[3])
19 cat("The 95% confidence interval for y when x1=86.25
      and x2=55.58 is",d[1])

```

R code Exa 11.27 Modeling a time trend

```
1 year = c
      (1970,1975,1980,1981,1982,1983,1984,1985,1986,1987,1988,1989,1990
2 suv = c
      (12.3,19.1,23.8,23.7,22.7,23.9,25.6,27.4,29.1,30.6,32.7,33.3,35.6
3
4 pol <- data.frame(year,suv)
5
6 l1 <- lm(suv~year,data = pol)
7 r1 = summary(l1)$r.squared
8 fit <- data.frame(year,log(suv))
9 l2 = lm(log(suv)~year,data=fit)
10 r2=summary(l2)$r.squared
11
12 cat(" Fitting log(FC) as a function of year produces
      a slightly better fit R^2= ",r2*100,"%", than the
      I modal")
```

R code Exa 11.28 Logistic regression

```
1 #Logistic Fit of Test by Pressure
2
3 pressure =c
      (3943,4163,3812,3888,3926,3900,3942,3732,4480,3940,4143,4146,3962
4 res =c
      (1,0,1,1,0,1,1,1,0,0,0,0,0,0,1,1,1,1,0,1,1,0,0,1,0)
5
```

```
6 dat=data.frame(pressure,res)
7 modal =glm(res~pressure,family = "binomial",data =
    dat)
8
9 summary(modal)
10 anova(modal,test = "Chisq")
```

Chapter 12

Analysis of Variance

R code Exa 12.2 Test to Compare k Treatment Means for a Completely Randomized Design

```
1 #H0 : the mean stopping times at the three types of
   signals are the same i.e  $\mu_1 = \mu_2 = \mu_3$ 
2
3 #H1 : The mean stopping times for at least two types
   of signals are different.
4
5
6 a =c
   (36.6,39.2,30.4,37.1,34.1,17.5,20.6,18.7,25.7,22.0,15.0,10.4,18.9

7 b =c(rep(1,5),rep(2,5),rep(3,5))
8 dat =data.frame(a,b)
9
10 c <- aov(a~factor(b),data = dat)
11 summary(c)
12
13 #From summary table we obtain data as:
14 cat("SST = ",1202.6)
15 cat("SSE = ",137.83)
16 cat("TSS = ",1202.6+137.83)
```



```

17
18 #F value=52.35
19 compare = qf(1-0.05,2,12)
20
21 cat("Since F0.05(2,12) < Fvalue, we we reject the
      null hypothesis of equal means and conclude that
      at least two types of signals have different mean
      stop times")

```

R code Exa 12.3 mean score for the three groups of managers

```

1 #H0 : mua =mub =muc
2 #H1 : At least two group means are different
3
4 a =c( 82, 114, 90, 80, 88, 93, 80, 105, 128, 90,
        130, 110, 133, 130, 104, 156, 128, 151, 140 )
5 b =c(rep(1,8),rep(2,7),rep(3,4))
6 dat =data.frame(a,b)
7
8 x =lm(a~factor(b),data = dat)
9 summary(x)
10
11 c <- aov(a~factor(b),data = dat)
12 summary(c)
13
14 cat("The p-value for the F-test is < 0.0001, which
      means we would reject the null hypothesis of
      equal population mean scores for three groups of
      managers and conclude that at least two groups
      differ in their mean scores. ")

```

R code Exa 12.4 test for mean counts show significant differences

```

1 #H0 : muAN = muLC = muEC
2 #H1 : the mean counts differ for the tree metals
3
4
5 a =c
      (10,9,9,9,10,11,14,11,8,11,7,6,9,7,8,10,12,14,9,8,8,9,8,7,10,
6      14,12,15,14,10,17,16,11,13,14,15,11,16,12,6,13,20,17,10,16,10,1
7      42,46,44,39,50,34,42,40,36,37,46,42,43,50,32,41,37,49,28,34,34,
8
9 b =c(rep(1,25),rep(2,25),rep(3,25))
10
11 dat =data.frame(a,b)
12
13 # the analysis must be done on the transformed data
      because of the lack of homogeneity of variances.
14
15 x =lm(sqrt(a)~factor(b),data = dat)
16 summary(x)
17 c <- aov(sqrt(a)~factor(b),data = dat)
18 summary(c)
19
20 cat(" the p-value <0.001 so, we reject the null
      hypothesis of equality of mean square root count
      for three metals, and conclude that at least two
      mean square root counts are different. ")

```

R code Exa 12.5 Equivalence between a t test and an F test

```

1 #H0 : mu1 =mu2
2 #H1 : mu1 not= mu2
3
4 a =c(253,261,258,255,256,264,265,261,257,256)

```

```

5 b =c(rep(1,5),rep(2,5))
6 dat =data.frame(a,b)
7 c <- aov(a~factor(b),data = dat)
8 summary(c)
9
10 t.test(a[1:5],a[6:10],alternative = "two.sided",var.
      equal = TRUE,mu=0,conf.level = 0.05)
11
12 # t=-1.768 and F value = 3.125 which is equal to t
      ^2.
13
14 cat("Because p-value = 0.1151 is larger than , we
      fail to reject the null (using either test) and
      conclude that there is no evidence of significant
      difference between the mean UTI for wires
      provided by the two suppliers")

```

R code Exa 12.6 common variance using a pooled sample variance

```

1 a =c
      (36.6,39.2,30.4,37.1,34.1,17.5,20.6,18.7,25.7,22.0,15.0,10.4,18.9
2 b =c(rep(1,5),rep(2,5),rep(3,5))
3 dat =data.frame(a,b)
4
5 c <- aov(a~factor(b),data = dat)
6 cat("MSE =" ,anova(c)[["Mean Sq"]][2])

```

R code Exa 12.7 modal for the test score of one manager

```

1 # ANOVA and regression analysis
2

```

```

3 a =c( 82, 114, 90, 80, 88, 93, 80, 105, 128, 90,
      130, 110, 133, 130, 104, 156, 128, 151, 140 )
4 b =c(rep(1,8),rep(2,7),rep(3,4))
5 dat =data.frame(a,b)
6
7 x =lm(a~factor(b),data = dat)
8 summary(x)
9 anova(x)

```

R code Exa 12.8 Confidence Intervals for Means in the Completely Randomized Design Bonferroni Method

```

1 a =c( 82, 114, 90, 80, 88, 93, 80, 105, 128, 90,
      130, 110, 133, 130, 104, 156, 128, 151, 140 )
2 b =c(rep(1,8),rep(2,7),rep(3,4))
3
4 dat =data.frame(a,b)
5
6 x =lm(a~factor(b),data = dat)
7 s =summary(x)$sigma
8
9 # Bonferroni Method ,c=3
10 alpha =0.05
11 c=3
12 k=(alpha/2)/c
13 t = qt(1-k,df=16)
14
15 #Three intervals are constructed as follows:
16
17 a_b = mean(a[1:8]) - mean(a[9:15])
18 x1= t*s*sqrt(1/8 + 1/7)
19 cat("Interval muA - muB =", a_b-x1,a_b +x1)
20
21 a_c = mean(a[1:8]) - mean(a[16:19])
22 x2 =t*s*sqrt(1/8 + 1/4)

```

```

23 cat("Interval muA - muC =", a_c-x2,a_c +x2)
24
25 b_c = mean(a[9:15]) - mean(a[16:19])
26 x3 =t*s*sqrt(1/7 + 1/4)
27 cat("Interval muB - muC =", b_c-x3,b_c +x3)

```

R code Exa 12.9 confidence intervals for the pairwise difference in mean stop times

```

1 a =c
    (36.6,39.2,30.4,37.1,34.1,17.5,20.6,18.7,25.7,22.0,15.0,10.4,18.9

2 b =c(rep(1,5),rep(2,5),rep(3,5))
3 dat =data.frame(a,b)
4
5 x =lm(a~factor(b),data = dat)
6 s =summary(x)$sigma
7
8 # Bonferroni Method ,c=3
9 alpha =0.05
10 c=3
11 k=(alpha/2)/c
12 t = qt(1-k,df=12)
13
14 #a
15 #Three intervals are constructed as follows:
16
17 a_b = mean(a[1:5]) - mean(a[6:10])
18 x1= t*s*sqrt(1/5 + 1/5)
19 cat("Interval muPre - muSA =", a_b-x1,a_b +x1)
20
21 a_c = mean(a[1:5]) - mean(a[11:15])
22 x2 =t*s*sqrt(1/5 + 1/5)
23 cat("Interval muPre - muFA =", a_c-x2,a_c +x2)
24

```

```

25 b_c = mean(a[6:10]) - mean(a[11:15])
26 x3 =t*s*sqrt(1/5 + 1/5)
27 cat("Interval muSA - muFA =", b_c-x3,b_c +x3)
28
29 #b
30 y = mean(a[11:15])
31 d =s/sqrt(5)
32 cat("a 95% confidence interval for the mean stop
      time of the best signal i.e FA =",
33     y - d," - ",y+d)

```

R code Exa 12.10 Test to Compare k Treatment Means for a Randomized Block Design

```

1  c1 =c(5,9)
2  c2 =c(3,8)
3  c3 =c(8,13)
4  c4 =c(4,6)
5
6  dat <- rbind(c1,c2,c3,c4)    # combining rows to make
      matrix..
7
8  a <- c(t(as.matrix(dat)))    # concatenate different
      rows into a vector..
9
10 b <- c("b1","b2") # treatment levels
11
12 n_tr <- 2 # no. of treatment levels
13
14 n_cont <- 4 # no. of control blocks..
15 block <- gl(n_tr,1,n_cont*n_tr,factor(b)) # vector
      of treatment factors corresponding to each
      element of vector a..
16
17 chemical <- gl(n_cont,n_tr,n_tr*n_cont) # vector of

```

```

        blocking factors corresponding to each element in
        vector a..
18
19 print("The Analysis of Variance table is:")
20
21 summary(aov(a~block+chemical)) # anova table display
    ..
22
23 #A
24 #H0 : mu1 =mu2 =mu3 =mu4
25 #H1 : the mean resistance differs for at least two
        treatments.
26 #F ratio for block is:
27
28 F_value_chemical = 12.33
29 compare = qf(1-0.05,3,3)
30
31 if(F_value_chemical < compare){
32     cat("Null hypothesis is accepted")
33 } else{
34     cat("Null hypothesis is rejected")
35 }
36
37
38 #B
39 #H0 : m1 =mu2
40 #H1 : there is evidence of significant difference
        between the block (fabric) means
41 #F ratio for block is:
42
43 F_value_block = 32
44 compare = qf(1-0.05,1,3)
45
46 if(F_value_block < compare){
47     cat("Null hypothesis is accepted")
48 } else{
49     cat("Null hypothesis is rejected")
50 }

```

R code Exa 12.11 difference in mean gains

```
1 #H0 :  $\mu_D=0$  that is , that there is no difference in
   the mean gain by modified and conventional
   systems
2 #H1 :  $\mu_D \neq 0$  , that is , that the mean gain by the
   modified system differs from the mean gain by
   the conventional system
3
4 M= c( 1.776, 1.637 ,1.554, 1.460 ,1.405)
5 C= c(1.901, 1.730 ,1.629 ,1.517 ,1.451)
6 p_value= t.test(M,C,paired = TRUE,alternative = "two
   .sided")$p.value
7
8 if(p_value <0.05){
9   cat("Null hypothesis is rejected")
10 } else{
11   cat("Null hypothesis is accepted")
12 }
```

R code Exa 12.12 ANOVA for RBD and regression analysis

```
1 c1 =c(5,9)
2 c2 =c(3,8)
3 c3 =c(8,13)
4 c4 =c(4,6)
5
6 dat <- rbind(c1,c2,c3,c4)    # combining rows to make
   matrix..
7
8 a <- c(t(as.matrix(dat)))    # concatenate different
   rows into a vector..
```



```

9
10 b <- c("b1","b2") # treatment levels
11
12 n_tr <- 2 # no. of treatment levels
13
14 n_cont <- 4 # no. of control blocks..
15 block <- gl(n_tr,1,n_cont*n_tr,factor(b)) # vector
      of treatment factors corresponding to each
      element of vector a..
16
17 chemical <- gl(n_cont,n_tr,n_tr*n_cont) # vector of
      blocking factors corresponding to each element in
      vector a..
18
19 print("The Analysis of Variance table is:")
20
21 summary(aov(a~block+chemical)) # anova table display
      ..
22
23 #F ratio for block is:
24
25 F_value_chemical = 12.33

```

R code Exa 12.13 Confidence Intervals for Means in the Randomized Block Design

```

1 c1 =c(5,9)
2 c2 =c(3,8)
3 c3 =c(8,13)
4 c4 =c(4,6)
5
6 dat <- rbind(c1,c2,c3,c4) # combining rows to make
      matrix..
7
8 a <- c(t(as.matrix(dat))) # concatenate different

```

```

        rows into a vector..
9
10 b <- c("b1","b2") # treatment levels
11
12 n_tr <- 2 # no. of treatment levels
13
14 n_cont <- 4 # no. of control blocks..
15 block <- gl(n_tr,1,n_cont*n_tr,factor(b)) # vector
        of treatment factors corresponding to each
        element of vector a..
16
17 chemical <- gl(n_cont,n_tr,n_tr*n_cont) # vector of
        blocking factors corresponding to each element in
        vector a..
18
19 print("The Analysis of Variance table is:")
20
21 summary(aov(a~block+chemical)) # anova table display
        ..
22
23 #MSE = 1
24 s=sqrt(1)
25
26
27 # Bonferroni Method ,c=3
28 alpha =0.10
29 c=6
30 k=(alpha/2)/c
31 t = qt(1-k,df=3)
32
33 x= t*s*sqrt(1/2 + 1/2)
34 cat("mu1 -mu2 =", mean(c1) -mean(c2) -x,mean(c1) -
        mean(c2) +x)
35 cat("mu1 -mu3 =", mean(c1) -mean(c3) -x,mean(c1) -
        mean(c3) +x)
36 cat("mu1 -mu4 =", mean(c1) -mean(c4) -x,mean(c1) -
        mean(c4) +x)
37 cat("mu2 -mu3 =", mean(c2) -mean(c3) -x,mean(c2) -

```

```

      mean(c3) +x)
38 cat("mu2 -mu4 =", mean(c2) -mean(c4) -x,mean(c2) -
      mean(c4) +x)
39 cat("mu3 -mu4 =", mean(c3) -mean(c4) -x,mean(c3) -
      mean(c4) +x)
40 #Simialrly , other differences can be computed

```

R code Exa 12.14 Analysis of variance for the factorial experiment

```

1 y1 =c(9,8,8,7)
2 y2 =c(5,6,3,4)
3
4 observations <- c(y1,y2)
5 A <- c(rep(1,4),rep(2,4))
6
7 a <- c(rep(1,2),rep(2,2))
8
9 B <- c(rep(a,2))
10 dat <- data.frame(observations,A,B)
11 d <- aov(observations~factor(B)*factor(A),data = dat
   )
12
13 summary(d)

```

R code Exa 12.15 confidence intervals for the six possible differences between treatment means

```

1 y1 =c(9,8,3,4)
2 y2 =c(5,6,8,7)
3
4 observations <- c(y1,y2)
5 A <- c(rep(1,4),rep(2,4))
6

```

```

7 a <- c(rep(1,2),rep(2,2))
8
9 B <- c(rep(a,2))
10 dat <- data.frame(observations,A,B)
11 d <- aov(observations~factor(A)*factor(B),data = dat
12      )
13 summary(d)
14 # NOTE: Sum Sq value for factor(A)*factor(B)= 24.5
15         and Mean Sq value for A is 0.5 . (misprinted in
16         textbook)
17
18 s=sqrt(anova(d)[["Mean Sq"]][4])
19 # Bonferroni Method ,c=3
20 alpha =0.05
21 c=6
22 k=(alpha/2)/c
23 t = qt(1-k,df=4)
24 x= t*s*sqrt(1/2 + 1/2)
25 cat(" Thus, all six intervals will be of the form :
26     yi_bar - yj_bar +- ",x)

```

R code Exa 12.16.a Fitting higher order models

```

1 x1 =c(14.05,14.93,16.56,15.85)
2 x2 =c(10.55,9.48,13.63,11.75)
3 x3 =c(7.55,6.59,9.23,8.78)
4
5 observations <- c(x1,x2,x3)
6 A <- c(rep(1,4),rep(2,4),rep(3,4))
7 a <- c(rep(1,2),rep(2,2))
8
9 B <- c(rep(a,3))
10

```

```

11 dat <- data.frame(observations,A,B)
12
13 d <- aov(observations~factor(A)*factor(B),data = dat
14 )
15 summary(d) # analysis of variance table..

```

R code Exa 12.17 Factorial Design

```

1 E1 =c(14,10)
2 E2 =c(14,11)
3 E3 =c(12,11)
4 E4 =c(11,12)
5 W1 =c(4,5)
6 W2 =c(4,5)
7 W3 =c(3,6)
8 W4 =c(5,6)
9
10 r=c(E1,E2,E3,E4,W1,W2,W3,W4)
11 r
12 f1 = c("X", "Y")           # 1st factor levels
13 f2 = c("Dry", "Damp")      # 2nd factor levels
14 k1 = length(f1)           # number of 1st factors
15 k2 = length(f2)           # number of 2nd factors
16 n = 4                      # observations per
    treatment
17
18 A = gl(k1, 1, n*k1*k2, factor(f1))
19 A
20
21 B = gl(k2, n*k1, n*k1*k2, factor(f2))
22 B
23
24 av = aov(r ~ A * B) # include interaction
25

```

26 `summary(av)`

R code Exa 12.18 effect on defrosted fish by freezing method defrosting method and duration

```
1 #H0: there is no 3 factor interaction Duration*
   Defrost*Freeze effect on the quality.
2 #H1: 3 factor interaction Duration*Defrost*Freeze
   affects quality.
3
4
5 A= c(73,70,65,65,68,75,74,69,67,67,74,73)
6 B= c(75,71,69,70,69,68,76,72,69,72,81,61)
7 C= c(74,70,70,72,70,70,74,65,65,69,80,74)
8
9 obs= c(A,B,C)
10 freeze= c(rep(1,12),rep(2,12),rep(3,12))
11 defrost= c(rep(1:3,12))
12 a= c(rep("1 day",3),rep("8 days",3))
13
14 duration= c(rep(a,6))
15 dat= data.frame(obs,freeze,defrost,duration)
16 d= aov(obs~(factor(freeze)*factor(defrost)*factor(
   duration)),data = dat)
17
18 print("The Analysis of Variance Table is shown as
   follows:")
19
20 summary(d)
21
22 cat("pvalue =0.20507 is higher than any reasonable
   level of significance. Therefore, fail to reject
   the null hypothesis and conclude that there is no
   evidence of the presence of a 3-factor
   interaction. ")
```

```

23
24 x = lm(obs~(factor(freeze)*factor(defrost)*factor(
      duration)),data = dat)
25 print("Summary of fit is shown below:")
26 summary(x)

```

R code Exa 12.19 yield differ significantly by temperature pressure and reaction time

```

1 p1 =c(68.5,72.8,72.5,74.5,72.0,75.5,70.5,69.5,65.0)
2 p2 =c(73.0,80.1,72.5,75.0,81.5,70.0,72.5,84.5,66.5)
3 p3 =c(68.7,72.0,73.1,74.6,76.0,76.0,74.7,76.0,70.5)
4
5 obs =c(p1,p2,p3)
6 pressure =c(rep(30,9),rep(70,9),rep(100,9))
7 temp =c(rep(1:3,9))
8 a =c(rep("1 hour",3),rep("2 hours",3),rep("3 hours",
      ,3))
9 time =c(rep(a,3))
10 dat= data.frame(obs,pressure,temp,time)
11 d= aov(obs~(factor(pressure)*factor(temp)*factor(
      time)),data = dat)
12
13 print("The Analysis of Variance Table is shown as
      follows:")
14
15 summary(d)

```
