R Textbook Companion for Elementary Statistics: A Step by Step Approach by Allan G. Bluman[1]

Created by
R. Janani
B.Tech.
Electrical Engineering
National Institute of Technology, Tiruchirappalli
Cross-Checked by
R TBC Team

May 15, 2020

[1]Funded by a grant from the National Mission on Education through ICT - http://spoken-tutorial.org/NMEICT-Intro. This Textbook Companion and R codes written in it can be downloaded from the "Textbook Companion Project" section at the website - https://r.fossee.in.

Book Description

Title: Elementary Statistics: A Step by Step Approach

Author: Allan G. Bluman

Publisher: McGraw-Hill, New York

Edition: 7

Year: 2009

ISBN: 9780073534978

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

Exa Example (Solved example)

Eqn Equation (Particular equation of the above book)

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

Contents

List of R Codes

Chapter 2

Frequency Distribution and graphs

R code Exa 2.1 Frequency table

```
a<-c("A","B","B","AB","O","O","B","AB","B","B","B","O","A","O","A","O","O","O","AB","AB","A","O","B","A","O")
results<-table(a)
results
results/length(a)*100
```

R code Exa 2.2 Frequency table

```
mpg<-c
   (112,110,107,116,120,100,118,112,108,113,127,117,114,110,120,120,

H = max(mpg)
L = min(mpg)
cat("Maximum value is: ", H)
cat("Minimum value is: ", L)
```

```
Range = H - L
cat("Range is: ", Range)
cat("\n")
breaks= seq(99.5,134.5, by=5)
mpg.div=cut(mpg, breaks, right=TRUE)
mpg.freq=table(mpg.div)
table<-data.frame(mpg.freq)
table
```

R code Exa 2.3 MPG for SUV

```
mpg<-c(12, 17, 12, 14, 16, 18, 16, 18, 12, 16, 17,
   15, 15, 16, 12, 15, 16, 16, 12, 14, 15, 12, 15,
   15, 19, 13, 16, 18, 16, 14)
H = max(mpg)
L = min(mpg)
cat("Maximum value is: ", H)
cat("\n")
cat("Minimum value is: ", L)
cat("\n")
Range = H - L
cat("Range is: ", Range)
cat("\n")
breaks= seq(12, 20, by=1)
mpg.div=cut(mpg, breaks, right=FALSE)
mpg.freq=table(mpg.div)
table<-data.frame(mpg.freq)
table
#for cumulative frequency
mpg.cumfreq = cumsum(mpg.freq)
tbl<-data.frame(mpg.cumfreq)
tbl
```

R code Exa 2.4 histogram Record high temperatures

```
temp<-c(112, 100, 127, 120, 134, 118, 105, 110, 109,
    112, 110, 118, 117, 116, 118, 122, 114, 114,
   105, 109, 107, 112, 114, 115, 118, 117, 118, 122,
    106, 110, 116, 108, 110, 121, 113, 120, 119,
   111, 104, 111, 120, 113, 120, 117, 105, 110, 118,
    112, 114, 114)
H = max(temp)
L = min(temp)
Range = H - L
Width = ceiling(Range/7)
#though the question specifies lower limit as 99.5
   and upper limit as 134.5, the considered upper
   and lower limits are 100 and 135
hist(temp,main="Histogram of Recorded Temparatures",
   xlab="temparature",border="red",col="yellow",xlim
   =c(100,135),las=1,breaks=Width)
```

R code Exa 2.5 frequency polygon

```
temp<-c(112, 100, 127, 120, 134, 118, 105, 110, 109,
    112, 110, 118, 117, 116, 118, 122, 114, 114,
   105, 109, 107, 112, 114, 115, 118, 117, 118, 122,
    106, 110, 116, 108, 110, 121, 113, 120, 119,
   111, 104, 111, 120, 113, 120, 117, 105, 110, 118,
    112, 114, 114)
H = max(temp)
L = min(temp)
Range = H - L
Width = ceiling(Range/7)
#using package "mosaic"
library(mosaic)
freqpoly(temp)
```

R code Exa 2.6 ogive

```
temp<-c(112, 100, 127, 120, 134, 118, 105, 110, 109,
    112, 110, 118, 117, 116, 118, 122, 114, 114,
   105, 109, 107, 112, 114, 115, 118, 117, 118, 122,
    106, 110, 116, 108, 110, 121, 113, 120, 119,
   111, 104, 111, 120, 113, 120, 117, 105, 110, 118,
    112, 114, 114)
H = max(temp)
L = min(temp)
Range = H - L
Width = ceiling(Range/7)
#installed package agricolae
library(agricolae)
#since the histogram division is from 100 to 135,
   ther is a shift in the ogive curve, when compared
    with the original
temp.h=hist(temp,main="Histogram of Recorded
   Temparatures",xlab="temparature",border="red",col
   ="yellow",xlim=c(100,135),las=1,breaks=7)
ogive.freq(temp.h, xlab="temparature", ylab="
   relative cumulative frequency")
```

R code Exa 2.7 Histogram frequency polygon and ogives

```
fr<-c
   (6,11,11,17,17,17,22,22,22,22,22,27,27,27,27,32,32,32,37,37)

H=40.5
L=5.5
Range = H - L
Width = ceiling(Range/7)
```

```
#though the question specifies lower limit as 99.5
   and upper limit as 134.5, the considered upper
   and lower limits are 100 and 135
fr.h=hist(fr,main="Histogram of Recorded
   Temparatures",xlab="temparature",border="red",col
   ="yellow",xlim=c(5,41),las=1,breaks=Width)
fr.h
#using package "mosaic"
library(mosaic)
freqpoly(fr)
library(agricolae)
ogive.freq(fr.h, xlab="miles", ylab="relative
   cumulative frequency")
#install package "agricolae"
```

R code Exa 2.8 bar graphs

```
#bar plot construction
amount<-c(728, 344, 141, 72)
barplot(amount, main="Expenditure of a first year
   college student", xlab="expenditure", ylab="
   amount", names.arg =c("electronics", "dorm decor"
   , "clothing", "shoes"), border="green")
barplot(amount, main="Expenditure of a first year
   college student", xlab="expenditure", ylab="
   amount", names.arg =c("electronics", "dorm decor"
   , "clothing", "shoes"), border="green", horiz=
   TRUE)
```

R code Exa 2.9 pareto chart

```
tp<-c(2.9, 4.3, 6, 3.8, 5.8)
```

```
2 names(tp)<-c("Indiana","Oklahoma","Florida","Maine",
    "Pennysylvania")
3 #install package qcc that contains pareto chart
4 library(qcc)
5 pareto.chart(tp, xlab="States", ylab="Number of
    Turnpikes", main="Turnpike Distribution", col="
    blue")
```

R code Exa 2.10 time series graph

```
1 damage<-c(2.8, 3.3, 3.4, 5.0, 8.5)
2 ts(damage, start=2001, end= 2005)
3 names(damage)<-c("2001", "2002", "2003", "2004", "
    2005")
4 plot.ts(damage,xlab="Year",ylab="Damage(in millions)
    ")
```

R code Exa 2.11 pie graph

```
1 snacks<-c(11.200, 8.200, 4.300, 3.800, 2.500)
2 food<-c("Potato chips", "Tortilla chips", "Pretzels"
    , "Popcorn", "Snack Nuts")
3 pct <- round(snacks/sum(snacks)*100, digits = 1)#
    round off to one decimal place
4 pct
5 food <- paste(food,"(", pct, ")")
6 food <- paste(food,"%",sep="")
7 pie(snacks, labels=food, main="Amount of food eaten
    during super bowl")
```

Chapter 3

Data Description

Chapter 4

Probability and counting rules

R code Exa 4.17 Find Probability

```
1 dough<-c(3,4,5) #glazed, jelly and chocolate
    doughnuts respectively
2 pr<-(dough[3]+dough[1])/sum(dough)
```

```
3 n<-paste("the answer is",round(pr,digits=2))
4 n
```

R code Exa 4.21 Probability

```
1 designation<-c(8,5)
2 gender<-c(10,3)
3 repnf<-7
4 reppf<-3
5 repnm<-designation[1]-repnf
6 pr<-(designation[1]+gender[2]-repnm)/sum(designation
    )
7 n<-paste("The probability that selected staff is",
    round(pr, digits=2))
8 n
```

R code Exa 4.22 Probability

```
1 p1<-0.32
2 p2<-0.09
3 p3<-0.06
4 pr<-p1+p2-p3
5 n<-paste("probability of having an accident while
    intoxicated or driving is",pr)
6 n
```

R code Exa 4.27 Probability

```
pr<-0.09
ne<-3
lbs<-paste("The probability is",(pr^ne))
lbs
```

R code Exa 4.28 University Crime

```
c1=5
c2=16
c3=32
t=c1+c2+c3
c21=c2-1
p=(c2/t)*(c21/(t-1))
l<-paste("the probability that both will have
   occurred in 2004 is",p)
l
```

R code Exa 4.29 Homeowners and Automobile Insurance

```
a=0.53
h=0.27
l<-paste("the probability that both will have
    occurred in 2004 is",a*h)
l
```

R code Exa 4.30 Drawing Cards

```
l<-paste("Probability of getting three jacks is",(4/
    52)*(3/51)*(2/50))
l
l<-paste("Probability of getting an ace, a king, and
     a queen in order is",(4/52)*(4/51)*(4/50))
l
l<-paste("Probability of getting a club, a spade,
    and a heart in order is",(13/52)*(13/51)*(13/50))
l
l<-paste("Probability of getting three clubs is",(13
    /52)*(12/51)*(11/50))
l
```

R code Exa 4.42 Permutation

```
l<-paste("There are",factorial(5),"different
    possible rankings")
l
```

R code Exa 4.43 Permutation

```
n=5
```

```
2 r=3
3 l<-paste("There are",factorial(n)/factorial(n-r),"
    different ways")
4 l
```

R code Exa 4.44 Permutation and Combination

```
1 n=8
2 r=3
3 l<-paste("There are",factorial(n)/factorial(n-r),"
    different ways")
4 l
```

R code Exa 4.45 Permutation and Combination

```
1 n=9
2 r=2
3 l<-paste("There are",factorial(n)/factorial(n-r),"
    different ways")
4 l
```

R code Exa 4.47 Combinations

```
1 n=4
2 r=2
3 l<-paste("There are",choose(n,r),"different ways")
4 l
```

R code Exa 4.48 Permutation and Combination

```
n=8
r=3
l<-paste("There are",choose(n,r),"different ways")
l
```

R code Exa 4.49 Permutation and Combination

```
n1=5
n2=7
r1=2
r2=3
l<-paste("There are",choose(n1,r1)*choose(n2,r2),"
    different ways")
l
```

R code Exa 4.50 Probability using combinations

```
n1=4
n2=52
r1=4
r2=5
l<-paste("Probability is",48*choose(n1,r1)/choose(n2
    ,r2))
l
```

R code Exa 4.51.a Probability using combinations

```
n1=4
```

```
2 n2=20
3 n3=24
4 r1=2
5 r2=2
6 r3=4
7 l<-paste("There are",choose(n1,r1)*choose(n2,r2)/
    choose(n3,r3),"different ways")
8 l
```

R code Exa 4.51.b Probability using combinations

```
1 n1=20
2 n2=24
3 r1=4
4 r2=4
5 l<-paste("There are",choose(n1,r1)/choose(n2,r2),"
    different ways")
6 l
```

R code Exa 4.51.c Probability using combinations

```
1 n1=24
2 r1=4
3 l<-paste("There are",1/choose(n1,r1),"different ways
    ")
4 l
```

R code Exa 4.51.d Probability using combinations

```
1 n1=20
```

```
2 n2=24
3 r1=4
4 r2=4
5 l<-paste("There are",1-(choose(n1,r1)/choose(n2,r2))
    ,"different ways")
6 l
```

R code Exa 4.52 Probability using combinations

```
1 n1=6
2 n2=8
3 n3=14
4 r1=1
5 r2=1
6 r3=2
7 l<-paste("There are",choose(n1,r1)*choose(n2,r2)/
    choose(n3,r3),"different ways")
8 l
```

Chapter 5

Discrete Probability Distrubution

R code Exa 5.2 Probability Distribution

```
h<-c(0,1,2,3)
ss<-8
pr<-c(1/ss,3/ss,3/ss,1/ss)
tbl<-data.frame(h,pr)
tbl
```

R code Exa 5.5 Mean of Probability distribution

```
out<-c(1,2,3,4,5,6)
pr<-c(1/6,1/6,1/6,1/6,1/6,1/6)
mea<-weighted.mean(out,pr)
n<-paste("the mean of the probability distribution
    is",round(mea,digits=3))
n
```

R code Exa 5.6 Mean of probability distribution

```
1 out<-c(0,1,2)
2 pr<-c(1/4,1/2,1/4)
3 mea<-weighted.mean(out,pr)
4 n<-paste("the mean of the probability distribution
     is",round(mea,digits=3))
5 n
```

R code Exa 5.7 Tossing Coins

```
1 h<-c(0,1,2,3)
2 p<-c(1/8,3/8,3/8,1/8)
3 weighted.mean(h,p)
```

R code Exa 5.8 Mean of Probability Distribution

```
1 out<-c(0,1,2,3,4)
2 pr<-c(0.06, 0.70, 0.20, 0.03, 0.01)
3 mea<-weighted.mean(out,pr)
4 n<-paste("the mean of the probability distribution
     is",round(mea,digits=3))
5 n
```

R code Exa 5.9 Variance and standard deviation of probability distribution

```
out<-c(1,2,3,4,5,6)
pr<-c(1/6,1/6,1/6,1/6,1/6,1/6)
mea<-weighted.mean(out,pr)
v<-round(sum(out*out*pr),digits=1)-(round(mea,digits
   =1)*round(mea,digits=1))
sd<-round(sqrt(v),digits=1)
n<-paste("the variance and standard deviation of the
    probability distribution is",v,"and",sd,"
   respectively")
n
```

R code Exa 5.10 Discrete Probability Distribution

```
out<-c(3,4,5)
pr<-c(2/5,1/5,2/5)
mea<-weighted.mean(out,pr)
v<-round(sum(out*out*pr),digits=1)-(round(mea,digits
   =1)*round(mea,digits=1))
sd<-round(sqrt(v),digits=1)
n<-paste("the variance and standard deviation of the
    probability distribution is",round(v,digits=1),"
   and",sd,"respectively")
n
```

R code Exa 5.11 Variance and standard deviation of probability distribution

```
out<-c(0,1,2,3,4)
pr<-c(0.18, 0.34, 0.23, 0.21, 0.04)
mea<-weighted.mean(out,pr)
v<-round(sum(out*out*pr),digits=1)-(round(mea,digits
   =1)*round(mea,digits=1))
sd<-round(sqrt(v),digits=1)
```

```
round(v,digits=1)
n<-paste("the variance and standard deviation of the
    probability distribution is",round(v,digits=1),"
   and",sd,"respectively")
n
k=2
ct<-round(mea,digits=1)+(k*round(sd,digits=1))
f<-paste("Most calls will be accomodated by",round(
   ct),"phone lines")
f
```

R code Exa 5.12 Expectation

```
gain<-c(350,-1)#win and lose
pr<-c(0.001,0.999)
e<-weighted.mean(gain,pr)
n<-paste("the expectation of the event is",e)
n
```

R code Exa 5.13 Discrete Probability Distribution

```
gain<-c(98,48,23,8,-2)#win and lose
pr<-c(0.002,0.002,0.002,0.002,0.992)
e<-weighted.mean(gain,pr)
n<-paste("the expectation of the event is",e)
n
```

R code Exa 5.14 Bond Investment

```
principle<-c(5000)
```

```
r1<-c(0.04)
r2<-c(0.025)
dr1<-c(0.02)
dr2<-c(0.01)
a<-((principle*r1)*(1-dr1))-(principle*dr1)
b<-((principle*r2)*(1-dr2))-(principle*dr2)
a
b
if (a>b)
  s<-paste("Bond X brings more returns")
if(a<b)
  s<-paste("Bond Y brings more returns")
s
```

R code Exa 5.16 Survey on Doctor Visits

```
n<-10
x<-n-3
q<-4/5
n<-paste("Probability is",round(dbinom(x,n,q),3))
n
```

R code Exa 5.17 Survey on Employment

```
n=5
x=n-3
q=0.7
n<-paste("Probability is",round(pbinom(x,n,q),2))
n
```

R code Exa 5.19.a Find Probability

```
1 dbinom(5,20,0.05)
```

R code Exa 5.19.b Find Probability

```
1 pbinom(3,20,0.05)
```

R code Exa 5.19.c Find Probability

```
1 1-pbinom(2,20,0.05)
```

R code Exa 5.20 Find Probability

```
1 dbinom(12,15,0.7)
```

R code Exa 5.24 Leisure Activities

```
n=5
x1=3
x2=1
x3=1
p1=0.5
p2=0.3
p3=0.2
P=dmultinom(c(x1,x2,x3),size = 5,prob = c(p1,p2,p3))
n<-paste("Probability is",P)
n
```

R code Exa 5.25 CD Purchases

```
n=6
x1=1
x2=3
x3=2
p1=0.3
p2=0.6
p3=0.1
P=dmultinom(c(x1,x2,x3),size = 6,prob = c(p1,p2,p3))
n<-paste("Probability is",P)
n
```

R code Exa 5.26 Selecting Colored Balls

```
n=5
x1=2
x2=2
x3=1
p1=0.4
p2=0.3
p3=0.3
```

```
P=dmultinom(c(x1,x2,x3),size = 5,prob = c(p1,p2,p3))
n<-paste("Probability is",P)
n
#answer is in decimals
```

R code Exa 5.27 Typographical Errors

```
np=500
ne=200
x=3
l=ne/np
P=dpois(x,l)
n<-paste("Probability that there are three errors in
    a page is less than a percentage of",round(P*
   100))
n
```

R code Exa 5.28.a Find Probability

```
x<-c(0,1,2,3)
l<-3
f<-dpois(x,l)
sum(f)
```

R code Exa 5.28.b Find Probability

```
x<-c(0,1,2)
l<-3
f<-dpois(x,l)
round(1-sum(f),4)
```

R code Exa 5.28.c Find Probability

```
x<-c(0,1,2,3,4)
l<-3
f<-dpois(x,l)
1-sum(f)
```

R code Exa 5.29 Left Handed People

```
np=0.02
ne=200
x=5
l=ne*np
P=dpois(x,l)
n<-paste("Probability is",round(P,4))
n
```

Chapter 6

The Normal Distribution

R code Exa 6.7 Monthly Newspaper Recycling

```
X1=27
X2=31
X3=30.2
m=28
sd=2
z3=round((X3-m)/sd,2)
z1=round((X1-m)/sd,2)
```

```
g<-paste(round((1-pnorm(z3))*100,2),"% is the
    probability that a randomly selected household
    generates More than 30.2 pounds per month")
g
z2=round((X2-m)/sd,2)
g<-paste(round((pnorm(z2)-pnorm(z1))*100,2),"% is
    the probability that a randomly selected
    household generates between 27 and 31 pounds of
    newspapers per month")
g
```

R code Exa 6.11 Checking for normal distribution

```
g<-c
    (5,29,34,44,45,63,68,74,74,81,88,91,97,98,113,118,151,158)

#since the formula for skewness in the inbuilt is
    not the same as in textbook, using manual methods
sk=3*(mean(g)-median(g))/sd(g)
l<-paste("Skewness is",round(sk,3))
l
iqr<-IQR(g,type = 2)
q1<-quantile(g,0.25,type = 2)
q3<-quantile(g,0.75, type = 2)
r1=q1-(1.5*iqr)
r2=q3+(1.5*iqr)
if(max(g)>r2){lb<-paste(max(g),"is a outliner");
lb}else{print("no outliner")}
```

R code Exa 6.12 Checking for normal distribution

```
g<-c
   (81,148,152,135,151,152,159,142,34,162,130,162,163,143,67,112,70)

#since the formula for skewness in the inbuilt is
   not the same as in textbook, using manual methods
sk=3*(mean(g)-median(g))/sd(g)
l<-paste("Skewness is",round(sk,3))
l
iqr<-IQR(g,type = 2)
q1<-quantile(g,0.25,type = 2)
q3<-quantile(g,0.75, type = 2)
r1=q1-(1.5*iqr)
r2=q3+(1.5*iqr)
if(max(g)>r2){lb<-paste(max(g),"is a outliner");
lb}else{print("no outliner")}
```

R code Exa 6.15 Meat Consumption

```
x=224
m=218.4
s=25
n=40
z1=(x-m)/s
l<-paste("the probability that a person selected at
   random consumes less than 224 pounds per year is"
   ,round(pnorm(z1)*100,1),"%")
l
z2=(x-m)/(s/sqrt(n))
l<-paste("the probability that the mean of a sample
```

```
    of 40 individuals is less than 224 pounds per
    year is",round(pnorm(z2)*100,1),"%")
10 1
```

Chapter 7

Confidence Intervals and Sample Size

R code Exa 7.1 Days it takes to sell an Aveo

```
1 m=54
2 z=round(qnorm(0.975),2)
3 z
4 s=6
5 n=50
6 x=round((z*s)/sqrt(n),2)
7 l<-paste("Hence one can say with 95% confidence that
      the interval between",m-x,"and",m+x,"days does
     contain the population mean, based on a sample of
      50 automobiles.")
8 l
```

R code Exa 7.2 Ages of Automobiles

```
1 m=5.6
2 z=round(qnorm(0.995),2)
```

```
s=0.8
n=30
x=round((z*s)/sqrt(n),1)
l<-paste(" one can be 99% confident that the mean
    age of all primary vehicles is between",m-x,"and"
    ,m+x,"years, based on 30 vehicles.")
l
```

R code Exa 7.3 Credit Union Assets

```
da<-c(12.23, 16.56, 4.39, 2.89, 1.24, 2.17, 13.19,
    9.16, 1.42, 73.25, 1.91, 14.64, 11.59, 6.69,
    1.06, 8.74, 3.17, 18.13, 7.92, 4.78, 16.85,
    40.22, 2.42, 21.58, 5.01, 1.47, 12.24, 2.27,
    12.77, 2.76)
m=round(mean(da),2)
s=round(sd(da),2)
c=0.9
a=1-c
z=round(qnorm(0.951),2)
n=30
x=round((z*s)/sqrt(n),3)
l<-paste("Hence, one can be 90% confident that the
    population mean of the assets of all credit
    unions is between",m-x,"and",m+x,"based on a
    sample of 30 credit unions")
l
```

R code Exa 7.4 Depth of a River

```
c=0.99
a=1-c
s=4.38
```

```
z=round(qnorm(0.995),2)
z
E=2
ss=(z*s/E)^2
l<-paste("to be 99% confident that the estimate is
    within 2 feet of the true mean depth, the
    scientist needs at least a sample of",round(ss),"
    measurements.")
l
```

R code Exa 7.5 Critical t value

```
round(qt(0.975,21),2)
```

R code Exa 7.6 Sleeping Time

```
t=round(qt(0.975,9),3)
c=0.95
s=0.78
n=10
m=7.1
lo=round(m-(t*(s/sqrt(n))),2)
hi=round(m+(t*(s/sqrt(n))),2)
l<-paste("Therefore, one can be 95% confident that
    the population mean is between",lo,"and",hi,"
    inches")
l
```

R code Exa 7.7 Home Fires Started by Candles

```
t=qt(0.995,6)
s=1610.3
n=7
m=7041.4
lo=round(m-(t*(s/sqrt(n))))
hi=round(m+(t*(s/sqrt(n))))
l<-paste("Therefore, one can be",c*100,"confident
   that the population mean is between",lo,"and",hi)
l
```

R code Exa 7.9 Male Nurses

```
x=60
n=500
p=round((x/n),2)
q=1-p
z=round(qnorm(0.951),2)
lo=round((p-(z*sqrt(p*q/n)))*100,1)
hi=round((p+(z*sqrt(p*q/n)))*100,1)
l<-paste("Hence, you can be 90% confident that the
   percentage of applicants who are men is between",
   lo,"and",hi,"%")
l
```

R code Exa 7.10 Religious Books

```
n=1721
p=0.159
q=1-p
z=round(qnorm(0.975),2)
lo=round((p-(z*sqrt(p*q/n)))*100,1)
```

```
6 hi=round((p+(z*sqrt(p*q/n)))*100,1)
7 l<-paste("Hence, you can be 95% confident that the
     true percentage is between",lo,"and",hi,"%")
8 l
```

R code Exa 7.11 Home Computers

```
1 z=round(qnorm(0.975),2)
2 E=0.02
3 p=0.40
4 q=1-p
5 n=p*q*((z/E)^2)
6 l<-paste("Minimum sample space required is",round(n)
     )
7 l
```

R code Exa 7.12 Car Phone Ownership

```
1 z=round(qnorm(0.951),2)
2 E=0.05
3 p=0.50
4 q=1-p
5 n=p*q*((z/E)^2)
6 if(n>round(n))
7   n=n+1
8 l<-paste("Minimum sample space required is",round(n)
     )
9 l
```

R code Exa 7.14 Nicotine Continent

```
n=20
s=1.6
Xr=round(qchisq(0.975,19),2)
Xl=round(qchisq(0.025,19),2)
lo=(n-1)*(s^2)/Xr
lo=sqrt(lo)
hi=(n-1)*(s^2)/Xl
hi=sqrt(hi)
l<-paste("Hence, you can be 95% confident that the
    true standard deviation for the nicotine content
    of all cigarettes manufactured is between",round(
    lo,1),"and",round(hi,1),"milligrams based on a
    sample of 20 cigarettes.")
l
```

R code Exa 7.15 Cost of Ski Lift Tickets

```
n=10
m<-c(59, 54, 53, 52, 51, 39, 49, 46, 49, 48)
s=sd(m)
Xr=round(qchisq(0.95,9),3)
Xl=round(qchisq(0.05,9),3)
lo=(n-1)*(s^2)/Xr
lo=sqrt(lo)
hi=(n-1)*(s^2)/Xl
hi=sqrt(hi)
l<-paste("Hence, you can be 90% confident that the
    true standard deviation price of all single-day
    ski lift tickets of the population is between",
    round(lo,2),"and",round(hi,2),"dollars on a
    sample of 10 nationwide ski resorts.")
l
```

Chapter 8

Hypothesis Testing

R code Exa 8.3 Professors Salaries

```
x1=43260
m=42000
s=5230
n=30
c=round(qnorm(0.951),2)
z=round((x1-m)/(s/sqrt(n)),2)
f=z
x=seq(-4,4,length=200)
y=dnorm(x)
plot(x,y,type="l", lwd=2, col="green")
if(z<c)
{ f=c
  x=seq(z,z,length=100)
  y=dnorm(x)
  polygon(c(z,x,z),c(0,y,0),col="black")}
```

```
16 x=seq(f,4,length=100)
17 y=dnorm(x)
18 polygon(c(f,x,4),c(0,y,0),col="gray")
19 #values are compared on the plot
```

R code Exa 8.4 Costs of Mens Athletic Shoes

```
1 g<-c(60, 70, 75, 55, 80, 55, 50, 40, 80, 70, 50, 95,
     120, 90, 75, 85, 80, 60, 110, 65, 80, 85, 85,
    45, 75, 60, 90, 90, 60, 95, 110, 85, 45, 90, 70,
    70)
2 t.test(g, mu=80, conf.level = 0.9)
```

R code Exa 8.5 Cost of Rehabilitation

```
1 m=25226
2 mu=24672
3 s=3251
4 n=35
5 c=round(qnorm(0.995),2)
6 z=round((m-mu)/(s/sqrt(n)),2)
7 x=seq(-4,4,length=200)
8 y=dnorm(x)
9 plot(x,y,type="l", lwd=2, col="green")
10 x=seq(z,z,length=100)
11 y=dnorm(x)
12 polygon(c(z,x,z),c(0,y,0),col="black")
13 x=seq(-c,-4,length=100)
14 y=dnorm(x)
15 polygon(c(-c,x,-4),c(0,y,0),col="yellow")
16 x=seq(c,4,length=100)
17 y=dnorm(x)
18 polygon(c(c,x,4),c(0,y,0),col="yellow")
```

```
19 #values are compared on the plot
```

R code Exa 8.13 Substitute Teachers Salaries

```
1 g<-c(60, 56, 60, 55, 70, 55, 60, 55)
2 t.test(g, mu=60, conf.level = 0.9)
```

R code Exa 8.17 people who are trying to avoid Trans Fats

```
1 mu=0.60
2 n=200
3 x=128
4 p=x/n
```

```
q=1-mu
c=1.96
z=(p-mu)/sqrt(mu*q/n)
if((z>-c)&&(z<c))
  l<-paste("Hypothesis rejected") else
    l<-paste("Hypothesis accepted")
l
```

R code Exa 8.18 Survey on Call Waiting Service

```
mu=0.40
n=100
p=0.37
q=1-mu
c=qnorm(0.995)
z=(p-mu)/sqrt(mu*q/n)
if((z>-c)&&(z<c))
  l<-paste("Hypothesis rejected") else
    l<-paste("Hypothesis accepted")
l
```

R code Exa 8.23 critical chisq value

```
l<-paste(round(qchisq(0.975,22),3),"and",round(
   qchisq(0.025,22),3),"is the critical value")
```

```
2 l
```

R code Exa 8.25 Outpatient Surgery

```
1 g<-c(25, 30, 5, 15, 18, 42, 16, 9, 10, 12, 12, 38,
     8, 14, 27)
2 library(EnvStats)
3 varTest(g,alternative = "greater",conf.level = 0.9,
     sigma.squared = 64)
4 #chisq.test didn't give the correct output for me.
     Hence i have opted for varTest which essentially
     performs the same on variance as given in the
     question.
5 #please install package "EnvStats" before proceeding
```

R code Exa 8.30 Sugar Production

```
1 mu1=5
2 n=50
3 x=4.6
4 s=0.7
5 sq=round(sqrt(n),3)
6 z=(x-mu1)*sq/s
```

```
c1=-round(qnorm(0.975),2)
c2=round(qnorm(0.975),2)
if(z<c1 || z>c2)
  l<-paste("Null hypothesis rejected")
l
m1=x-(c2*s/sq)
m2=x+(c2*s/sq)
p<-paste("confidence interval is between",round(m1
   ,1),"and",round(m2,1))
p
```

R code Exa 8.31 Hog Weight

```
mu1=200
n=10
x=198.2
s=3.3
sq=round(sqrt(n),4)
z=(x-mu1)*sq/s
c2=round(-qt(0.025,9),3)
c1=-round(-qt(0.025,9),3)
if(z>c1 || z<c2)
  l<-paste("Null hypothesis accepted")else
    l<-paste("Null hypothesis rejected")
l
m1=x-round((c2*s/sq),3)
m2=x+round((c2*s/sq),3)
p<-paste("confidence interval is between",round(m1
   ,1),"and",round(m2,1))
p
```

Chapter 9

Testing the Difference Between Two Means and Two Proportions and Two Variances

R code Exa 9.1 Hotel Room Cost

```
c1=-round(qnorm(0.975),2)
c2=round(qnorm(0.975),2)
s1=5.62*5.62
s2=4.83*4.83
n1=n2=50
x1=88.42
x2=80.61
z=(x1-x2)/sqrt(s1/n1 + s2/n2)
if(z>c2 || z<c1)
  l<-paste("Null hypothesis rejected. There is
     enough evidence to support the claim that the
     means are not equal. Hence, there is a
     significant difference in the rates.")
l
```

R code Exa 9.3 Confidence Interval

```
c1=-round(qnorm(0.975),2)
c2=round(qnorm(0.975),2)
s1=5.62*5.62
s2=4.83*4.83
n1=n2=50
x1=88.42
x2=80.61
z=1.96
lo=(x1-x2)-(z*sqrt(s1/n1 + s2/n2))
hi=(x1-x2)+(z*sqrt(s1/n1 + s2/n2))
p<-paste("Confidence interval is between",round(lo
   ,2),"and",round(hi,2))
if(lo>0)
  l<-paste("Since the confidence interval does not
     contain zero, the decision is to reject the
     null hypothesis, which agrees with the previous
      result.")else
  l<-paste("Hypothesis accepted")
p
l
```

R code Exa 9.4 Farm Sizes

```
c1=-round(qt(0.975,7),3)
c1
c2=round(qt(0.975,7),3)
s1=38*38
s2=12*12
```

```
n1=8
n2=10
x1=191
x2=199
z=(x1-x2)/sqrt(s1/n1 + s2/n2)
z
if(z>c2 || z<c1)
  l<-paste("Null hypothesis rejected.") else
    l<-paste("Null hypothesis accepted.There is not
       enough evidence to support the claim that the
        average size of the farms is different. ")
l
#qf function with alpha=0.05 and df=7 doesnt give
   2.365
```

R code Exa 9.5 Confidence Interval

```
c2=round(qt(0.975,7),3)
c2
s1=38*38
s2=12*12
n1=8
n2=10
x1=191
x2=199
lo=(x1-x2)-(c2*sqrt(s1/n1 + s2/n2))
hi=(x1-x2)+(c2*sqrt(s1/n1 + s2/n2))
p<-paste("Confidence interval is between",round(lo
   ,2),"and",round(hi,2))
if(lo<0 && hi>0)
  l<-paste("Since 0 is contained in the interval,
     the decision is to not reject the null
     hypothesis")else
    l<-paste("Hypothesis rejected")
p
```

```
16 l
17 #qf doesn't give the exact answer as in the tb
```

R code Exa 9.6 Vitamin for Increased Strength

```
1 X1<-c(210,230,182,205,262,253,219,216)
2 X2<-c(219,236,179,204,270,250,222,216)
3 t.test(X1,X2,paired =  TRUE,alternative = "less")
```

R code Exa 9.7 Cholesterol Levels

```
1 X1<-c(210,235,208,190,172,244)
2 X2<-c(190,170,210,188,173,228)
3 t.test(X1,X2, paired =  TRUE)
```

R code Exa 9.8 Confidence Interval

```
1 n=6
2 mu=0
3 c=2.015
4 X1<-c(210,235,208,190,172,244)
5 X2<-c(190,170,210,188,173,228)
6 t.test(X1,X2, paired = TRUE, conf.level = 0.90)
```

R code Exa 9.9 Vaccination Rates in Nursing Homes

```
x1=12
n1=34
x2=17
n2=24
p1=x1/n1
p2=x2/n2
p=(x1+x2)/(n1+n2)
q=1-p
c1=round(qnorm(0.975),2)
c2=-round(qnorm(0.975),2)
z=round((p1-p2)/sqrt(p*q*(1/n1+1/n2)),1)
if(z>c1 || z<c2)
  l<-paste("There is enough evidence to reject the
     claim that there is no difference in the
     proportions of small and large nursing homes
     with a resident vaccination rate of less than
     80%.") else
    l<-paste("Null hypothesis accepted")
l
```

R code Exa 9.10 Missing Work

```
p1=0.45
n1=200
p2=0.35
n2=200
x1=p1*n1
x2=p2*n2
p=(x1+x2)/(n1+n2)
q=1-p
c1=round(qnorm(0.995),2)
c2=-round(qnorm(0.995),2)
z=round((p1-p2)/sqrt(p*q*(1/n1+1/n2)),1)
if(z>c1 || z<c2)
  l<-paste("Null hypothesis rejected") else
```

```
14      l<-paste("Null hypothesis accepted. There is not
            enough evidence to support the claim that
           there is a difference in proportions.")
15  l
```

R code Exa 9.11 Confidence Interval

```
1   x1=12
2   n1=34
3   x2=17
4   n2=24
5   p1=round(x1/n1,2)
6   q1=1-p1
7   p2=round(x2/n2,2)
8   q2=1-p2
9   p=(x1+x2)/(n1+n2)
10  q=1-p
11  c1=round(qnorm(0.975),2)
12  lo=(p1-p2)-(c1*sqrt((p1*q1)/n1 + (p2*q2)/n2))
13  hi=(p1-p2)+(c1*sqrt((p1*q1)/n1 + (p2*q2)/n2))
14  p<-paste("Confidence intervals are in between",round
       (hi,3),"and",round(lo,3))
15  p
16  if(lo<0 && hi<0)
17    l<-paste("Since 0 is not contained in the interval
         , the decision is to reject the null hypothesis
        ") else
18      l<-paste("Null hypothesis accepted")
19  l
```

Chapter 10

Correlation and Regression

R code Exa 10.12 Copy Machine Maintenance Costs

```
1 c<-c(1,2,3,4,4,6)
```

```
2 r<-c(62,78,70,90,93,103)
3 mo<-lm(r ~ c)
4 mo
5 summary(mo)
```

R code Exa 10.14 Standard error estimation

```
1 c<-c(1,2,3,4,4,6)
2 r<-c(62,78,70,90,93,103)
3 mo<-lm(r~c)
4 mo
5 predict(mo,interval = "prediction", level = 0.95)
```

R code Exa 10.15 State board scores

```
1 gpa<-c(3.2,2.7,2.5,3.4,2.2)
2 x<-c(22,27,24,28,23)
3 sbs<-c(550,570,525,670,490)
4 dt<-data.frame("Student"=c("a","b","c","d","e"),"GPA
    "=gpa,"Age"=x,"State board scores"=sbs)
5 mo<-lm(dt$State.board.scores ~ dt$GPA +dt$Age)
6 mo
7 summary(mo)
```

R code Exa 10.16 State board scores

```
1 gpa<-c(3.2,2.7,2.5,3.4,2.2)
2 x<-c(22,27,24,28,23)
```

```
3 sbs<-c(550,570,525,670,490)
4 dt<-data.frame("Student"=c("a","b","c","d","e"),"GPA
    "=gpa,"Age"=x,"State board scores"=sbs)
5 mo<-lm(dt$State.board.scores ~ dt$GPA +dt$Age)
6 mo
7 summary(mo)
```

R code Exa 10.17 State board scores

```
1 gpa<-c(3.2,2.7,2.5,3.4,2.2)
2 x<-c(22,27,24,28,23)
3 sbs<-c(550,570,525,670,490)
4 dt<-data.frame("Student"=c("a","b","c","d","e"),"GPA
    "=gpa,"Age"=x,"State board scores"=sbs)
5 mo<-lm(dt$State.board.scores ~ dt$GPA +dt$Age)
6 mo
7 summary(mo)
```

Chapter 11

Other Chi Square Tests

R code Exa 11.1 Fruit Soda Flavor Preference

```
o<-c(32,28,16,14,10)
chisq.test(o)
```

R code Exa 11.5 College Education and Place of Residence

```
u<-c(15,12,8,8,15,9,6,8,7)
m<-matrix(u,ncol=3,byrow = TRUE)
colnames(m)<-c("no college", "four year degree", "
   advanced college")
rownames(m)<-c("Urban","Suburban","Rural")
m<-as.table(m)
m
chisq.test(m)
```

R code Exa 11.6 Alcohol and Gender

```
u<-c(10,9,8,13,16,12)
m<-matrix(u,ncol=3,byrow = TRUE)
colnames(m)<-c("low", "moderate", "high")
rownames(m)<-c("Male","Female")
m<-as.table(m)
m
chisq.test(m)
```

R code Exa 11.7 Lost Luggage on Airline Flights

```
u<-c(10,7,4,90,93,96)
m<-matrix(u,ncol=3,byrow = TRUE)
colnames(m)<-c("Airline 1", "Airline 2", "Airline 3"
   )
rownames(m)<-c("Yes","No")
m<-as.table(m)
m
chisq.test(m)
```

Chapter 12

Analysis of Variance

R code Exa 12.1 Lowering Blood Pressure

```
1 x<-data.frame("meal"=c(10,12,9,15,13),"exercise"=c
    (6,8,3,0,2),"diet"=c(5,9,12,8,4))
2 mo<-c(x$meal,x$exercise,x$diet)
3 gr<-factor(rep(letters[1:3],each = 5))
4 fit=aov(formula = mo ~ gr)
5 anova(fit)
```

R code Exa 12.2 Employees at Toll Road Interchanges

```
1 x<-data.frame("tp"=c(7,14,32,19,10,11),"mf"=c
    (10,1,1,0,11,1),"bv"=c(1,12,1,9,1,11))
2 mo<-c(x$tp,x$mf,x$bv)
3 gr<-factor(rep(letters[1:3],each = 6))
4 fit=aov(formula = mo ~ gr)
5 anova(fit)
```

R code Exa 12.3 Scheffe Test

```
me<-c(10,12,9,15,13)
ex<-c(6,8,3,0,2)
di<-c(5,9,12,8,4)
l<-c(me,ex,di)
fa<-factor(rep(c("meat","exercise","diet"),15))
pr=factor(rep(1:15, rep(3, 15)))
d<-data.frame(l,fa,pr)
fml<- aov(d$l ~ d$fa + d$pr, data = d)
library(DescTools)
ScheffeTest(x=fml)
```

R code Exa 12.4 Turkey Test

```
me<-c(10,12,9,15,13)
ex<-c(6,8,3,0,2)
di<-c(5,9,12,8,4)
l<-c(me,ex,di)
fa<-factor(rep(c("meat","exercise","diet"),15))
pr=factor(rep(1:15, rep(3, 15)))
d<-data.frame(l,fa,pr)
a1<- aov(d$l ~ d$fa + d$pr)
TukeyHSD(x=a1,"d$fa", conf.level=0.95)
```

R code Exa 12.5 Gasoline Consumption

```
ss<-c(26.7,25.2,32.3,32.8,28.6,29.3,26.1,24.2)
gas<-factor(c("reg","reg","ho","ho","reg","reg","ho"
    ,"ho"))
auto<-factor(c("tw","tw","tw","tw","fw","fw","fw","
    fw"))
```

```
4 dat<-data.frame("value"=ss,"Gas"=gas,"Automobile"=
    auto)
5 dat
6 results<-lm(dat$value ~ dat$Gas + dat$Automobile +
    dat$Gas*dat$Automobile, data=dat)
7 anova(results)
```

Chapter 13

Nonparametric Statistics

R code Exa 13.1 snow Cone Sales

```
1 x<-c
    (18,43,40,16,22,30,29,32,37,36,39,34,39,45,28,36,40,34,39,52)

2 a=0.05
3 SIGN.test(x)
4 #package BSDA to be downloaded for SIGN.test
```

R code Exa 13.3 Ear infections in Swimmers

```
1 x<-c(3,0,5,4,2,4,3,5,2,1)
2 y<-c(2,1,4,0,1,3,1,3,2,3)
3 SIGN.test(x,y)
4 #package BSDA to be downloaded for SIGN.test
```

R code Exa 13.6 Milliequivalents of Potassium in Breakfast Drinks

```
a<-c(4.7,3.2,5.1,5.2,5.0)
b<-c(5.3,6.4,7.3,6.8,7.2)
c<-c(6.3,8.2,6.2,7.1,6.6)
kruskal.test(list(a,b,c))
```

R code Exa 13.10 Ages of Drug Program Participants

```
d<-c(18, 36, 19, 22, 25, 44, 23, 27, 27, 35, 19, 43,
    37, 32, 28, 43, 46, 19, 20, 22)
runs.test(d)
l<-paste("Since there are 9 runs between 5 and 15,
   the null hypothesis is not rejected.")
l
#snpar package to be downloaded for runs.test
```