Statistical Methods Assignment

Greeshma Jeev Koothuparambil

2023-11-06

This report is submitted by Greeshma Jeev Koothuparambil as a part of Statistical Methods (732A93) Course for the 2023 Autumn Semester.

Monte Carlo Methods

1-Derive a formula and explain how to generate a random variable with the density f(x) = (1.5)√x for 0< x < 1 if your random number generator produces a Standard Uniform random variable U. Use the inverse transform method. Compute this variable if U = 0.001.

```
##1.1
f <- function(u){
  return(u^(2/3))
}


U= 0.001
X = f(U)
```

Analysis *The given was a density function: f(x) = 1.5 X $\sqrt{x}$. This is converted to a cumulative density function through integration.*

$$\int 1.5 * \sqrt{x}$$

$$= \int 1.5 * x^{0.5}$$

$$= \frac{1.5 * x^{3.5}}{1.5}$$

$$U = x^{3.5}$$

$$x = U^{\frac{2}{3}}$$

When The value for U = 0.001 is substituted in the equation we get x = 0.01

2-Twenty computers are connected in a network. One computer becomes infected with a virus. Every day, this virus spreads from any infected computer to any uninfected computer with probability 0.1. Also, every day, a computer technician takes 5 infected computers at random (or all infected computers, if their number is less than 5) and removes the virus from them. Estimate:
(a) the expected time it takes to remove the virus from the whole network;

```
## 1.2
N <- 1000

#Probability of infection
```

```r
p <- 0.1

#Number of Computers
C <- 20

#expected time to disinfect the network function
expectedtime <- function(comp, pro, sim){
  totaldays <- 0

  for (i in 1:sim) {
    infected <- 1
    days <-0
    while(infected >0){

      randinf <- runif(comp-infected)
      infected <- infected + length(which(randinf<pro))
      if(infected >0){
        infected <- infected -5
      }
      days <- days+1
      totaldays <- totaldays+days
    }


  }
  expected <- totaldays/sim
  return(expected)
}
```

(b) the probability that each computer gets infected at least once;

```r
#probability of systems getting infected atlest once function
infectionprobability <- function(comp, pro, sim){


  proiter <- 0
  for(i in 1:sim){
    iteration <- 0
    infected <- rep(0,20)
    atleastonce <- rep(0,20)
    #t <- ceiling(runif(1,min = 1, max = comp))
    #infected[t] <-1
    while(length(which(atleastonce==0))>0){

      randinf <- runif(comp-length(which(infected>0)))
      infected[which(randinf<pro)] <- 1
      atleastonce[which(randinf<pro)] <- atleastonce[which(randinf<pro)]+1
      if(length(which(infected>0))>5){
        maximum = 5
      }else{
        maximum =length(which(infected>0))
      }
      t <- sample(which(infected>0),maximum)
      infected[t] <- 0
```

2

```
        iteration <- iteration+1

    }
    proiter <- proiter +1/iteration
  }
  probabilities <- proiter/sim
  return(probabilities)
}
```

(c) the expected number of computers that get infected.

```
#expected number of infected computers
expectedcomp <- function(comp, pro, sim){
  compnum <- 0

  for (i in 1:sim) {
    infected <- 1

    while(infected >0){

      randinf <- runif(comp-infected)
      infected <- infected + length(which(randinf<pro))
      compnum <- compnum + infected

      if(infected >0){
        infected <- infected -5
      }
    }


  }
  expectedcompnum <- compnum/sim
  return(expectedcompnum)
}
```

The functions are called as follows:

```
expecteddisinfecttime <- expectedtime(C, p, N)
atleastonceprobabilty <- infectionprobability(C, p, N)
expectedinfectedcomp <- expectedcomp(C, p, N)
```

Analysis   *The expected disinfecting time is 1.068 days.*
The probability that every system gets infected atleast once is 0.0314172.
The expected number of infected systems is 3.071.
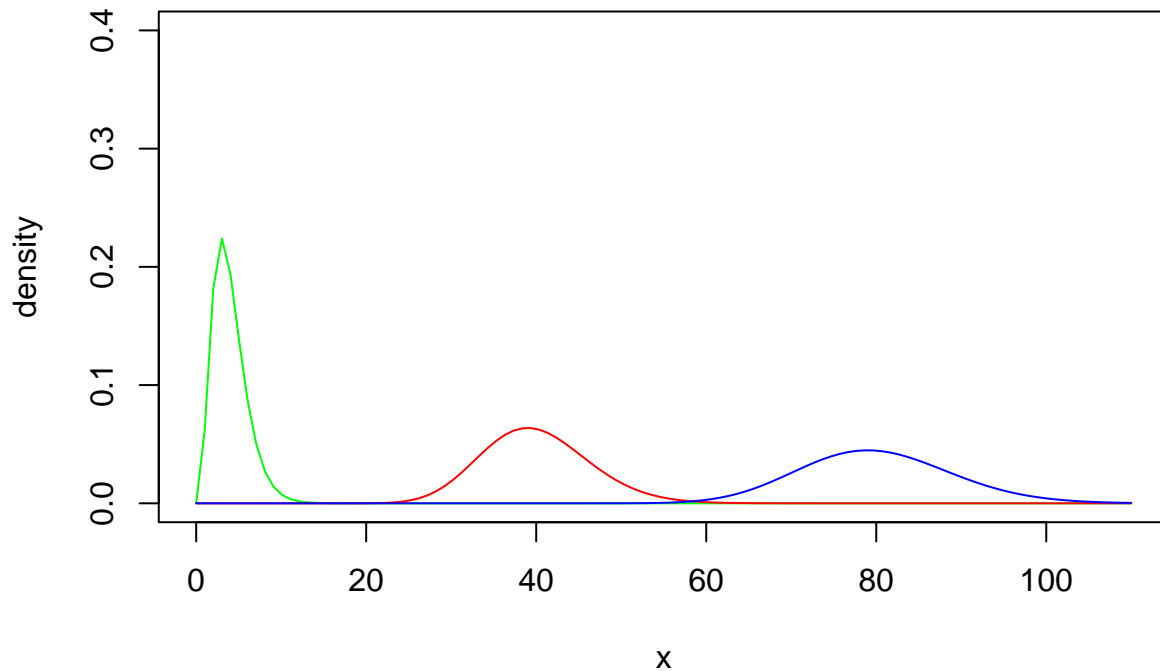

     Computer Exercises from Course book

4.84

Comparison of Gamma Density Functions to compare gamma density functions with ($\alpha = 4$, $\beta = 1$), ($\alpha = 40$, $\beta = 1$), and ($\alpha = 80$, $\beta = 1$).

```
library(plotly)
library(ggplot2)
x = seq(0, 110, length=110)
```

```
plot(x, dgamma(x, 4, 1), ylab='density',ylim = c(0,.4) ,type ='l', col='green')
lines(x, dgamma(x, 40, 1), col='red')
lines(x, dgamma(x, 80, 1), col='blue')
```


a What do you observe about the shapes of these three density functions? Which are less skewed and more symmetric?

Analysis *The x value is from 0 to 110. From the graph the gamma(4,1) seems to be left skewed while gamma(80,1) seems right skewed and gamma(40,1) appears to be less skewed and balanced. We can also see that with the increase of the alpha value the shape of the bell widens.*

b What differences do you observe about the location of the centers of these density functions? *The centers of all the gamma distribution decreases in an exponential way. The difference in peaks of gamma(4,1) and gamma(40,1) was too high while difference of peaks in gamma(40,1) and gamma(80,1) seems less.*

c Give an explanation for what you observed in part (b). *The gamma function is as follows:*

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

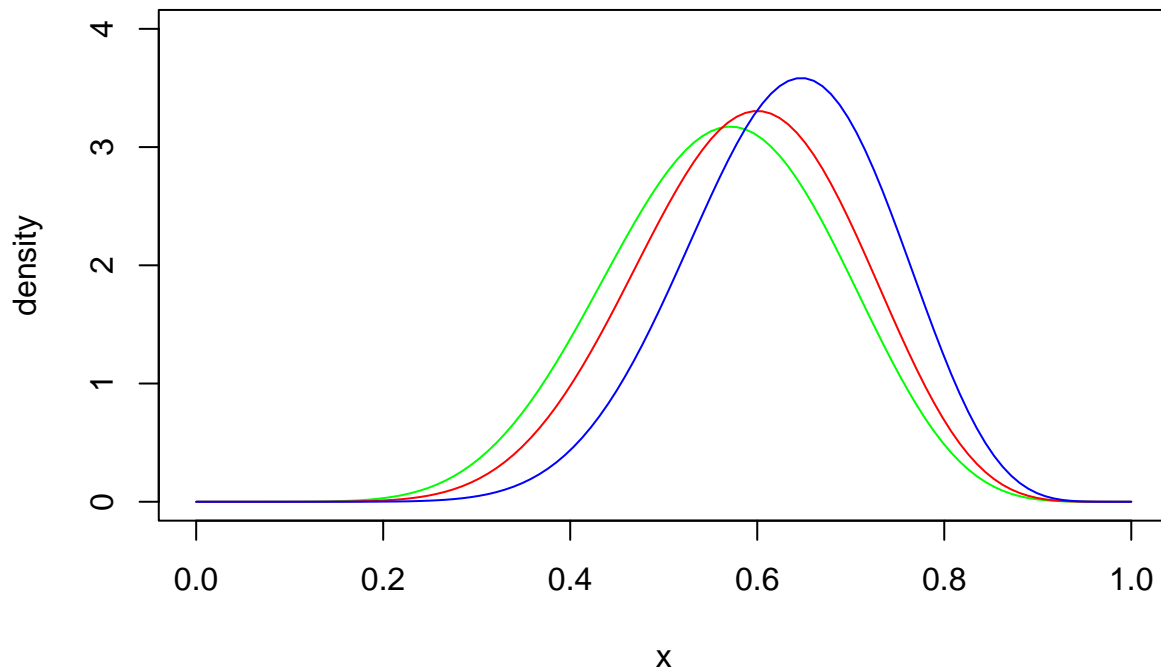*Since the function has an exponential trend the distribution plot will also exhibit that nature*

4.117

Comparison of Beta Density Functions to compare beta density functions with ($\alpha = 9$, $\beta = 7$), ($\alpha = 10$, $\beta = 7$), and ($\alpha = 12$, $\beta = 7$).

```
##4.117
x = seq(0, 1, length=110)

plot(x, dbeta(x, 9, 7), ylab='density',ylim = c(0,4) ,type ='l', col='green')
```

```
lines(x, dbeta(x, 10, 7), col='red')
lines(x, dbeta(x, 12, 7), col='blue')
```


a Are these densities symmetric? Skewed left? Skewed right?

Analysis   *All the densities seem skewed to the right.*

b What do you observe as the value of α gets closer to 12?

Analysis   *As the alpha value grows higher an evident increase in the peak value of each density is visible. We can notice the width of the peak reducing as well*
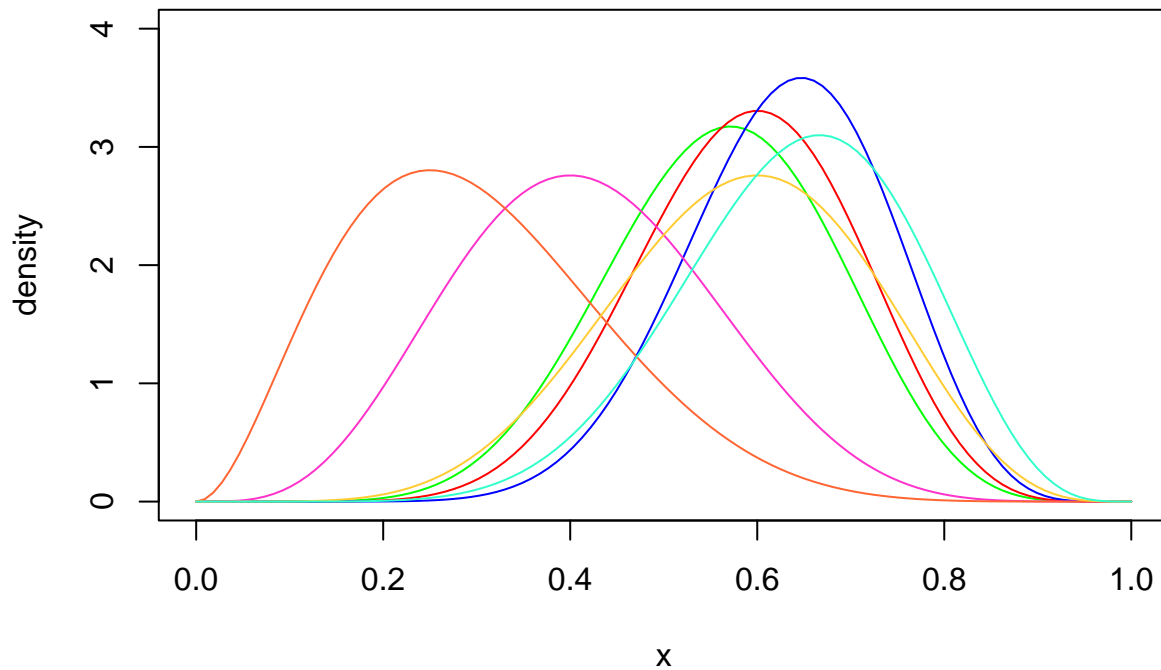
c Graph some more beta densities with $\alpha > 1$, $\beta > 1$, and $\alpha > \beta$. What do you conjecture about the shape of beta densities with $\alpha > \beta$ and both $\alpha > 1$ and $\beta > 1$?

```
##4.117
x = seq(0, 1, length=110)

plot(x, dbeta(x, 9, 7), ylab='density',ylim = c(0,4) ,type ='l', col='green')
lines(x, dbeta(x, 10, 7), col='red')
lines(x, dbeta(x, 12, 7), col='blue')

lines(x, dbeta(x, 5, 7), col='#FF33CC')
lines(x, dbeta(x, 3, 7), col='#FF6633')

lines(x, dbeta(x, 7, 5), col='#FFCC33')
lines(x, dbeta(x, 9, 5), col='#33FFCC')
```

Analysis *For the criteria $\alpha >1$ and $\beta >1$ we took values (5, 7) and (3, 7). This criteria gave right skewing tendency to the graph. For the criteria $\alpha > \beta$ we took values (7,5) and (9,5). These plots gave us right skewed graphs. From this experiment we can derive that the skeweness of the graph depends on the value relation between α and β.*
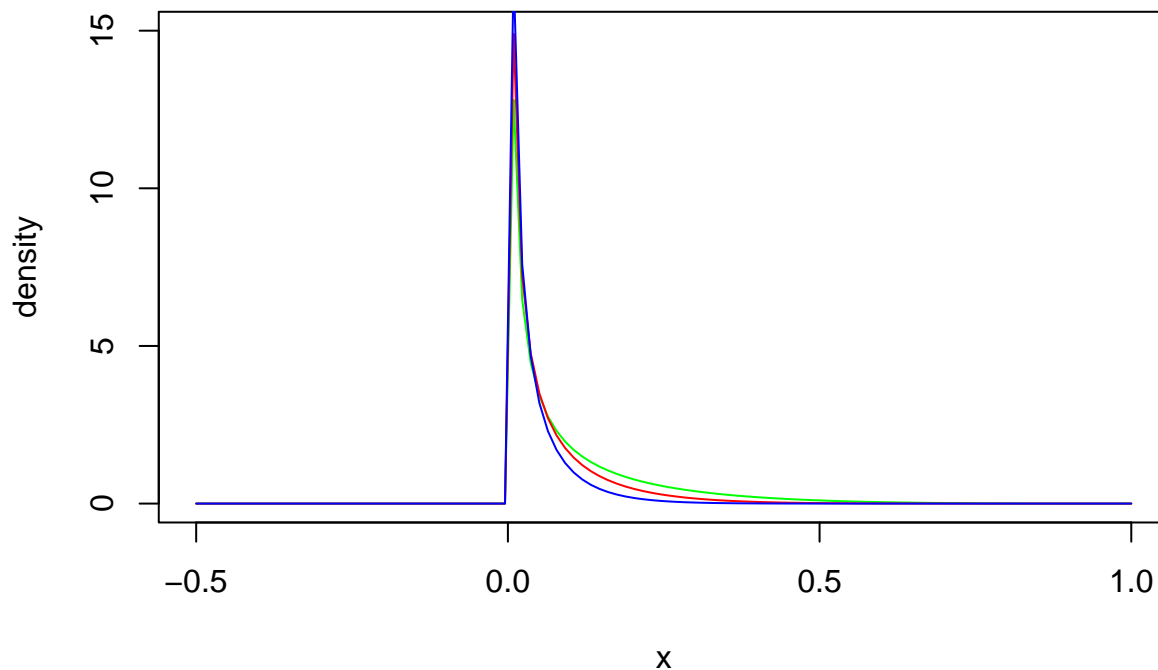

4.118

Comparison of Beta Density Functions to compare beta density functions with ($\alpha = .3$, $\beta = 4$), ($\alpha = .3$, $\beta = 7$), and ($\alpha = .3$, $\beta = 12$).

```
##4.118

x = seq(-.5, 1, length=110)

plot(x, dbeta(x, .3, 4), ylab='density',ylim = c(0,15) ,type ='l', col='green')
lines(x, dbeta(x, .3, 7), col='red')
lines(x, dbeta(x, .3, 12), col='blue')
```

a Are these densities symmetric? Skewed left? Skewed right?

Analysis *The densities are not symmetric. They are skewed to left.*

b What do you observe as the value of β gets closer to 12?

Analysis *The dip of the densities grow more as β grows larger.*

c Which of these beta distributions gives the highest probability of observing a value larger than 0.2?

Analysis *The distributions with beta(.3, 4) shows a higher probability at 0.2*

d Graph some more beta densities with $\alpha < 1$ and $\beta > 1$. What do you conjecture about the shape of beta densities with $\alpha < 1$ and $\beta > 1$?

```
##4.118

x = seq(-.5, 1, length=110)

plot(x, dbeta(x, .3, 4), ylab='density',ylim = c(0,15) ,type ='l', col='green')
lines(x, dbeta(x, .3, 7), col='red')
lines(x, dbeta(x, .3, 12), col='blue')

lines(x, dbeta(x, .5, .7), col='#FF33CC')
lines(x, dbeta(x, 2, 9), col='#FF6633')
```
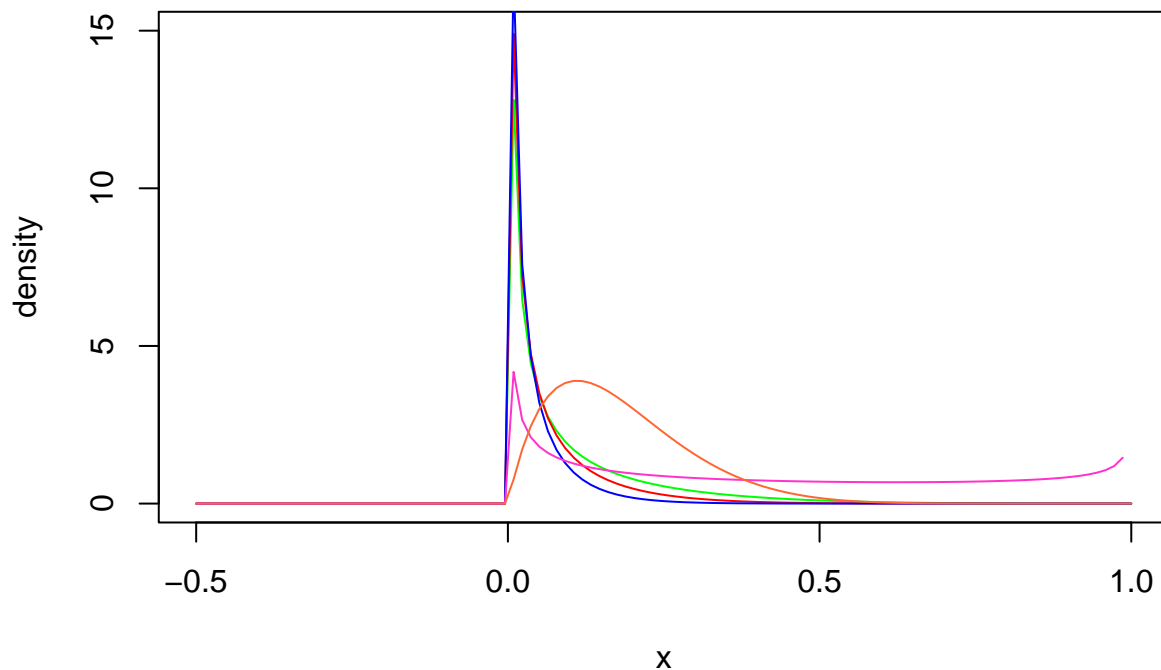
Analysis *For the criteria $\alpha < 1$ we took values (.5, .7). This criteria gave left skewing tendency to the graphbut it is lesser skewed than earlier values.For the criteria $\beta > 1$ we took values (2, 9). This plot gives us left skewed convex (bell shaped) graph.*

10.19

The output voltage for an electric circuit is specified to be 130. A sample of 40 independent readings on the voltage for this circuit gave a sample mean 128.6 and standard deviation 2.1. Test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. Use a test with level α .05.

```
##10.19

meanval <- 128.6
sdval <- 2.1
n <- 40
alpha <- 0.05



hypotest <- function(Ho, H1)
{
  teststat <- (meanval-Ho)/(sdval*sqrt(n))
  if(H1==">"){
    rejection <- teststat> qnorm(1-alpha)
  }else{
    if(H1 == "<"){
      rejection <- teststat> qnorm(1-alpha)
    }else{
      if(H1 == "!="){
        rejection <- abs(teststat)> qnorm(1-(alpha/2))
```

```
      }
    }
  }
  if(!rejection){
    retval <- "Null Hypothesis is true"
    return(retval)
  }else{
    retval <- "Null Hypothesis is rejected"
    return(retval)
  }
}

Nullhypo <- 130
Althypo <- "<"
print(paste0("Null Hypothesis is Ho : mean = ", Nullhypo))
```

```
## [1] "Null Hypothesis is Ho : mean = 130"
```

```
print(paste0("Alternative Hypothesis is H1: mean ", Althypo, Nullhypo))
```

```
## [1] "Alternative Hypothesis is H1: mean <130"
```

```
print(hypotest(Nullhypo, Althypo))
```

```
## [1] "Null Hypothesis is true"
```

10.21

Shear strength measurements derived from unconfined compression tests for two types of soils gave the results shown in the following table (measurements in tons per square foot). Do the soils appear to differ with respect to average shear strength, at the 1% significance level?

```
Soil_Type_I  <- c("n1 = 30", "y1 = 1.65", "s1 = 0.26")
Soil_Type_II <- c("n2 = 35", "y2 = 1.43", "s2 = 0.22")
dfsoil <- data.frame(Soil_Type_I,Soil_Type_II)
```

Soil_Type_I	Soil_Type_II
n1 = 30	n2 = 35
y1 = 1.65	y2 = 1.43
s1 = 0.26	s2 = 0.22

```
##10.21

y1 <- 1.65
s1 <- .26
n1 <- 30

y2 <- 1.43
s2 <- .22
n2 <- 35

alpha <- 0.1
```

9

```
meanval <- y1 -y2
deno <- sqrt((s1^2/n1)+(s2^2/n2))

hypotest <- function(Ho, H1)
{
  teststat <- (meanval-Ho)/deno
  if(H1==">"){
    rejection <- teststat> qnorm(1-alpha)
  }else{
    if(H1 == "<"){
      rejection <- teststat> qnorm(1-alpha)
    }else{
      if(H1 == "!="){
        rejection <- abs(teststat)> qnorm(1-(alpha/2))
      }
    }
  }
  if(!rejection){
    retval <- "Null Hypothesis is true"
    return(retval)
  }else{
    retval <- "Null Hypothesis is rejected"
    return(retval)
  }
}

Nullhypo <- 0
Althypo <- "!="
print(paste0("Null Hypothesis is Ho : (y1-y2) =", Nullhypo))

## [1] "Null Hypothesis is Ho : (y1-y2) =0"
print(paste0("Alternative Hypothesis is H1:(y1-y2) ",Althypo,Nullhypo))

## [1] "Alternative Hypothesis is H1:(y1-y2) !=0"
print(hypotest(Nullhypo, Althypo))

## [1] "Null Hypothesis is rejected"
```

11.31

Using a chemical procedure called differential pulse polarography, a chemist measured the peak current generated (in microamperes, μA) when solutions containing different amounts of nickel (measured in parts per billion, ppb) are added to different portions of the same buffer. 8 Is there sufficient evidence to indicate that peak current increases as nickel concentrations increase? Use $\alpha= .05$.

```
## 11.31

Ni <- c(19.1,38.2,57.3,76.2,95,114,131,150,170)
PeakCurrent <- c(0.095,0.174,0.256,0.348,0.429,0.500,0.580,0.651,0.722)

dfnp <- data.frame(Ni,PeakCurrent)
```

Ni	PeakCurrent
19.1	0.095
38.2	0.174
57.3	0.256
76.2	0.348
95.0	0.429
114.0	0.500
131.0	0.580
150.0	0.651
170.0	0.722

```r
lmmodel <- lm(PeakCurrent~ Ni, dfnp)
s <- summary(lmmodel)
```

The fitted function looks like this:

$$Y = \beta_0 + \beta_1 x$$

~ Y = 0.018749+ 0.0042152x

The summary of the fitted model is as follows:

```r
s
```

```
##
## Call:
## lm(formula = PeakCurrent ~ Ni, data = dfnp)
##
## Residuals:
##         Min         1Q      Median         3Q         Max
## -0.0133264 -0.0042777 -0.0000231   0.0080557   0.0098107
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.875e-02  6.129e-03    3.059   0.0183 *
## Ni          4.215e-03  5.771e-05   73.040 2.37e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008376 on 7 degrees of freedom
## Multiple R-squared:  0.9987, Adjusted R-squared:  0.9985
## F-statistic:  5335 on 1 and 7 DF,  p-value: 2.372e-11
```

```r
alpha <- 0.05
sd <- as.numeric(s$coefficients[,2][2])
sxx <- sum((dfnp$Ni^2))-1/9*(sum(dfnp$Ni)^2)
c11 <- 1/sxx
slope <- as.numeric(s$coefficients[,1][2])
n= nrow(dfnp)

hypotest <- function(Ho, H1)
{
  teststat <- (slope-Ho)/(sd*c11)
  if(H1==">"){
    rejection <- teststat> qt(p=alpha, df=n-2, lower.tail=FALSE)
```

```
  }else{
    if(H1 == "<"){
      rejection <- teststat> qt(p=alpha, df=n-2, lower.tail=TRUE)
    }else{
      if(H1 == "!="){
        rejection <- abs(teststat)> qt(p=alpha/2, df=n-2, lower.tail=FALSE)
      }
    }
  }
  if(!rejection){
    retval <- "Null Hypothesis is true"
    return(retval)
  }else{
    retval <- "Null Hypothesis is rejected"
    return(retval)
  }
}

Nullhypo <- 0
Althypo <- ">"
print(paste0("Null Hypothesis is Ho : slope = ", Nullhypo))
```

```
## [1] "Null Hypothesis is Ho : slope = 0"
```

```
print(paste0("Alternative Hypothesis is H1: slope ", Althypo, Nullhypo))
```

```
## [1] "Alternative Hypothesis is H1: slope >0"
```

```
print(hypotest(Nullhypo, Althypo))
```

```
## [1] "Null Hypothesis is rejected"
```

11.69

The manufacturer of Lexus automobiles has steadily increased sales since the 1989 launch of that brand in the United States. However, the rate of increase changed in 1996 when Lexus introduced a line of trucks. The sales of Lexus vehicles from 1996 to 2003 are shown in the accompanying table.

```
codedyear <- c(-7, -5, -3, -1, 1, 3, 5, 7)
sales <- c(18.5, 22.6,27.2,31.2, 33.0, 44.9, 49.4, 35.0)
codedyearsq <- codedyear^2

dflexus <- data.frame(codedyear, sales, codedyearsq)
```

codedyear	sales	codedyearsq
-7	18.5	49
-5	22.6	25
-3	27.2	9
-1	31.2	1
1	33.0	1
3	44.9	9
5	49.4	25
7	35.0	49

codedyear	sales	codedyearsq

a Letting Y denote sales and x denote the coded year (-7 for 1996, -5 for 1997, through 7 for 2003), fit the model

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

.

```r
lmmodellex1 <- lm(sales~codedyear, dflexus)
slexus1 <- summary(lmmodellex1)
```

The fitted function looks like this:

$$Y = \beta_0 + \beta_1 x$$

~ $Y = 32.725 + 1.8119048x$

The summary of the fitted model is as follows:

```r
slexus1
```

```
##
## Call:
## lm(formula = sales ~ codedyear, data = dflexus)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -10.4083  -1.5381  -0.5774  1.9000  7.6155
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  32.7250     2.1297  15.366  4.8e-06 ***
## codedyear     1.8119     0.4647   3.899    0.008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.024 on 6 degrees of freedom
## Multiple R-squared:  0.717,  Adjusted R-squared:  0.6698
## F-statistic:  15.2 on 1 and 6 DF,  p-value: 0.007996
```

b For the same data, fit the model

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

.

```r
lmmodellex2 <- lm(sales~codedyear+codedyearsq, dflexus)
slexus2 <- summary(lmmodellex2)
```

The fitted function looks like this:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2$$

~ $Y = 35.5625 + 1.8119048x + -0.135119x^2$

The summary of the fitted model is as follows:

```r
slexus2
```

```
##
## Call:
## lm(formula = sales ~ codedyear + codedyearsq, data = dflexus)
##
## Residuals:
##      1      2      3      4      5      6      7      8
##  2.242 -0.525 -1.711 -2.415 -4.239  5.118  8.156 -6.625
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.5625     3.1224  11.390 9.13e-05 ***
## codedyear    1.8119     0.4481   4.044  0.00988 **
## codedyearsq -0.1351     0.1120  -1.206  0.28167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.808 on 5 degrees of freedom
## Multiple R-squared:  0.7808, Adjusted R-squared:  0.6931
## F-statistic: 8.904 on 2 and 5 DF,  p-value: 0.0225
```