#Assignment-1

#STP 530: Applied Regression Analysis

#Part 1

#a

x = c(1, 0, 2, 0, 3, 1, 0, 1, 2, 0)

y = c(16, 9, 17, 12, 22, 13, 8, 15, 19,11)

z=lm(y~x) #estimated regression fucntion

abline(y~x)

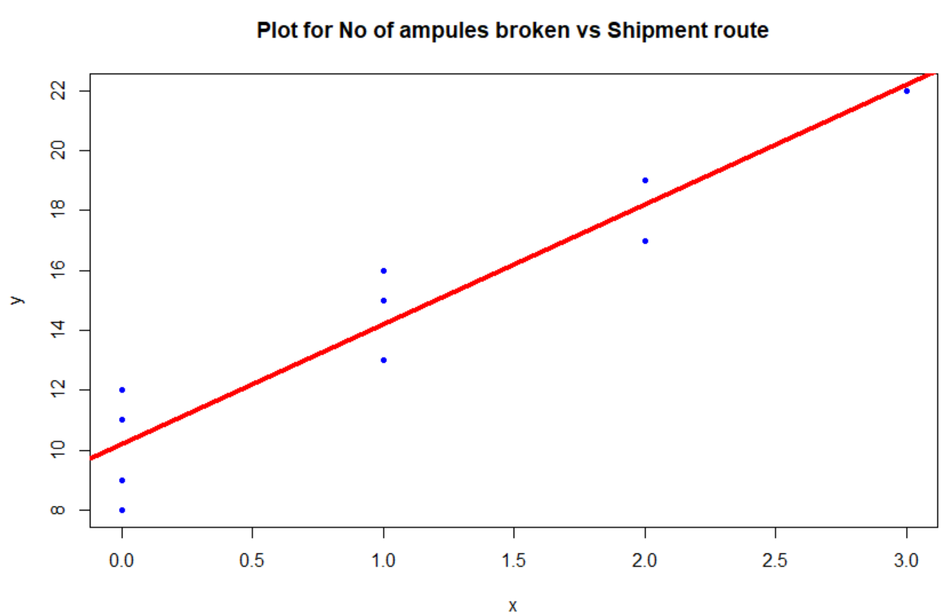
plot(y~x, col="blue", pch=20, main="Plot for No of ampules broken vs Shipment route")

line = lm(y~x)

summary(line)

abline(line, col="Red", lwd=4)

#the linear regression function does appear to give a good fit here because all the data are around the linear regression function like p-value of 2.75e-05 shows that it is statistically significant and Multiple R-squared is 0.9009 as well.



#b

coefficients((lm(y~x)))

```
(Intercept)           x
       10.2         4.0
```

a=data.frame(x=1)

predict(z,a)

#expected no of broken ampules are 14.2 when x=1 transfer is made

#c

coefficients((lm(y~x)))

b=data.frame(x=2)

predict(z,b)

predict(z,b) - predict(z,a)

#increase in no of broken ampules is 4 when there are 2 transfers as comapred to 1 trasnfer

#d

xbar = mean(x)

ybar = mean(y)

xbar

ybar

a=data.frame(x=1)

predict(z,a)

#thus on inputting values of xbar as 1, we get ans as ybar which is 14.2. Thus it passes through (xbar, ybar)

#Part 2

#a

GPAdata=read.table("D:\\ASU Stuff\\SEM-1\\STP 530\\120 students.txt")

ACT=GPAdata$V2

GPA=GPAdata$V1

z1=lm(GPA~ACT)

coef=coefficients(z1)

coef

```
(Intercept)         ACT
 2.11404929  0.03882713
```

#thus the model is 2.114+0.00388x

#b

z1=lm(GPA~ACT)

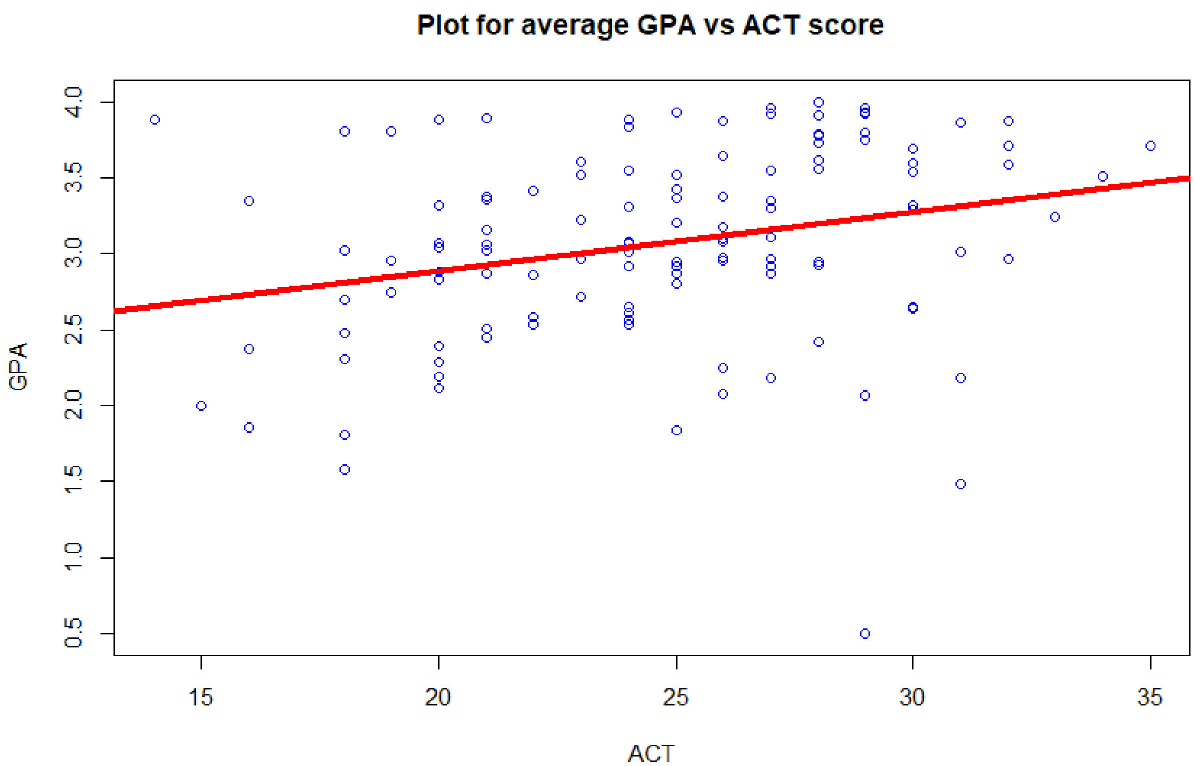
summary(z1)

line=(z1)

plot(GPA~ACT, main="Plot for average GPA vs ACT score", col="blue")

abline(z1, col="red", lwd=4)

#the line doesn't fit the data properly because there is no particular shape in data through which a line can pass



#c

meangpa=data.frame(ACT=30)

predict(z1,meangpa)

#point estimate of mean is 3.2788 when ACT is 30

#d

meangpa1=data.frame(ACT=31)

predict(z1,meangpa1)   #point estimate of mean is 3.3176 when ACT is 31

predict(z1,meangpa1) - predict(z1,meangpa)

#point estimate of the change of the mean response is 0.0388

#e

zresid=resid(z1)

zresid

sum(zresid)

#from the result, it is very much clear that the value doesn't sum up to zero

#f

MSE = sum(zresid^2)/(length(zresid)-2) #value of sigma square

sigma = sqrt(MSE) #value of sigma

sigma

#value of sigma is 0.6231

#Part 3

musclemass=read.table("D:\\ASU Stuff\\SEM-1\\STP 530\\60 students.txt")

age=musclemass$V2

measure=musclemass$V1

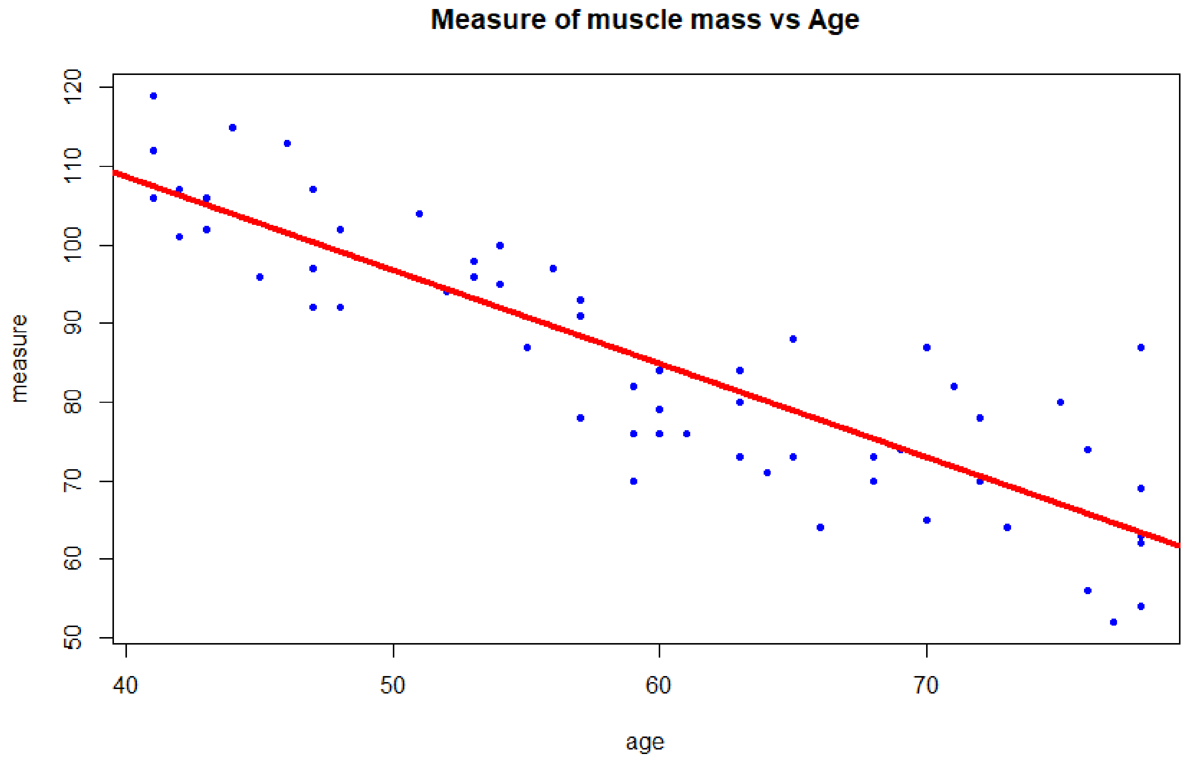
#a

z2=lm(measure~age) #function is 156.3466-1.1900*age

plot(measure~age, col="blue", pch=20, main="Measure of muscle mass vs Age")

abline(z2, col="red", lwd=4)

summary(z2)



#Yes, it does appear that the model does seem to give a good fit here as the R-squared is 0.7501 and the plot clearly shows that the muscle mass decreases with age

#b

#1

z2$coefficients[2]

#mean muscle mass difference for women differing in age by one year is -1.1899

#2

z2$coefficients[1] + z2$coefficients[2]*60

#point estimate for women aged 60 years is 84.9468

#3

Residuals=resid(z2)

#thus the residual at eighth value is 4.4432

#4

MSE = sum(Residuals^2)/(length(Residuals)-2)

#point estimate sigma^2 is 66.8008

#Part 4

#a

manuscript=c(7,12,4,14,25,30)

errors=c(128,213,75,250,446,540)

sumx=sum(manuscript)

sumy=sum(errors)

sumxx=sum(manuscript*manuscript)

sumyy=sum(errors*errors)

sumxy=sum(manuscript*errors)

lm(errors~manuscript)

```
Call:
lm(formula = errors ~ manuscript)

Coefficients:
(Intercept)   manuscript
      1.597       17.852
```

#a

likelihood=function(errors, manuscript, beta1)

{

lf=c()

for (i in 1:length(errors))

{

lf[i]=(1/(sqrt(32*pi)))*exp((-1/32)*(errors[i]-beta1*manuscript[i])^2) #likelihood function for six observations

}

likefunction=prod(lf)

return(likefunction)

}

#b

likelihood(errors, manuscript, 17)   # 9.45133e-30

likelihood(errors, manuscript, 18)   # 2.649043e-07

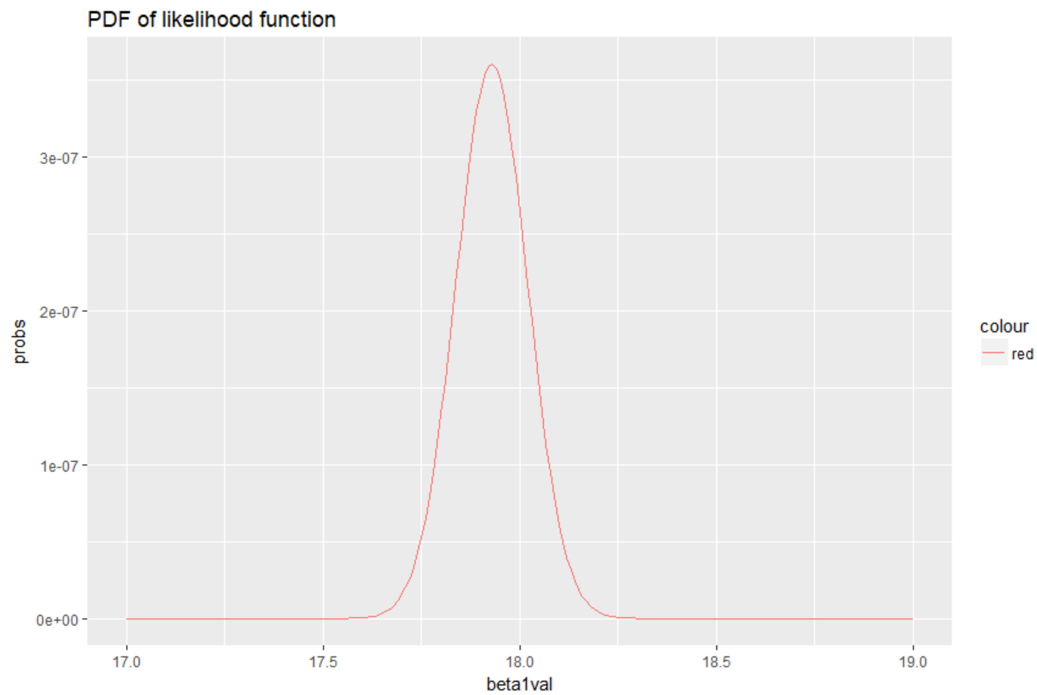
likelihood(errors, manuscript, 19)   # 3.047285e-37

#c

b1 = sumxy/sumxx   # 17.928

#d

```
beta1val=seq(17,19, by=0.01)
lf1=c()
for (i in 1:length(beta1val)) {
  lf1[i]=likelihood(errors, manuscript, beta1val[i])
}
lf1=data.frame(beta1val,lf1)
colnames(lf1)=c("beta1", "probs")
qplot(beta1val, probs, data=lf1, geom="line", col="red", main="PDF of likelihood function")
```



#thus the plot shows that the max. value of likelihood function is slightly less than 18 which is consistent with our max. likelihood estimate in part c.

#Part 5

#a

```
likelihood1=function(errors, manuscript, beta0, beta1)

{

  lf2=c()

  for ( i in 1:length(errors))

  {

    lf2[i]=(1/(sqrt(32*pi)))*exp((-1/32)*(errors[i]-beta0-(beta1*manuscript[i]))^2) #likelihood function for six observations

  }

  likefunction1=prod(lf2)

  return(likefunction1)

}
```

#b

coefficients(lm(errors~manuscript))

#beta0=1.5969   #beta1=17.8523

#c

```
b0val=seq(-10,10, by=0.05)

b1val=seq(17,19, by=0.05)

f=function(b0val,b1val) {

  (32*pi)^(-3)*exp(-19387.31-60.3125*b1val^2-0.1875*b0val^2+2162.24*b1val+103.25*b0val-5.75*b0val*b1val)

}

z=outer(b0val, b1val, f)

persp(b0val, b1val, z, col="red", theta=120, phi=30, zlab="Likelihood value")
```