Practice Midterm Solutions

Question 1

1-) The regression model we would like to study is:

$$Y_i = \beta_1 X_i + \varepsilon_i$$

a-)write the likelihood function

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(Y_i - \beta_1 X_i)^2}{2\sigma^2}\right)$$

a-)find the MLS estimations for b_1 and σ^2 log likelihood function

$$LogL = -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 X_i)^2$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{n}(Y_i X_i - \beta_1 X_i^2) = 0$$

$$\beta_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(Y_i - \beta_1 X_i)^2 = 0$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(Y_i - \beta_1 X_i)^2}{n}$$

Problem 2

n Xbar Ybar SXX SYY SXY SSE

###120 28 3 2340 49 92 46

a-) calculate b_0 and b_1 and 95% confidence intervals for b_0 and b_1

Solution:

```r
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.6.1
```

```r
n=120
Xbar=28
Ybar=3
SXX=2340
SYY=49
SXY=92
SSE=46

b1= SXY/SXX
b0 =  Ybar-b1*Xbar
MSE = SSE/(n-2)
Var.b1 = MSE/SXX
Var.bo= MSE*{1/n + Xbar^2 / SXX}
sb1 = sqrt(Var.b1)
sb0 = sqrt(Var.bo)
tc=qt(1-0.05/2,n-2)
cbind(b1-tc*sb1,b1+tc*sb1)
```

```
##            [,1]       [,2]
## [1,] 0.01375659 0.06487589
```

```r
cbind(b0-tc*sb0,b0+tc*sb0)
```

```
##         [,1]     [,2]
## [1,] 1.17463 2.623661
```

b) for Xh=3, predict Yh and calculate 95% prediction interval

```r
Yhat=b0+b1*3
Yhat
```

```
## [1] 2.017094
```

```r
s.yhat=sqrt(MSE*((1/n)+((3-Xbar)^2/SXX)))
cbind(Yhat-tc*s.yhat,Yhat+tc*s.yhat)
```

```
##          [,1]     [,2]
## [1,] 1.368211 2.665977
```

Problem 3

3. Based on the table above, write the ANOVA table and perform the F test, and perform a General Linear F test?

```r
SST = SYY
SSR= SST - SSE
cbind(SSR,SSE,SST)
```

```
##      SSR SSE SST
## [1,]   3  46  49
```

```r
Df.r=1
Df.e=n-2
cbind(1,n-2,n-1)
```

```
##      [,1] [,2] [,3]
## [1,]    1  118  119
```

```r
MSR=SSR
MSE=SSE/(n-2)
Ftest=MSR/MSE
cbind(MSR,MSE)
```

```
##      MSR       MSE
## [1,]   3 0.3898305
```

```r
Ftest
```

```
## [1] 7.695652
```

```r
1-pf(Ftest,1,n-2)
```

```
## [1] 0.006436837
```

Source DF SS MS F

Regression 1 3 3

Error 118 46 0.39 7.7

Total 119 49

Ho:B_1=0

Ha:B_1

F = (49-46)/(46/118) = 3/0.39=7.7, pavalue=0.006, reject null, model is significant.

Problem 4

a) Fit a regression model to predict house price. Obtain, the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

b) Conduct the Brown-Forsythe test to determine whether or not the error variance varies with the level of X.

c) Calculate the simultaneous interval for b_o,and b_1

d) Calculate the simultaneous confidence intervals for the predicted house prices for 1200,1400,1500 square feet

Solution:

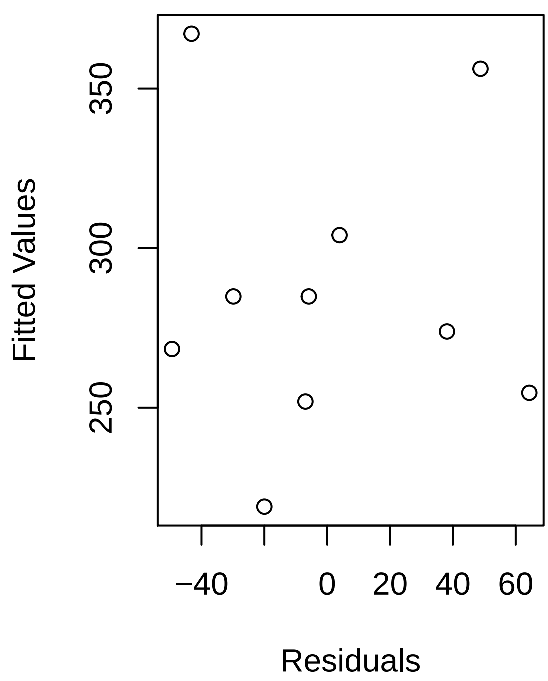
a)

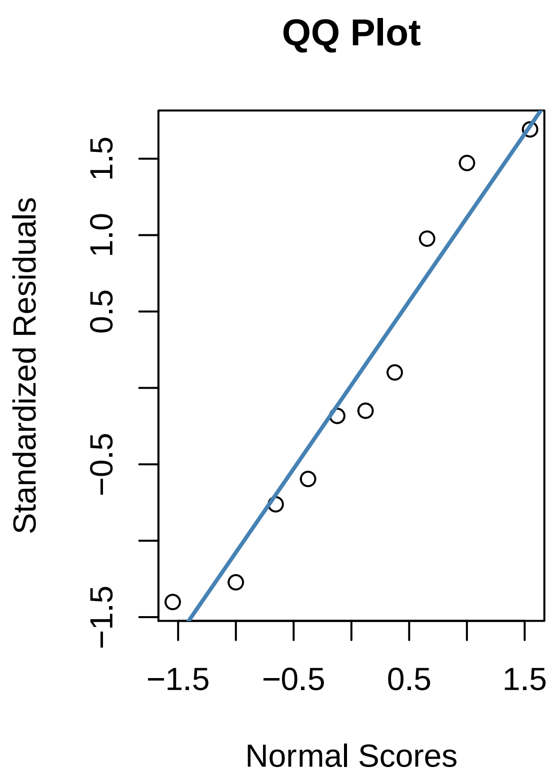
```
House.Price<-c(245,312,279,308,199,219,405,324,319,255)
Square.Feet<-c(1400,1600,1700,1875,1100,1550,2350,2450,1425,1700)
f4<-lm(House.Price~Square.Feet)
summary(f4)
```

```
##
## Call:
## lm(formula = House.Price ~ Square.Feet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.388 -27.388  -6.388  29.577  64.333
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 98.24833   58.03348   1.693   0.1289
## Square.Feet  0.10977    0.03297   3.329   0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.33 on 8 degrees of freedom
## Multiple R-squared:  0.5808, Adjusted R-squared:  0.5284
## F-statistic: 11.08 on 1 and 8 DF,  p-value: 0.01039
```

```
ei<-f4$residuals
yhat=f4$fitted.values
stdei<- rstandard(f4)
par(mfrow=c(1,2))
plot(ei,yhat,xlab="Residuals",ylab="Fitted Values")
qqnorm(stdei,ylab="Standardized Residuals",xlab="Normal Scores", main="QQ Plot")
qqline(stdei,col = "steelblue", lwd = 2)
```





The QQ plot roughly looks normal, error vs fitted values do not indicate unequal variance.

b) Conduct the Brown-Forsythe test to determine whether or not the error variance varies with the level of X.

```r
ei<-f4$residuals
M=median(Square.Feet)
DM<-data.frame(cbind(House.Price,Square.Feet,ei))
DM1<-DM[DM[,2]< M,]
DM2<-DM[DM[,2]>=M,]

M1<-median(DM1[,3])
M2<-median(DM2[,3])
N1<-length(DM1[,3])
N2<-length(DM2[,3])

d1<-abs(DM1[,3]-M1)
d2<-abs(DM2[,3]-M2)
s2<-sqrt((var(d1)*(N1-1)+var(d2)*(N2-1))/(N1+N2-2))
Den<- s2*sqrt(1/N1+1/N2)
Num<- mean(d1)-mean(d2)
T= Num/Den
T
```

```
## [1] 0.579081
```

t stat is less than 1, indicates that null shouls be accepted. There is No unequal variance.

c)

```r
confint(f4,level=1-0.05/2)
```

```
##                    1.25 %     98.75 %
## (Intercept) -61.43215607 257.9288153
## Square.Feet   0.01905154   0.2004839
```

###d) Calculate the simultaneous confidence intervals for the predicted house prices for 1200,1400,1500 square feet

```r
Xh<-c(1200,1400,1500)
predict.lm(f4,data.frame(Square.Feet  = c(Xh)),interval = "confidence", level = 1-0.05/3)
```

```
##        fit      lwr      upr
## 1 229.9696 165.3510 294.5882
## 2 251.9232 201.5793 302.2670
## 3 262.8999 218.0608 307.7391
```

Problem 5

a) Fit a linear regression function. Obtain, the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

###b) Use the Box-Cox procedure to find an appropriate power transformation by using λ = -.2, -.1,0, .1, .2. What transformation of Y is suggested? ###c) Use the transformation suggested by part b and obtain the estimated linear regression function for the transformed data. ###d) Express the estimated regression function in the original units. Predict Y for X=0.5 and calculate the 95% confidence interval

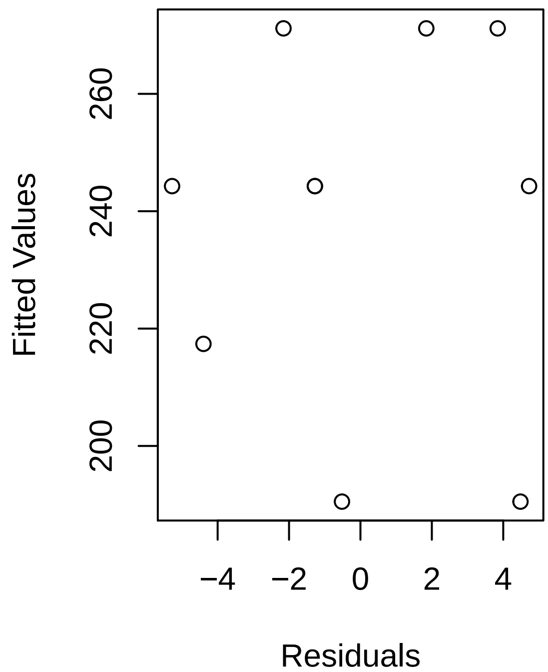
Solution:

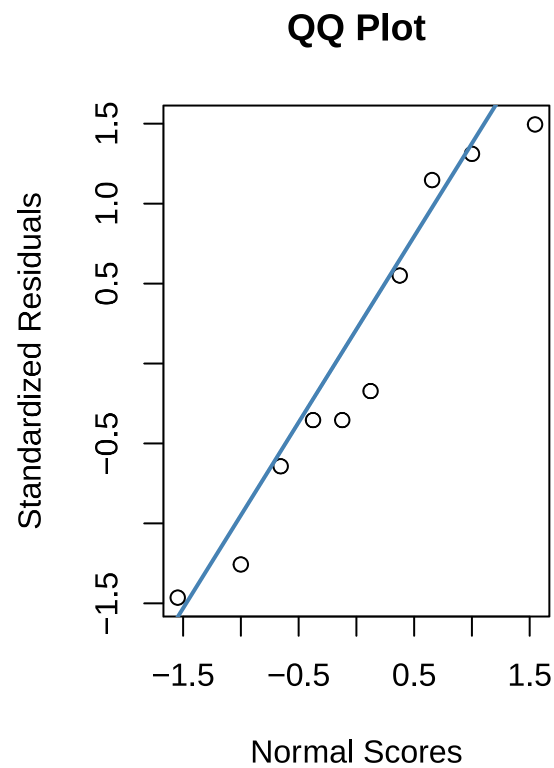
a)

```r
y<-c(243,195,275,190,213,249,239,243,269,273)
x<-c(5.5,5.3,5.6,5.3,5.4,5.5,5.5,5.5,5.6,5.6)
f5<-lm(y~x)
summary(f5)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2759 -1.9353 -0.8966  3.3448  4.7241
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1234.09      61.26  -20.14 3.85e-08 ***
## x             268.79      11.18   24.05 9.53e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.807 on 8 degrees of freedom
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.9847
## F-statistic: 578.3 on 1 and 8 DF,  p-value: 9.528e-09
```

```r
ei.5<-f5$residuals
yhat.5=f5$fitted.values
stdei.5<- rstandard(f5)
par(mfrow=c(1,2))
plot(ei.5,yhat.5,xlab="Residuals",ylab="Fitted Values")
qqnorm(stdei.5,ylab="Standardized Residuals",xlab="Normal Scores", main="QQ Plot")
qqline(stdei.5,col = "steelblue", lwd = 2)
```



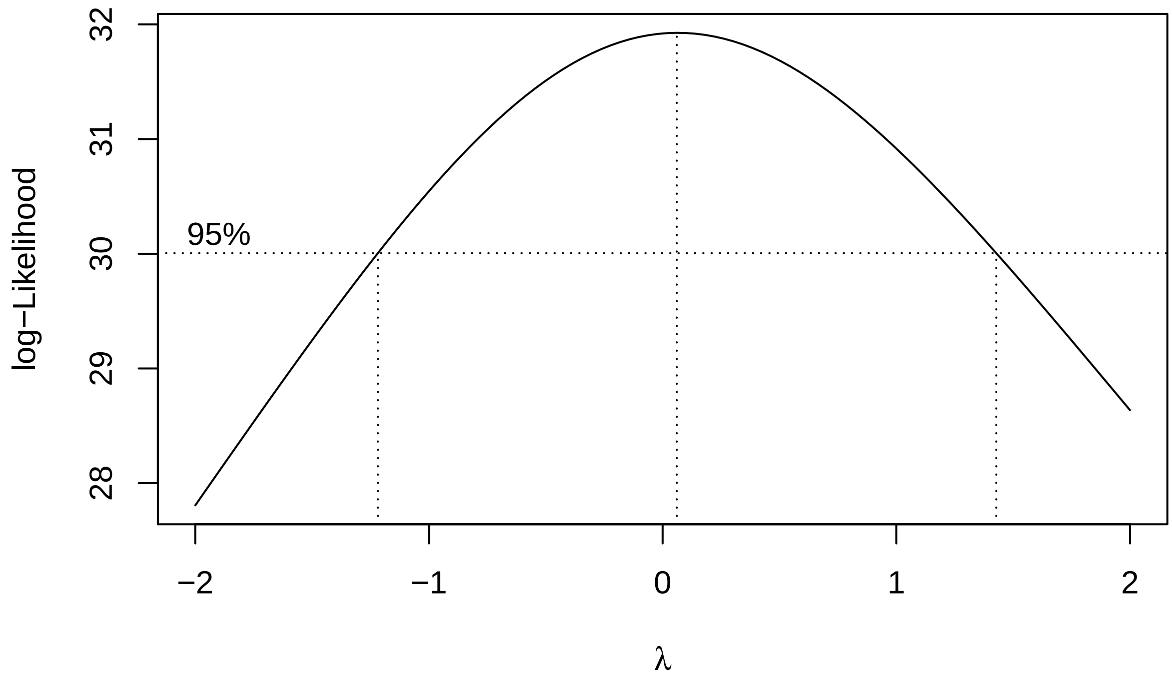


QQ plot shows S pattern, indicates problem with linearity. Log transformation might be needed. box-cox transformation should be performed.

b)

```
library(MASS)
boxcox(f5,lamda=seq(-2,2,by=0.1))
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'lamda' will be disregarded
```



boxcox plot shows that lamda is around zero, indicating log transformation

```
y1<-log(y)
f6<-lm(y1~x)
summary(f6)
```

```
##
## Call:
## lm(formula = y1 ~ x)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.015148 -0.012757  0.001203  0.006555  0.025841
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.93806    0.23377  -4.013  0.00388 **
## x            1.16903    0.04265  27.409 3.38e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01453 on 8 degrees of freedom
## Multiple R-squared:  0.9895, Adjusted R-squared:  0.9881
## F-statistic: 751.3 on 1 and 8 DF,  p-value: 3.385e-09
```

d)

log(Y) = -0.9380 + 1.16903 X Y= exp(-0.9380 + 1.16903 X)

```
exp(predict.lm(f6,data.frame(x=0.5),interval = "confidence", level = 1-0.05))
```

```
##         fit       lwr      upr
## 1 0.7021932 0.4302176 1.146107
```