HW4

Tuesday, June 3, 2025 12:21 PM

1. A company sells an imported copier and performs preventive maintenance and repair service on this copier. Data were collected from 45 recent calls on users to perform routine preventive maintenance service. For each call, x is the number of copiers serviced and Y is the total number of minutes spent by the service person. The data is available on Canvas as `Copier.txt`. Assume a simple linear regression model is appropriate.

   (a) Two service calls for preventive maintenance are scheduled in which the numbers of copiers to be serviced are 4 and 7, respectively. Provide a family of prediction intervals for the minutes to be spent on these calls with a 90% family-wise level, using the Bonferroni procedure.

   (b) Get the joint confidence intervals for the intercept and slope parameters at a 90% family-wise confidence level using the Bonferroni procedure.

```
> copierdata<-read.table("C:\\Users\\jacob\\OneDrive\\Documents\\Stats\\8050\\HW4\\Copier.txt")
> mins<-as.numeric(copierdata[-1,1])
> cops<-as.numeric(copierdata[-1,2])
> model<-lm(mins~cops)
> modelcoef<-as.numeric(coef(model))
> y4<-modelcoef[1]+modelcoef[2]*4
> y7<-modelcoef[1]+modelcoef[2]*7
> nmodel<-length(cops)
> xbar<-mean(cops)
> radius4<-qt(0.975,nmodel-2)*sqrt(sum(residuals(model)^2)*(1/nmodel+(4-xbar)^2/sum((cops-xbar)^2)))
> radius7<-qt(0.975,nmodel-2)*sqrt(sum(residuals(model)^2)*(1/nmodel+(7-xbar)^2/sum((cops-xbar)^2)))
> cat("Family of Intervals: y4 (",y4-radius4,",",y4+radius4,") and y7 (",y7-radius7,",",y7+radius7,")")
Family of Intervals: y4 ( 40.60946 , 78.51221 ) and y7 ( 83.35027 , 125.9829 )
```

```
> radiusb0<-qt(0.975,nmodel-2)*sqrt(sum(residuals(model)^2)*(1/nmodel+xbar^2/sum((cops-xbar)^2)))
> radiusb1<-qt(0.975,nmodel-2)*sqrt(sum(residuals(model)^2/sum((cops-xbar)^2)))
> cat("Family of Intervals: b0 (",modelcoef[1]-radiusb0,",",modelcoef[1]+radiusb0,") and b1 (",modelcoef[2]-r
adiusb1,",",modelcoef[2]+radiusb1,")")
Family of Intervals: b0 ( -37.66041 , 36.5001 ) and b1 ( 8.646741 , 21.42376 )
```

2. The data below show, for a consumer finance company operating in six cities, the number of competing loan companies operating in the city (x) and the number per thousand of the company's loans made in that city that are currently delinquent (Y). Assume that the simple linear regression model is applicable.

i	x_i	Y_i
1	4	16
2	1	5
3	2	10
4	3	15
5	3	13
6	4	22

(a) Using matrix methods, find:

i. $\boldsymbol{Y}^\top \boldsymbol{Y}$

ii. $\boldsymbol{X}^\top \boldsymbol{X}$

iii. $\boldsymbol{X}^\top \boldsymbol{Y}$

iv. $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$

(b) Using matrix methods, obtain the following:

i. Vector of estimated regression coefficients $\hat{\boldsymbol{\beta}}$

ii. Vector of residuals

iii. SSR

iv. SSE

v. Estimated variance-covariance matrix of $\hat{\boldsymbol{\beta}}$

vi. Point estimate of $E(Y_h)$ when $x_h = 4$

vii. $s^2(pred)$ when $x_h = 4$

(c) Find the hat matrix $\boldsymbol{H}$.

(d) Find $\boldsymbol{s}^2(\boldsymbol{e})$.

a)

$$Y = \begin{bmatrix} 16 \\ 5 \\ 10 \\ 15 \\ 13 \\ 22 \end{bmatrix} \qquad Y^T Y = 16^2 + 5^2 + 10^2 + 15^2 + 13^2 + 22^2$$
$$= 1259$$

$$X = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \qquad X^T X = \begin{bmatrix} 6 & 4+1+2+3+3+4 \\ 4+1+2+3+3+4 & 4^2+1^2+2^2+3^2+3^2+4^2 \end{bmatrix}$$
$$= \begin{bmatrix} 6 & 17 \\ 17 & 55 \end{bmatrix}$$

$$17 \quad 55 \Big]$$

$$X^T Y = \begin{bmatrix} 16+5+10+15+13+22 \\ 4\cdot 16+1\cdot 5+2\cdot 10+3\cdot 15+3\cdot 13+4\cdot 22 \end{bmatrix}$$

$$= \begin{bmatrix} 81 \\ 261 \end{bmatrix}$$

$$\det(X^T X) = 6\cdot 55 - 17^2 = 41$$

$$(X^T X)^{-1} = \frac{1}{41}\begin{bmatrix} 55 & -17 \\ -17 & 6 \end{bmatrix}$$

```
y<-c(16,5,10,15,13,22)
xi<-c(4,1,2,3,3,4)
x0<-c(1,1,1,1,1,1)
n<-length(x0)
X<-matrix(c(x0,xi),nrow=n,ncol=2,byrow="False")
J<-outer(x0,x0)
H<-X%*%solve(t(X)%*%X)%*%t(X)
xh<-c(1,4)
```

```
> b<-solve(t(X)%*%X)%*%t(X)%*%y
> b
          [,1]
[1,] 0.4390244
[2,] 4.6097561
> r<-y-H%*%y
> r
            [,1]
[1,] -2.87804878
[2,] -0.04878049
[3,]  0.34146341
[4,]  0.73170732
[5,] -1.26829268
[6,]  3.12195122
> SSR<-drop(t(y)%*%(H-1/n*J)%*%y)
> SSR
[1] 145.2073
> SSE<-drop(t(y)%*%(diag(n)-H)%*%y)
> SSE
[1] 20.29268
```

```
> MSE<-SSE/(n-2)
> Vb<-MSE*solve(t(X)%*%X)
> Vb
          [,1]       [,2]
[1,]  6.805473 -2.1035098
[2,] -2.103510  0.7424152
```

```
> yh<-drop(xh%*%b)
> yh
[1] 18.87805
> s2pred<-drop(MSE*(1+t(xh)%*%solve(t(X)%*%X)%*%xh))
> s2pred
[1] 6.929209
> H
            [,1]       [,2]       [,3]      [,4]      [,5]        [,6]
[1,]  0.36585366 -0.1463415 0.02439024 0.1951220 0.1951220  0.36585366
[2,] -0.14634146  0.6585366 0.39024390 0.1219512 0.1219512 -0.14634146
[3,]  0.02439024  0.3902439 0.26829268 0.1463415 0.1463415  0.02439024
[4,]  0.19512195  0.1219512 0.14634146 0.1707317 0.1707317  0.19512195
[5,]  0.19512195  0.1219512 0.14634146 0.1707317 0.1707317  0.19512195
[6,]  0.36585366 -0.1463415 0.02439024 0.1951220 0.1951220  0.36585366

> MSE*(diag(n)-H)
           [,1]       [,2]       [,3]       [,4]       [,5]       [,6]
[1,]  3.2171327  0.7424152 -0.1237359 -0.9898870 -0.9898870 -1.8560381
[2,]  0.7424152  1.7323022 -1.9797739 -0.6186794 -0.6186794  0.7424152
[3,] -0.1237359 -1.9797739  3.7120761 -0.7424152 -0.7424152 -0.1237359
[4,] -0.9898870 -0.6186794 -0.7424152  4.2070196 -0.8661511 -0.9898870
[5,] -0.9898870 -0.6186794 -0.7424152 -0.8661511  4.2070196 -0.9898870
[6,] -1.8560381  0.7424152 -0.1237359 -0.9898870 -0.9898870  3.2171327
```

3. A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data is named `SalesGrowth.txt`, where Y is the sales in thousands of units and x is the year (coded).

   (a) Prepare a scatter plot of the data. Does a linear relation appear adequate here?

   (b) Fit a normal simple linear regression model on the data and plot the residuals against x. What problem appears to be present here?

   (c) Use the Box-Cox procedure to find an appropriate power transformation of Y. Provide the Box-Cox plot of log-likelihood against λ with λ between 0 and 1. What transformation of Y is suggested?

   (d) Now use the transformation $Y' = \sqrt{Y}$ and obtain the estimated linear regression function for the transformed data.

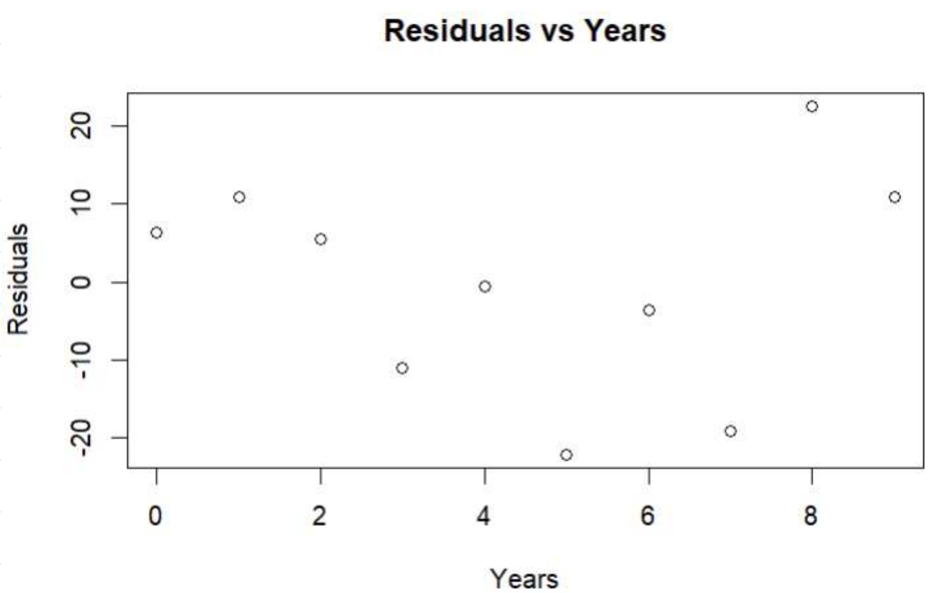
   (e) Plot the transformed data and add the estimated regression line on the plotted data. Does the regression line appear to be a good fit to the transformed data?

   (f) For the transformed data, plot the residuals against the fitted values. Also provide a normal QQ-plot. What do your plots show?

```
> salesdata<-read.table("C:\\Users\\jacob\\OneDrive\\Documents\\Stats\\8050\\HW4\\SalesGrowth.txt")
> sales<-as.numeric(salesdata[-1,1])
> years<-as.numeric(salesdata[-1,2])
> plot(years,sales,main="Sales vs Years",xlab="Years",ylab="Sales")
> nlin<-lm(sales~years)
> abline(nlin)
```

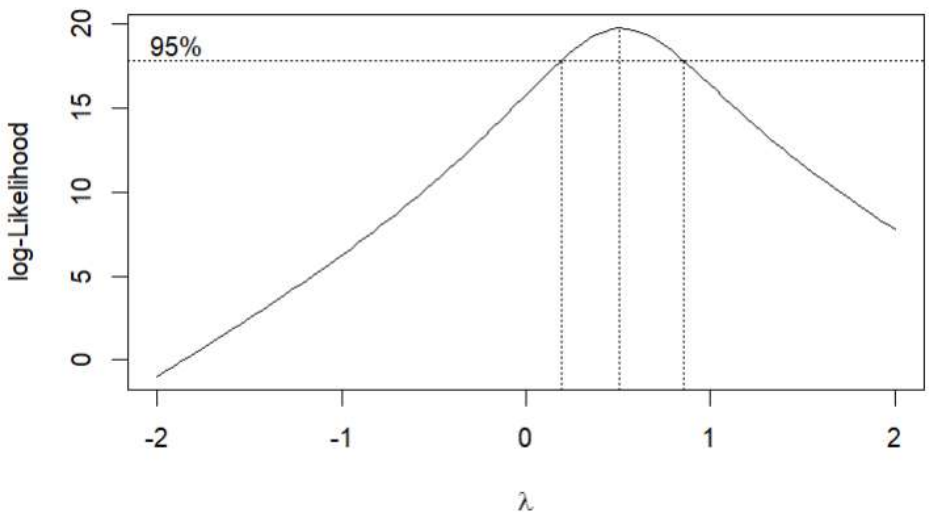
I feel like this is an okay scatter plot, but the professor said that it is not a good regression

```
> plot(years,residuals(nlin),main="Residuals vs Years",xlab="Years",ylab="Residuals")
```

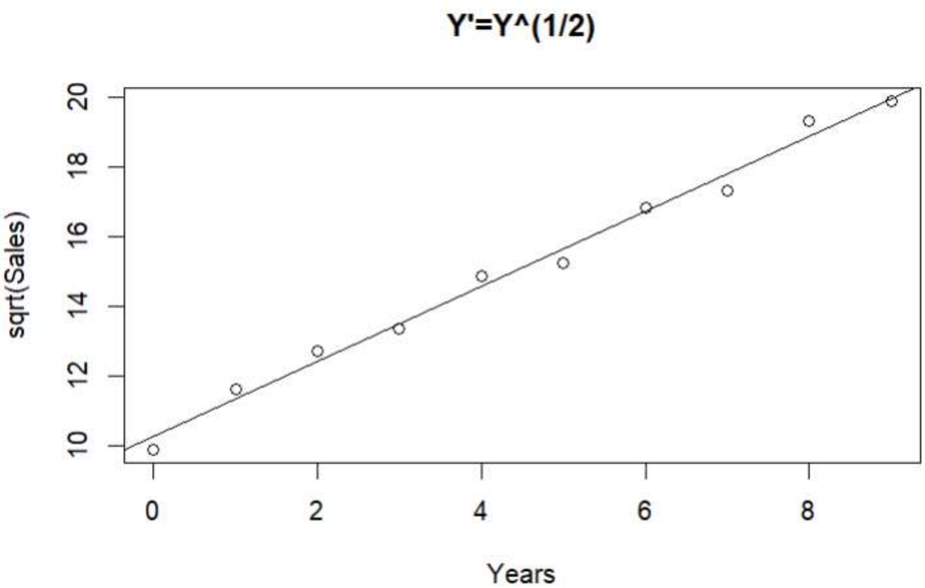


Again, I feel like this isn't terrible, but the professor said that this isn't good

```
> bc<-boxcox(sales~years)
> (lambda <- bc$x[which.max(bc$y)])
[1] 0.5050505
```

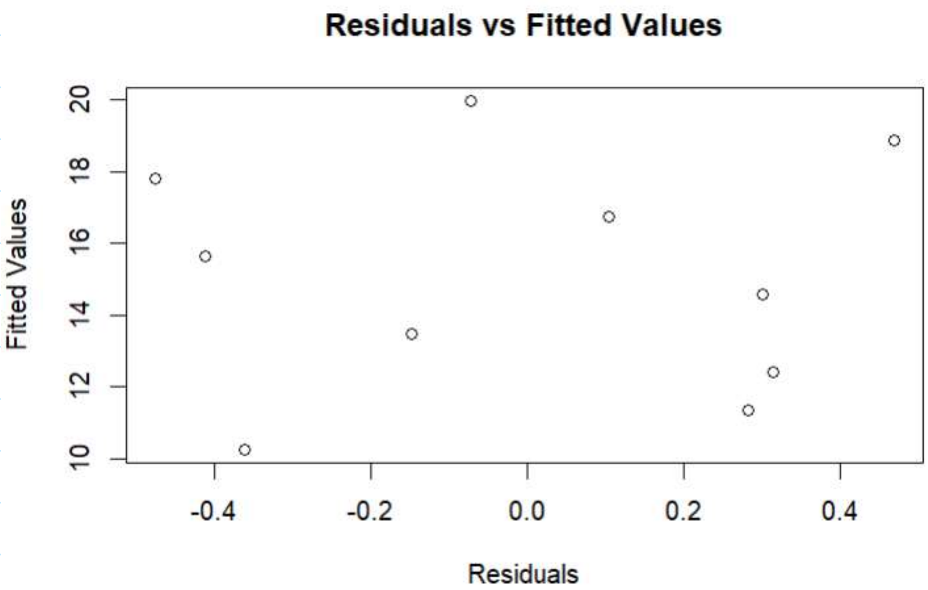



```
> sqrtsales<-sqrt(sales)
> nsqrt<-lm(sqrtsales~years)
> plot(years,sqrtsales,main="Y'=Y^(1/2)",xlab="Years",ylab="sqrt(Sales)")
> abline(nsqrt)
> coefsqrt<-as.numeric(coef(nsqrt))
> cat("Y'=",coefsqrt[1],"+",coefsqrt[2],"x")
Y'= 10.26093 + 1.076292 x
```



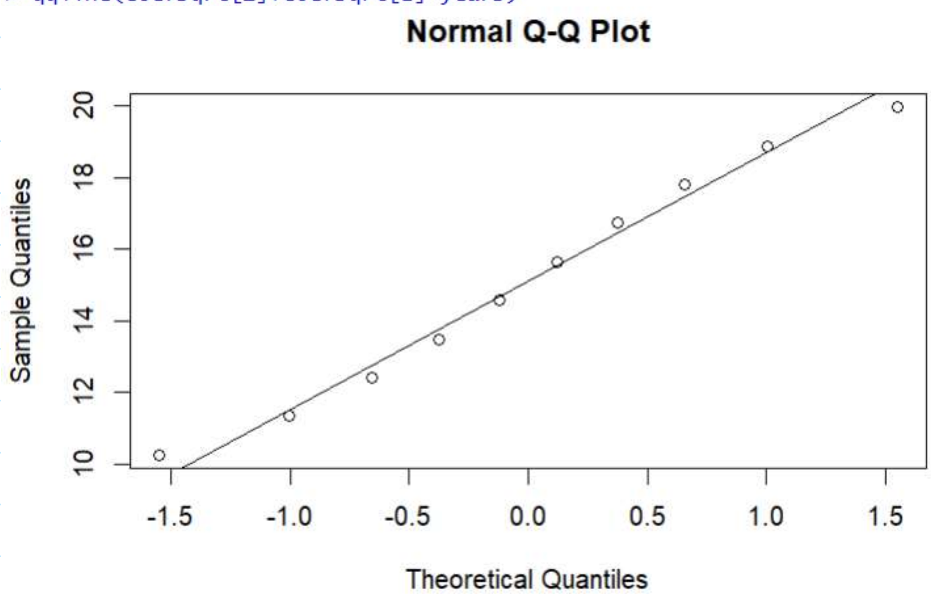
This is a better regression with smaller residuals visually speaking

```
> plot(residuals(nsqrt),coefsqrt[1]+coefsqrt[2]*years,main="Residuals vs Fitted Values",xlab="Residuals",ylab
="Fitted Values")
```



I think that this does not look any different than the other, but again the professor said the other one was bad

```
> qqnorm(coefsqrt[1]+coefsqrt[2]*years)
> qqline(coefsqrt[1]+coefsqrt[2]*years)
```



These residuals look fairly normal