80

STAT 8003, Homework 9

Group # 8

Members: Nooreen Dabbish, Yinghui Lu, Anastasia Vishnyakova

November 14, 2013

Problem 1. Consider a multiple linear model:

$$Y = X\beta + \epsilon,\ \epsilon \sim MVN(0, \sigma^2 I)$$

a). Write out the log-likelihood function of β and σ^2.

In the multiple linear model, as

$$Y = X\beta + \epsilon,\ \epsilon \sim MVN(0, \sigma^2 I)$$

then

$$Y \sim MVN(X\beta, \sigma^2 I)$$

Define the joint density of Y as $f(y; \beta, \sigma^2)$, and the $\log$-likelihood function as $l(\beta, \sigma^2; y)$, then

$$\begin{aligned}
l(\beta, \sigma^2; y) &= \log f(y; \beta, \sigma^2) \\
&= \log\{[(2\pi)^{-\frac{1}{2}n}|\sigma^2 I|^{-\frac{1}{2}}] \exp[-\frac{1}{2}(y - X\beta)^T(\sigma^2 I)^{-1}(y - X\beta)]\} \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta) \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y^T y - 2y^T X\beta + \beta^T X^T X\beta)
\end{aligned}$$

b). Get MLEs for β and σ^2.

Find first derivative with respect to β

$$\frac{\partial l}{\partial \beta} = \frac{1}{2\sigma^2}(2X^Ty - 2X^TX\beta)$$

Set the first derivative to zero, we have:

$$\begin{aligned} X^Ty &= X^TX\hat{\beta} \\ \hat{\beta} &= (X^TX)^{-1}X^Ty \end{aligned}$$

Find first derivative with respect to σ^2:

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(y - X\beta)^T(y - X\beta)}{2\sigma^4}$$

Set partial derivative to zero:

$$\begin{aligned} -\frac{n}{2\hat{\sigma}^2} + \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{2\hat{\sigma}^4} &= 0 \\ \frac{n}{2\hat{\sigma}^2} &= \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{2\hat{\sigma}^4} \\ \hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{n} \end{aligned}$$

Then check if the matrix below is negative definite:

$$P = \begin{pmatrix} \frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l}{\partial \sigma^4} \end{pmatrix}_{|\beta=\hat{\beta}, \sigma^2=\hat{\sigma}^2}$$

i)$\frac{\partial^2 l}{\partial \beta^2} = -\frac{1}{\sigma^2}(X^TX)$
As X has full column rank, X^TX is $p.d$, then $-\frac{1}{\sigma^2}(X^TX)$ is negative definite.

ii)$\frac{\partial^2 l}{\partial \beta \partial \sigma^2}|_{\beta=\hat{\beta},\sigma^2=\hat{\sigma}^2} = -\frac{1}{\sigma^4}(X^Ty - X^TX\beta) = 0_{p\times 1}$
iii)$\frac{\partial^2 l}{\partial \sigma^4}|_{\beta=\hat{\beta},\sigma^2=\hat{\sigma}^2} = \frac{n}{2\sigma^4} - \frac{(y-X\beta)^T(y-X\beta)}{\sigma^6} = \frac{-n}{2\hat{\sigma^4}} < 0$
Vi)$\frac{\partial^2 l}{\partial \sigma^2 \partial \beta}|_{\beta=\hat{\beta},\sigma^2=\hat{\sigma}^2} = \frac{1}{\sigma^4}(X^Ty - X^TX\beta)^T = 0_{1\times p}$

From the above we can see that, for matrix P, it is partitioned into four block matrices, of which all of the off diagonal blocks are 0, one of the diagonal block matrix is a negative definite matrix, and the other one diagonal element is a negative number. Hence the matrix P is negative definite. Combining the above results, we can generate the conclusion that

$$\hat{\beta}_{MLE} = (X^T X)^{-1} X^T y$$

$$\hat{\sigma}^2_{MLE} = \frac{(y - X\hat{\beta}_{MLE})^T (y - X\hat{\beta}_{MLE})}{n}$$

Problem 2.

Suppose we are interested in studying the relationship between the volume of a cherry tree and its height and diameters. We treat the volume as outcome, and height and diameter as covariates.

a). How to build up a linear regression model to study the relationship between volume and height diameters? Write out your model and interprete all the parameters in your model.

Denote the volume as V, height as H, diameter as D. If we assume the volume of the tree to be cylindrical, then

$$V = \frac{1}{4}\pi D^2 H$$

This gives us a hint to take $\log$ on both sides to construct a linear model. Then we write our model as

$$\log V = \beta_0 + \beta_1 \log H + \beta_2 \log D + \epsilon$$

To write it further into a matrix form, we then denote

$$Y = \begin{pmatrix} \log V_1 \\ \log V_2 \\ \vdots \\ \log V_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & \log H_1 & \log D_1 \\ 1 & \log H_2 & \log D_2 \\ \vdots & \vdots & \vdots \\ 1 & \log H_n & \log D_n \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \qquad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Then the linear model is

$$Y = X\beta + \epsilon$$

And consider that the model satisfy the below assumptions

$$E(\epsilon) = 0$$

and

$$Var(\epsilon) = \sigma^2 I$$

Interpretation of the parameters:
β_0 can be interpreted as the expected $\log V$ when both the $\log H$ and the $\log D$ are 0, that is when height=1 feet and diameter=1 inch.
β_1 can be interpreted as the average $\log V$ changes by β_1 for each unit change in $\log H$.
β_2 can be interpreted as the average $\log V$ changes by β_2 for each unit change in $\log D$.

b). The data give the volume (cubic feet), height (feet) and diameter (inches) (at 54 inches above ground) for a sample of 31 black cherry trees in the Allegheny National Forest, Pennsylvania. Please check the tab-separated dataset on blackboard called "cherry.txt". Plot the scatter plots of volume versus height, and volume versus diameter. Do you see any patterns?

The plots are as follows.
From the plots we can see that there is a positive correlation beween Height and Volume, but the linearity of the correlation is weak. And we can further see that the correlation is improved by taking the $\log$ transformation.
There is also a positive correlation beween Diameter and Volume, and the linearity seems good. After taking the $\log$, we can see that the linearity is still good.

```
R code for 2b:

> plot(cherry$Height,cherry$Volume,col='blue',xlab='Height',
ylab='Volume')
> plot(log(cherry$Height),log(cherry$Volume),col='red',xlab=
'log(Height)',ylab='log(Volume)')
> plot(cherry$Diam,cherry$Volume,col='blue',xlab='Diameter',
ylab='Volume')
> plot(log(cherry$Diam),log(cherry$Volume),col='red',xlab=
'log(Diam)',ylab='log(Volume)')
```

c). Fit the model using the data. What is your LSE estimator? And what's it's variance?

Define $Q(\beta) = ||Y - X\beta||_2^2$, and we will find the $\hat{\beta}$, which minimize $Q(\beta)$.

$$Q(\beta) = (Y - X\beta)^T (Y - X\beta)$$

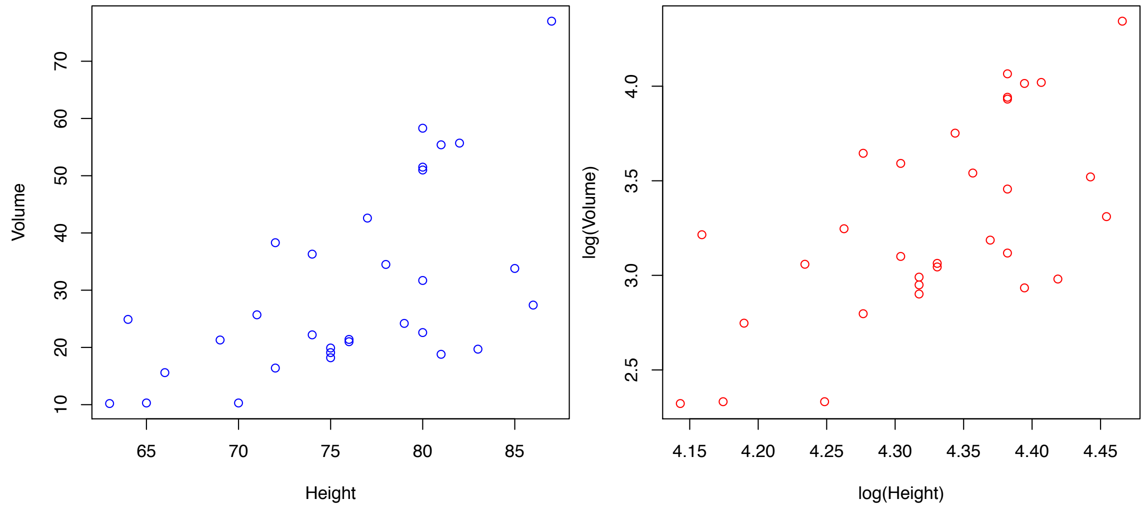

Figure 1: Scatter plot of Volume vs. Height and log(Height)

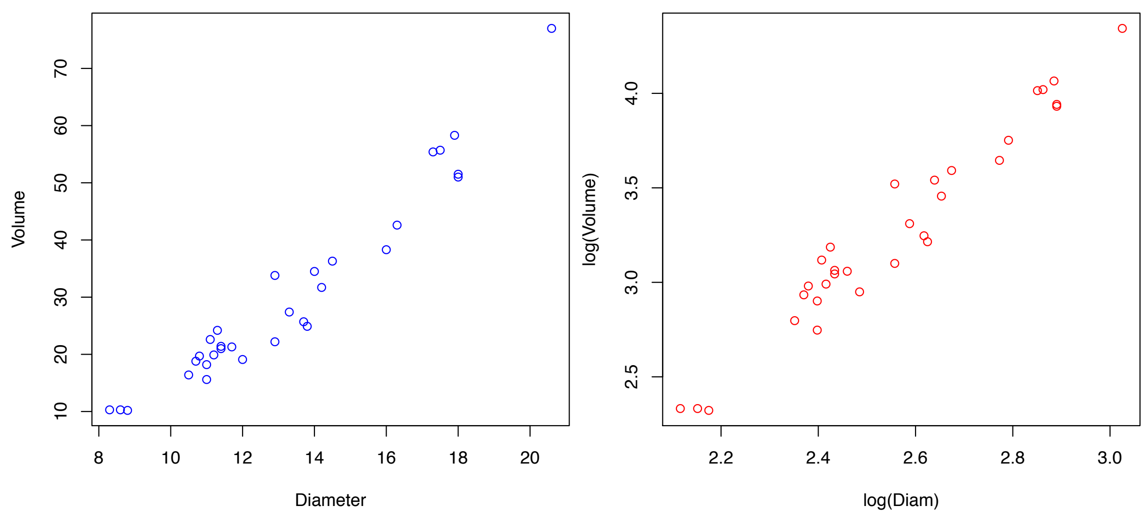

Figure 2: Scatter plot of Volume vs. Diameter and log(Diameter)

Take derivative of the above equation and set it to 0,

$$\frac{\partial Q(\beta)}{\partial \beta} = \frac{\partial(Y^TY - 2Y^TX\beta + \beta^TX^TX\beta)}{\partial \beta}$$
$$= -2X^TY + 2X^TX\beta$$

Set $\frac{\partial Q(\beta)}{\partial \beta} = 0$, we have

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

Check the second derivative,

$$\frac{\partial^2 Q(\beta)}{\partial \beta^2} = 2X^TX$$

As X is with full column rank, X^TX is $p.d$, then $\hat{\beta}$ we obtain minimizes $Q(\beta)$, hence $\hat{\beta}_{LSE} = (X^TX)^{-1}X^TY$.
Solving it in R, we have,

$$\hat{\beta}_{LSE} = \begin{pmatrix} -6.63 \\ 1.12 \\ 1.98 \end{pmatrix}$$

$$\begin{aligned} E(\hat{\beta}_{LSE}) &= E[(X^TX)^{-1}X^TY] \\ &= (X^TX)^{-1}X^TE(Y) \\ &= (X^TX)^{-1}X^TX\beta \\ &= \beta \end{aligned}$$

$$\begin{aligned} Var(\hat{\beta}_{LSE}) &= Var((X^TX)^{-1}X^TY) \\ &= (X^TX)^{-1}X^TVar(Y)[(X^TX)^{-1}X^T]^T \\ &= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1} \end{aligned}$$

As σ^2 is unknown, we use $\hat{\sigma}^2$ to estimate it,

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{RSS}{n-3} \\
&= \frac{\sum_{i=1}^{n}(y^i - \hat{y}_i)^2}{n-3} \\
&= \frac{(Y-\hat{Y})^T(Y-\hat{Y})}{n-3} \\
&= \frac{(Y-P_XY)^T(Y-P_XY)}{n-3} \\
&= \frac{Y^TY - Y^TP_XY}{n-3}
\end{aligned}
$$

Where $P_X = X(X^TX)^{-1}X^T$.
Hence,

$$
\begin{aligned}
\widehat{Var}(\hat{\beta}_{LSE}) &= \hat{\sigma}^2(X^TX)^{-1} \\
&= \frac{Y^TY - Y^TP_XY}{n-3}(X^TX)^{-1}
\end{aligned}
$$

Plugging in the data, and solving it in R, we have

$$
\hat{\sigma} = 0.0814
$$

$$
\widehat{Var}(\hat{\beta}_{LSE}) = \begin{pmatrix} 0.6397 & -0.1601 & 0.0208 \\ -0.1601 & 0.0418 & -0.0081 \\ 0.0208 & -0.0081 & 0.0056 \end{pmatrix}
$$

R code:

```
> n=31
> Y=log(cherry$Volume)
> X=cbind(rep(1,31),log(cherry$Height),log(cherry$Diam))
> betahat=solve(crossprod(X))%*%t(X)%*%Y
> betahat
          [,1]
[1,] -6.631617
[2,]  1.117123
[3,]  1.982650
> sigmahat=sqrt((t(Y)%*%Y-t(Y)%*%X%*%solve(crossprod(X))%*%t(X)%*%Y)/(n-3))
> sigmahat
```

```
          [,1]
[1,] 0.08138607
> Var=drop(sigmahat^2)*solve(crossprod(X))
> Var
            [,1]         [,2]         [,3]
[1,]  0.63966361 -0.160062124  0.020793539
[2,] -0.16006212  0.041794512 -0.008130512
[3,]  0.02079354 -0.008130512  0.005626592
```

d). Construct a 95% condence interval for the effect of height on volume? What would be the result?

From part c), we know that $\hat{\beta}_{LSE} \sim N(\beta, \sigma^2(X^TX)^{-1})$, then,

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2(X^TX)^{-1}_{11})$$

As σ^2 is unknown, we use $\hat{\sigma}$ to estimate, then we have

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{11}}} \sim T_{n-3}$$

Let the 97.5%-th quantile of t distribution with n-3 degree of freedom to be $t_{n-3}(2.5)$, then according to pivot method, we can construct a 95% confidence interval with

$$L = \hat{\beta}_1 - t_{n-3}(2.5)\hat{\sigma}\sqrt{(X^TX)^{-1}_{11}}$$

$$U = \hat{\beta}_1 + t_{n-3}(2.5)\hat{\sigma}\sqrt{(X^TX)^{-1}_{11}}$$

Solving it in R, we get

$$L = 0.6984 \qquad U = 1.5359$$

It means that we are 95% confident that the CI we constructed would cover the true value of β_1. From the result we can see that 0 is not including in the confidence interval, which suggest that height might has a significant effect on volume.

```
R code:
> L=betahat[2]-qt(0.975,n-3)*sqrt(Var[2,2])
> U=betahat[2]+qt(0.975,n-3)*sqrt(Var[2,2])
> L
[1] 0.698353
> U
```

```
[1] 1.535894
```

e). Test whether the effect of diameter is significant or not, controlling type I error at 5%. What's your p-value. What's your conclusion?

Consider the hypothesis testing

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

Under H_0, we have

$$\frac{\hat{\beta}_2 - 0}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{22}}} \sim T_{n-3}$$

Denote the t random variable with n-3 degree of freedom as T, then our p-value is

$$\begin{aligned} p\text{-value} &= P\left(T > \left|\frac{\hat{\beta}_2}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{22}}}\right|\right) \\ &= P\left(T < -\frac{\hat{\beta}_2}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{22}}}\right) + P\left(T > \frac{\hat{\beta}_2}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{22}}}\right) \end{aligned}$$

Solving it in R, we have

$$p\text{-value} = 2.42255e - 21 \approx 0 < 0.05$$

From this result we can say that the effect of diameter on volume is significant.It appears that volume is affected by tree's diameter - the bigger the diameter, the higher the volume of the tree.

```
> pvalue=2*pt(-betahat[3]/sqrt(Var[3,3]),n-3)
> pvalue
[1] 2.42255e-21
```

f). Now comes a new tree, with (height, diameter) = (76; 22). Can you predict the volume of the tree? How to construct a 95% CI of the predicted volume?

We can predict the volumn of the tree using the linear model we just constructed.

$$\begin{aligned}\hat{y_0} &= \hat{\beta}_0 + \hat{\beta}_1 \log 76 + \hat{\beta}_2 \log 22 \\ &= 4.33\end{aligned}$$

So we predict the volume of the new tree would be $\hat{V}_0 = \exp(4.33) = 76.31$ cubic feet.

Write $x_0 = (1 \quad \log 76 \quad \log 22)$, then the predicted interval would be

$$\left(x_0\hat{\beta} - t_{n-3}(2.5)\hat{\sigma}\sqrt{1 + x_0(X^TX)^{-1}x_0^T}, x_0\hat{\beta} + t_{n-3}(2.5)\hat{\sigma}\sqrt{1 + x_0(X^TX)^{-1}x_0^T} \right)$$

Plugging in the data, and solving it using R, we have the predicted interval of

$$L_1 = 4.15$$

$$U_1 = 4.52$$

So the 95% prediction interval for V is

$$L = \exp(4.15) = 63.24$$

$$U = \exp(4.52) = 92.08$$

```
R code for 2e:

> x_0=c(1,log(76),log(22))
> y_0hat=x_0%*%betahat
> y_0hat
[1] 4.334801
> V_0=exp(y_0hat).
> V_0
[1] 76.30979

> L_1=x_0%*%betahat-qt(0.975,n-3)*sigmahat*sqrt(1+x_0%*%solve(crossprod(X))%*%x_0)
> U_1=x_0%*%betahat+qt(0.975,n-3)*sigmahat*sqrt(1+x_0%*%solve(crossprod(X))%*%x_0)
> L_1
         [,1]
[1,] 4.146948

> U_1
         [,1]
[1,] 4.522655
```

```
> L=exp(L_1)
> U=exp(U_1)
> L
         [,1]
[1,] 63.24067
> U
         [,1]
[1,] 92.07972
```