Homework 02
Joseph Blubaugh
jblubau1@tamu.edu
STAT 608-720

1. For this to be true we must assume that the variance of $var(x_i) = 0$. Variance of constants are 0 as well so the variance of $var(Y_i) = var(e_i)$

$$
\begin{aligned}
Y_i =& B_0 + B_1 X_i + e_i \\
var(Y_i) =& var(B_0) + var(B_1 var(X_i)) + var(e_i) \\
var(Y_i) =& var(e_i)
\end{aligned}
$$

2. The sum of squares is calculated by squaring and summing the difference between the predicted value and actual value of y for each observation. The coefficients are set so that the total sum of squared errors are minimized.

3. a) Since there are not predictor variables there will only be a constant coefficient plus the error term. $B_0 = \bar{y}$ because the average value of why will minimize the RSS.

$$
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \ldots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix} + \begin{bmatrix} B_0 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ \ldots \\ e_n \end{bmatrix}
$$

$$Y_i = B_0 + e_i$$

   b)

$$For, \hat{B} = (X'X)^{-1} X'Y$$

$$
\begin{aligned}
E(\hat{B}|X) &= (X'X)^{-1} X'Y|X \\
&= (X'X)^{-1} X'E(Y|X) \\
&= (X'X)^{-1} X'XB \\
&= B
\end{aligned}
$$

$$
\begin{aligned}
Y_i &= B_0 n \\
\bar{Y} &= B_0
\end{aligned}
$$

4. a) i.

$$
\begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \ldots & \ldots \\ 1 & x_n \end{bmatrix} + \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{bmatrix} \rightarrow y_i = B_0 + B_1 x + e_i
$$

ii.

$$
\begin{aligned}
Y_i &= X_iB + e_i \\
e_i &= Y - X_iB
\end{aligned}
$$

$$
\sum e_i^2 = e'e = \begin{bmatrix} e_1 & e_2 & ... & ei \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ ... \\ e_i \end{bmatrix} = (Y - X_iB)'(Y - X_iB)
$$

$$
\begin{aligned}
\frac{d}{dB}(Y - X_iB)'(Y - X_iB) &= -2X'(Y - XB) \\
-2X'(Y - XB) &= 0 \\
X'Y &= (X'X)B \\
B &= (X'X)^{-1}X'Y
\end{aligned}
$$

b)

$$
\begin{aligned}
E(\hat{B}|X) &= (X'X)^{-1}X'Y|X \\
&= (X'X)^{-1}X'E(Y|X) \\
&= (X'X)^{-1}X'XB \\
&= B
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{B}|X) &= Var((X'X)^{-1}X'Y) \\
&= (X'X)^{-1}X'Var(Y)X(X'X)^{-1} \\
&= (X'X)^{-1}X'Var(XB + e)X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sum X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
\frac{\sigma^2}{\sum_{i=1}^n x_i^2} &= \sigma^2(X'X)^{-1}
\end{aligned}
$$

c) For $\hat{B}|X \to N(B, \frac{\sigma^2}{\sum_{i=1}^n x_i^2})$
Since we know that the errors are normally distributed, the Ys and the sum of the Ys must be normally distributed so $\hat{B}$ is a linear combination of the Ys. The distribution therefore must by normal.

5.

$$X'X = \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix} X'Y = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} (X'X(^{-1}) = \frac{1}{SXX} \begin{bmatrix} \frac{1}{n} & -\bar{x} \\ -x & 1 \end{bmatrix}$$

$$\begin{aligned}
\hat{B} &= (X'X)^{-1}X'Y \\
&= \frac{1}{SXX} \begin{bmatrix} \frac{1}{n} & -\bar{x} \\ -x & 1 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i \end{bmatrix} \\
&= \begin{bmatrix} \frac{\bar{y}SXX - \bar{x}SXY}{SXX} \\ \\ \frac{SXY}{SXX} \end{bmatrix} \\
&= \begin{bmatrix} \bar{y} - \hat{B}\bar{x} \\ B \end{bmatrix}
\end{aligned}$$

6.

$$\begin{aligned}
Var(a'\hat{B}|X) &= Var(a'(X'X)^{-1}X'Y) \\
&= (X'X)^{-1}X'aVar(Y)Xa'(X'X)^{-1} \\
&= (X'X)^{-1}X'aVar(XB+e)Xa'(X'X)^{-1} \\
&= (X'X)^{-1}X'\sum Xa'(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IXa'(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'Xa'(X'X)^{-1} \\
\frac{a'\sigma^2}{\sum_{i=1}^{n} x_i^2} &= a'\sigma^2(X'X)^{-1}
\end{aligned}$$

7. Confidence Intervals are an indication of the mean of the prediction. A model that explains variance well could have a small confidence interval meaning there is high certainty that the mean of all values at that particular x is within that range. The prediction interval is the interval in which would expect 95 percent of predicted values to be for a particular x.

8. The line plotted over the data points is not least squares because it is not intersecting the averages of both x and y. Simple least squares regression will fit models through the averages of two vectors because that is the point where the sum of squared errors are minimized.