STAT 3008 Applied Regression Anylysis

Midterm Solution

2012-2013 1st term, Oct. 29, 2012

Q1-11.(By Yang Ping @ LSB G30)

1. C) is incorrect, since
$$\frac{(\bar{X}-\mu)\sqrt{n}}{S} \sim t(n-1). \qquad \text{(\textit{3 marks})}$$

2. Mean function $= 0$; (*1 mark*) Variance function = constant; (*1 mark*) No separated points. (*1 mark*)

3. Since there are only two data points, the fitted line must past through these two points.
$$\frac{y-4}{x+2} = \frac{3-4}{-1+2}. \qquad \text{(\textit{2 marks})}$$
$$\Rightarrow y = 2 - x. \qquad \text{(\textit{1 mark})}$$
Hence, $\hat{\beta}_0 = 2$ (*1 mark*), $\hat{\beta}_1 = -1$ (*1 mark*).

4. a), b), d), and e) are correct. (*4 correct answers for 3 marks, 3 for 2 and so on*)
Error term is a random variable and it can not be consistently estimated.

5. $Var(\hat{e}) = Var[(I-H)Y] = (I-H)Var(Y)(I-H) = \sigma^2(I-H) \neq \sigma^2 I.$
(*2 marks for each of the first two equality signs; 1 mark for each of the others*)

6.
$$\hat{e}_5 = 0 - \sum_{i=1} 4\hat{e}_i = -0.1. \qquad \text{(\textit{2 marks})}$$
$$\therefore RSS = 3 \times 0.1^2 + 0.2^2 + 0.3^2 = 0.16. \qquad \text{(\textit{2 marks})}$$
$$R^2 = 1 - \frac{RSS}{SYY} = 1 - \frac{0.16}{3.3} = 0.952. \qquad \text{(\textit{2 marks})}$$

7. c). (*1 mark*) Compare their absolute value.

8. b). (*6 marks*)
$$\hat{\beta}_{1_{(new)}} = \frac{c\sum(x_i-\bar{x})(y_i-\bar{y})}{c^2\sum(x_i-\bar{x})^2} = \frac{1}{c}\hat{\beta}_1.$$

$$\hat{\beta}_{0_{(new)}} = \bar{y} - (\frac{1}{c}\hat{\beta}_1)(c\bar{x}) = \bar{y} - \hat{\beta}_1\bar{x} = \hat{\beta}_0.$$

$$\hat{e}_{i_{(new)}} = y_i - \hat{\beta}_0 - \hat{\beta}_{1_{(new)}} x_{i_{(new)}} = y_i - \hat{\beta}_0 - (\frac{1}{c}\hat{\beta}_1)(cx_i) = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = \hat{e}_i,$$

$$\therefore \hat{\sigma}^2_{(new)} = \frac{\sum \hat{e}_i^2}{n-2} = \hat{\sigma}^2.$$

$$\because \hat{e}_{i_{(new)}} = \hat{e}_i, RSS_{(new)} = RSS.$$

$$\because SYY = \sum (y_i - \bar{y})^2 \text{ is unaffected,}$$

$$\therefore R^2_{(new)} = 1 - \frac{RSS}{SYY} = R^2.$$

$$t_{(new)} = \frac{\hat{\beta}_0}{se(\hat{\beta}_0)_{(new)}}, \text{ where } se(\hat{\beta}_0)_{(new)} = \hat{\sigma}(\frac{1}{n} + \frac{\bar{x}^2_{(new)}}{SXX_{(new)}})^{1/2}.$$

$$\because \frac{\bar{x}^2_{(new)}}{SXX_{(new)}} = \frac{c^2\bar{x}^2}{c^2\sum(x_i - \bar{x})} = \frac{\bar{x}^2}{\sum(x_i - \bar{x})} = \frac{\bar{x}^2}{SXX},$$

$$\therefore t_{(new)} = t.$$

9. $tr(I_n - X(X'X)^{-1}X') = tr(I_n) - tr(X(X'X)^{-1}X') = n - tr(X(X'X)^{-1}X')$

$$tr(X(X'X)^{-1}X') = tr[(X'X)^{-1}(X'X)],$$ (*1 mark*)

$$tr[(X'X)^{-1}(X'X)] = tr(I_r) = r,$$ (*1 mark*)

$$\therefore tr(I_n - X(X'X)^{-1}X') = n - r.$$ (*1 mark*)

10.

$$\frac{df(\beta)}{d\beta} = X^TY - 2X^TX\beta,$$ (*1 mark*)

$$\Rightarrow \hat{\beta} = \frac{1}{2}(X^TX)^{-1}X^TY,$$ (*1 mark*)

$$\therefore \text{ The minimum value of } f(\beta) = \frac{1}{2}Y^TX(X^TX)^{-1}X^TY - \frac{1}{4}Y^TX(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^TY = \frac{1}{4}Y^TX(X^TX)^{-1}X^TY.$$ (*1 mark*)

11. C.I. for $E(Y|X)$ is for a point, while C.B. for $E(Y|X)$ is C.I. for the whole line.

(*2 marks*)

C.I. for $E(Y|X)$ uses t_{n-p-1}, while C.B. for $E(Y|X)$ uses $\sqrt{(p+1)F_{(p+1,n-p-1)}}$.

(*1 mark*)

Q12.(By Pan Deng@LSB126)

$\bar{Y} = 12/4 = 3$, $SYY = 25 + 16 + 9 + 36 = 86$.

ANOVA Table

Source	Sum of Squares	d.f.	Mean Square	F-statistics
Regression	74(half mark)	2(1 mark)	37(1 mark)	37/12(1 mark)
Residuals	12	1(1 mark)	12(1 mark)	
Total	86(half mark)			

Q13.(By Pan Deng@LSB126)

Let $H = X(X'X)^{-1}X'$, then $\hat{Y} = HY$ and $\sum_{i=1}^{n} \hat{y}_i^2 = \hat{Y}'\hat{Y}$.

$$\begin{aligned}
E(\sum_{i=1}^{n} \hat{y}_i^2) &= E[(HY)'(HY)] && \text{(1 mark)} \\
&= E(Y'HY) && \text{(1 mark)} \\
&= E[tr(HYY')] && \text{(1 mark)} \\
&= tr[H \cdot E(YY')] && \text{(1 mark)} \\
&= tr[H(X\beta\beta'X' + \sigma^2 I)] \\
&= tr(\beta'X'X\beta) + tr(\sigma^2 H) \\
&= \beta'X'X\beta + p\sigma^2 && \text{(1 mark)}
\end{aligned}$$

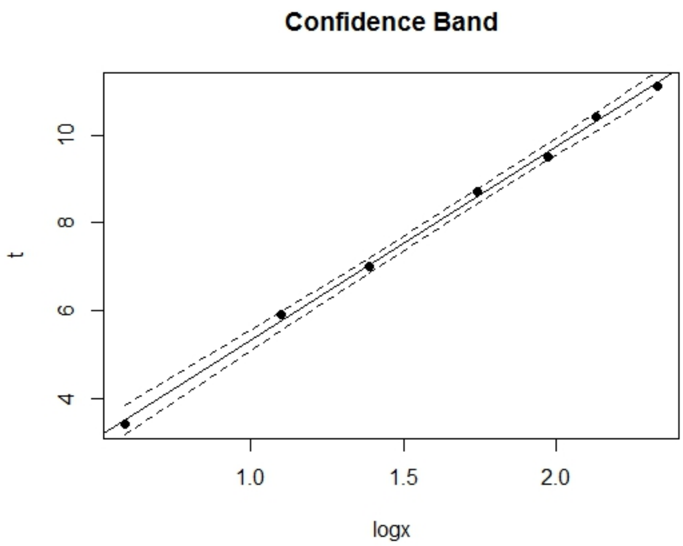
Q14.(By Feng Xiangnan @ LSB G24)

(1.)

$$\begin{aligned} \exp(t) &= Ax^B \\ t &= \log(Ax^B) = \log A + B\log x \text{.....(2pt)} \\ t &= a + b\log x \\ a &= \log A; \quad b = B \text{.....(2pt)} \end{aligned}$$

(2.)

$$\begin{aligned} t &= a + b\log x \\ \hat{b} &= \frac{\sum t \log x - n\bar{t}\,\overline{\log x}}{\sum (\log x)^2 - n\overline{\log x}^2} \text{.....(2pt)} \\ \hat{b} &= 4.4074 \text{.....(1pt)} \\ \hat{a} &= \bar{t} - \overline{\log x}\,\hat{b} \text{.....(2pt)} \\ \hat{a} &= 0.9181 \text{.....(1pt)} \end{aligned}$$



plot.....(1pt)
line.....(1pt)

This fitting is appropriate as the data set is approximately a straight line......(2pt)
You may answer this question with R^2 or other relevant concepts.

(3.)

W = $\sqrt{2 * F(\alpha, 2, 5)}$ = 3.401804

sefit:

```
> W*sqrt(sigma2 * t(xmatrix[,1])%*%invM%*%xmatrix[,1])
0.3030883
> W*sqrt(sigma2 * t(xmatrix[,2])%*%invM%*%xmatrix[,2])
0.1724965
> W*sqrt(sigma2 * t(xmatrix[,3])%*%invM%*%xmatrix[,3])
0.1691638
> W*sqrt(sigma2 * t(xmatrix[,4])%*%invM%*%xmatrix[,4])
0.2087156
> sqrt(sigma2 * t(xmatrix[,1])%*%invM%*%xmatrix[,1])
0.08909634
> sqrt(sigma2 * t(xmatrix[,2])%*%invM%*%xmatrix[,2])
0.05070736
> sqrt(sigma2 * t(xmatrix[,3])%*%invM%*%xmatrix[,3])
0.04972767
> sqrt(sigma2 * t(xmatrix[,4])%*%invM%*%xmatrix[,4])
0.0613544

> yhat [1] 3.973127 7.028126 8.815186 10.083125
```

.....Some points are given based on the above information

$$\hat{y} \pm \hat{\sigma}\sqrt{(p+1)F(\alpha, p+1, n-(p+1))(\mathbf{x}'(X'X)^{-1}\mathbf{x})}$$ .....(2pt)

$$(X'X)^{-1} = \begin{pmatrix} 1.2673 & -0.6998 \\ -0.6998 & 0.4355 \end{pmatrix}$$

$$
\begin{aligned}
\hat{\sigma} &= 0.1252; \\
\mathbf{x_1} &= (1, \log(2))' \\
\hat{y}_1 &\pm \hat{\sigma}\sqrt{2F(95\%, 2, 5)(\mathbf{x_1'}(X'X)^{-1}\mathbf{x_1})} \text{.....(1pt)} \\
&(3.6700, 4.2762)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x_2} &= (1, \log(4))' \\
\hat{y}_2 &\pm \hat{\sigma}\sqrt{2F(95\%, 2, 5)(\mathbf{x_2'}(X'X)^{-1}\mathbf{x_2})} \text{.....(1pt)} \\
&(6.8556, 7.2006)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x_3} &= (1, \log(6))' \\
\hat{y}_3 &\pm \hat{\sigma}\sqrt{2F(95\%, 2, 5)(\mathbf{x_3'}(X'X)^{-1}\mathbf{x_3})} \text{.....(1pt)} \\
&(8.6460, 8.9844)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x_4} &= (1, \log(8))' \\
\hat{y}_4 &\pm \hat{\sigma}\sqrt{2F(95\%, 2, 5)(\mathbf{x_4'}(X'X)^{-1}\mathbf{x_4})} \text{.....(1pt)} \\
&(9.8744, 10.2918)
\end{aligned}
$$

If all the calculation and plot(should be a curve) are right .....(2pt)

(4.)

$$
\frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2} \leq F(\alpha, p+1, n-p-1) \text{.....(4pt)}
$$

$$
(X'X) = \begin{pmatrix} 7 & 11.2476 \\ 11.2476 & 20.3687 \end{pmatrix}
$$

$$
\begin{aligned}
&\hat{\sigma^2} = 0.01568 \\
&223.285 * a^2 - 3572.52 * a + 15706.8 + 717.548 * a * b - 6385.91 * b + 649.718 * b^2 \leq F(0.95, 2, 5) \\
&= 5.786
\end{aligned}
$$

.....(2pt)

plot and three points.....(2pt)

x=1

4.4552

4.2692

x=0.5

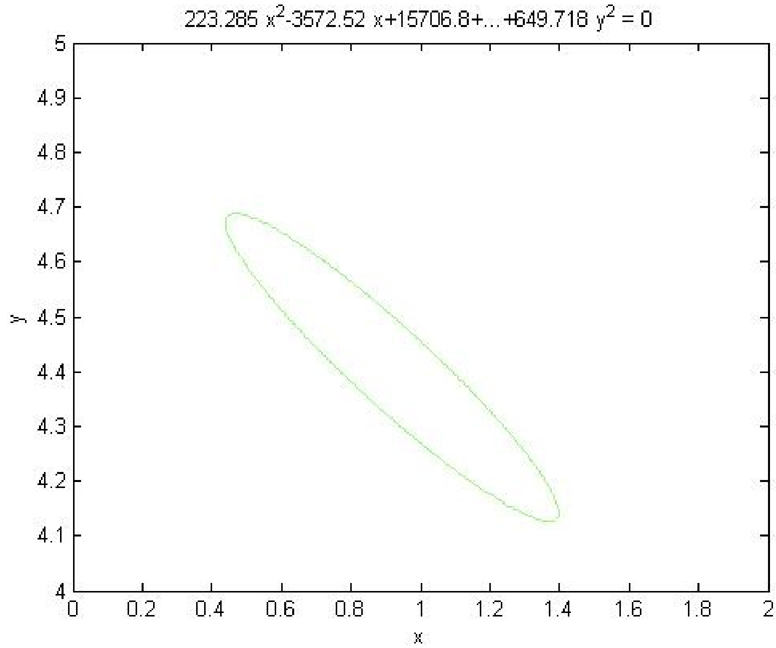
4.6844

4.5921

x=0.7

4.6119

4.4438



Q15.(By Pan Deng@LSB126)

Set $Y = \beta_1 X_1 + \beta_2 X_2 + e$, then

(i)

$$\begin{aligned} \hat{e}_{Y|X_2} &= Y - \bar{Y} && \text{(2 marks)} \\ \hat{e}_{X_1|X_2} &= X_1 - \bar{X}_1 && \text{(2 marks)} \\ slope &= \frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X}_1)}{\sum_{i=1}^{n}(x_i - \bar{X}_1)^2} = \hat{\beta}_1 && \text{(1 mark)} \end{aligned}$$

(ii)

$$\begin{aligned} \hat{e}_{Y|X_1} &= Y - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} X_1 && \text{(2 marks)} \\ \hat{e}_{X_2|X_1} &= X_2 - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} X_1 && \text{(2 marks)} \\ slope &= \frac{\sum_{i=1}^{n}(y_i - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} x_i)(1 - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} x_i)}{\sum_{i=1}^{n}(1 - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} x_i)^2} \\ &= \bar{Y} - \hat{\beta}_1 \bar{X} && \text{(1 mark)} \end{aligned}$$

Q16.(By Pan Deng@LSB126)

$$X'Y = |Y| \cdot |X| \cdot \cos\theta \qquad \text{(1 mark)}$$

where θ is the angle between the vectors Y and X.

$$\begin{aligned} OA &= |Y| \cdot \cos\theta = \frac{X'Y}{|X|} = (\sqrt{X'X})^{-1} X'Y && \text{(2 marks)} \\ \hat{Y} &= OA \cdot \frac{X}{|X|} = X(X'X)^{-1}X'Y = cX && \text{(2 marks)} \end{aligned}$$

Where $c = (X'X)^{-1}X'Y$.