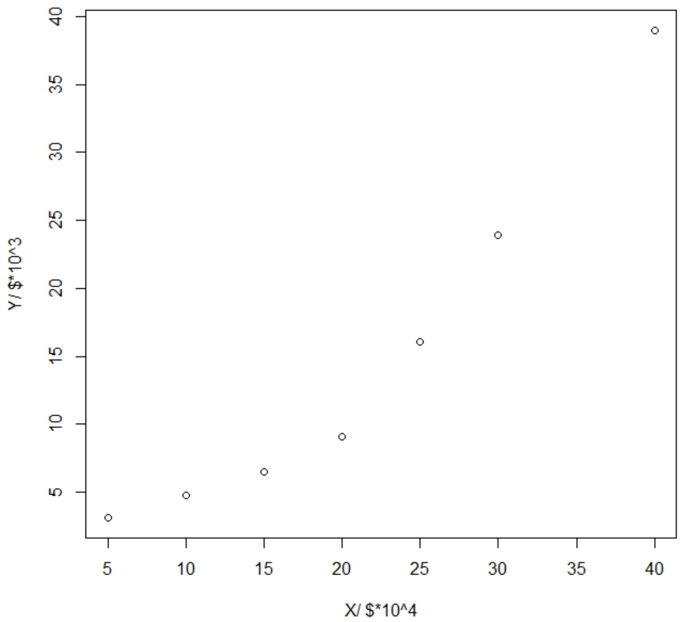
STAT 3008 Applied Regression Anylysis

Midterm Solution

March 2013

1.

a) Scatter plot....1pt

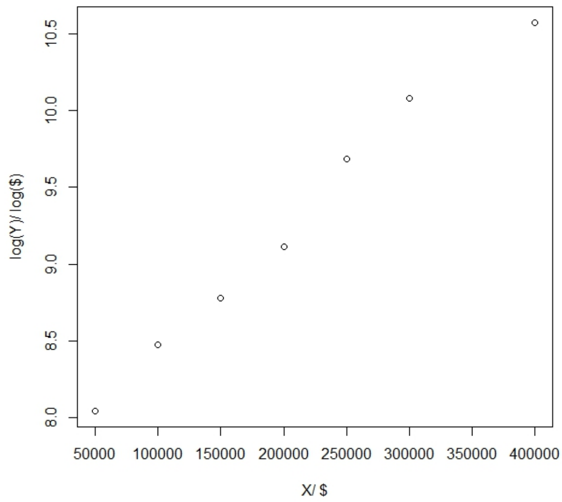


axes ....2pt; shape 1pt

Linear regression is not suitable....1pt, because the scatter plot is not a straight line....1pt.

b)

$$\begin{aligned}
\log y &= \log A + Bx + e \\
\therefore f(y) &= \log y \\
a &= \log A \\
b &= B \\
\epsilon &= e. \qquad \text{....4pt}
\end{aligned}$$



axes ....2pt; shape 1pt

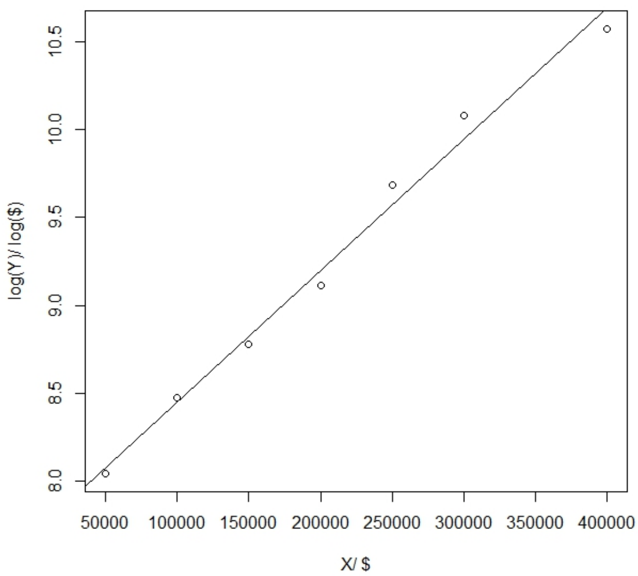
The actuary's suggestion is good. We can use linear regression now....1pt.

c)

$$n = 7\text{....1pt}$$

$$\hat{b} = \frac{\sum x_i f(y_i) - \frac{1}{n}\sum x_i \sum f(y_i)}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2} = 7.468 \times 10^{-6}\text{....formular 2pt, result 1pt}$$

$$\hat{a} = \overline{f(y)} - \hat{b}\bar{x} = \frac{1}{7} \times 64.757 - \hat{b} \cdot \frac{1450,000}{7} = 7.704\text{....formular 2pt, result 1pt}$$



draw the line....2pt

When $x = 250,000$,

$\hat{\log y} = \hat{a} + \hat{b} \cdot 250,000 = 9.571$....formular 1pt, result 1pt

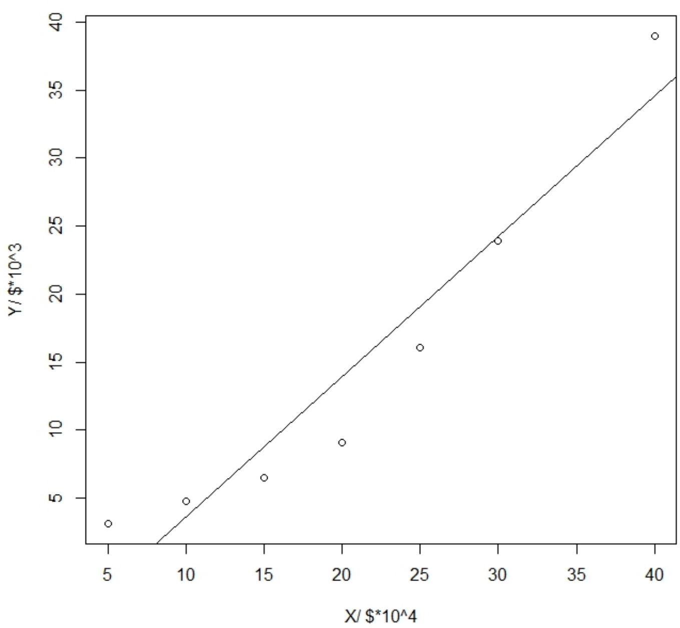
$\hat{y} = 14342.75$....1pt.

d)

	SS	DF	MS	F	p-value
SSreg	4.8596(1pt)	1(0.5pt)	4.8596(1pt)	426.2807(1pt)	<0.001(1pt)
RSS	0.0569(1pt)	5(0.5pt)	0.0114(1pt)		
TSS	4.9165(1pt)				

table headers....2pt

e)



....1pt

$\hat{y} = -6723 + 0.1032 \times 250000 = 19077$....1pt

The estimation in part c) is more reliable. As in part c), the regression is valid/in part e) the regression is not valid....2pt.

2.

a)

$$\hat{\beta} = (X'X)^{-1} X'Y = \begin{pmatrix} 1.313 & -0.1417 & -0.2204 \\ -0.1417 & 0.0284 & 0.0164 \\ -0.2204 & 0.0164 & 0.0532 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 3 & 5 \\ 1 & 0 & 6 \\ 1 & 8 & 3 \\ 1 & 5 & 0 \end{pmatrix}' \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \\ 8 \\ 2 \end{pmatrix} = \begin{pmatrix} -0.6205 \\ 0.6508 \\ 0.8108 \end{pmatrix}.$$

....formular 2pt, result 2pt

$$\hat{Y} = X\hat{\beta} = \begin{pmatrix} 3.1135 \\ 1.6519 \\ 5.3859 \\ 4.2443 \\ 7.0183 \\ 2.6335 \end{pmatrix},$$

$$\hat{e} = Y - \hat{Y} = \begin{pmatrix} -1.1135 \\ 1.3481 \\ -1.3859 \\ 0.7557 \\ 0.9817 \\ -0.6335 \end{pmatrix}.$$

....3pt

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-3} = 2.3047.$$

....formular 2pt, result 1pt.

b) $H_0 : \beta_1 = 0 \leftrightarrow H_A : \beta_1 \neq 0$ 1pt

t-statistic $= \dfrac{\hat{\beta}_1}{\hat{\sigma}\sqrt{V_{11}}} = \dfrac{0.6508}{\sqrt{2.3047}\sqrt{0.0284}} = 2.5438,$ where V_{11} is the $(2, 2)$ entry of $(X'X)^{-1}$.

....formular 2pt, result 1pt

Since t-critical value $= t_{0.05,3} = 2.3534 < 2.5438$, we reject H_0....1pt

c) $H_0 : Y = \beta_0 \leftrightarrow H_1 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X2$....1pt

$$\begin{aligned} RSS_{H_0} &= \sum (Y_i - \bar{Y})^2 = 26\text{....formular 2pt, result 1pt} \\ RSS_{H_A} &= \hat{e}'\hat{e} = 6.9141\text{....2pt} \\ \text{F-statistic} &= \frac{RSS_{H_0} - RSS_{H_A}/(df_{H_0} - df_{H_A})}{RSS_{H_A}/df_{H_A}} \\ &= \frac{(26 - 6.9141)/(5-3)}{6.9141/3} = 4.1406\text{....formular 2pt, result 1pt.} \end{aligned}$$

Since F-critical value $= F_{0.05,2,3} = 9.5521 > 4.1406$, we do not reject H_0....1pt.

d) $X_* = (1, 6, 4)'$, $\hat{Y} = X_*'\hat{\beta} = \begin{pmatrix} 1 & 6 & 4 \end{pmatrix} \begin{pmatrix} -0.6205 \\ 0.6508 \\ 0.8108 \end{pmatrix} = 6.5275$....2pt

The 95% prediction interval is

$$\hat{Y} \pm \hat{\sigma} t_{0.025,3} \sqrt{1 + X_*'(X'X)^{-1}X_*} \Rightarrow [0.5902, 12.4648].$$

....formular 2pt, result 1pt.

e)

$$\begin{aligned} &Var(\hat{\beta}_1 - \hat{\beta}_2) \\ &= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2) \\ &= (0.0284 + 0.0532 - 2 \times 0.0164)\sigma^2 \\ &= 0.0488\sigma^2. \end{aligned}$$

....formular 3pt, result 1pt.

$$\therefore \hat{Var}(\hat{\beta}_1 - \hat{\beta}_2) = 0.0488\hat{\sigma}^2 = 0.1125\text{....1pt}$$

f)

$$\text{t-statistic} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\hat{Var}(\hat{\beta}_1 - \hat{\beta}_2)}} = \frac{0.6508 - 0.8108}{\sqrt{0.1125}} = -0.4770.$$

....formular 3pt, result 1pt

Since t-critical value $= t_{0.025,3} = 3.1824 > |\text{t-statistic}| = 0.477$, we do not reject H_0....1pt.

3.

$$E(\hat{\beta}) = (X'X + kI)^{-1}X'E(Y) = (X'X + kI)^{-1}X'X\beta\text{....5pt}$$

4.

1) $20 - 2 - 1 = 17$....2pt

2) $1, 17$....3pt

5.

Since $\sum \hat{e}_i = 0$, the residuals of model 1) are $(-0.415, 0.108, -1.13, 0.994, 0.443)$....1pt
Similarly, the residuals of model 2) are $(0.3798, -0.373, -0.0868, -1.21, 1.29)$....1pt

$$\therefore \hat{\beta}_1 = \frac{\sum \hat{e}_{1i}\hat{e}_{2i}}{\sum \hat{e}_{2i}^2} = -0.2138.$$

....formular 2pt, result 1pt

6.

a)

$$\begin{aligned} RSS &= Y'(I - H)Y \\ &= (\beta'X' + e')(I - H)(X\beta + e) \\ &= e'(I - H)e \qquad\quad \because X'H = X' \text{ and } HX = X \end{aligned}$$

....2pt

b)

$$RSS = e'(I - H)e = e'QDQ^Te\text{....1pt}$$

Denote Q^Te as $\tilde{e}$. So $RSS = \tilde{e}'D\tilde{e}$, where $\tilde{e}\ N(\mathbf{0}, I\sigma^2)$....1pt
Since D is a diagnoal matrix with $n - p$ ones and p zeros in the diagonal entries,

$$RSS = \sum_{i=1}^{n-p} \tilde{e}_i^2$$

is the sum of $n - p$ mutually independent normally distributed variables with mean 0 and variance σ^2. So RSS is $\sigma^2\chi^2_{n-p}$ distributed....1pt.