Name: CHAN King Yeung
SID: 1155119394
STAT3008 Assignment 1

Question 1

a) $y_i = \beta x_i^2 + e_i \Rightarrow e_i = y_i - \beta x_i^2$ $\qquad g(\beta) = \sum_{i=1}^n \left(y_i - \beta x_i^2\right)^2$ $\qquad \frac{dg}{d\beta} = -2\sum_{i=1}^n \left(y_i - \beta x_i^2\right)x_i^2$

$$\left.\frac{dg}{d\beta}\right|_{\hat\beta} = 0$$
$$0 = -2\sum_{i=1}^n \left(y_i - \hat\beta x_i^2\right)x_i^2$$
$$= \sum_{i=1}^n x_i^2 y_i - \hat\beta \sum_{i=1}^n x_i^4$$
$$\hat\beta = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}$$

$$RSS = \sum_{i=1}^n \left(y_i - \hat\beta x_i^2\right)^2$$
$$E(RSS) = (n-1)\sigma^2$$
$$\hat\sigma^2 = \frac{1}{n-1}\sum_{i=1}^n \left(y_i - \hat\beta x_i^2\right)^2$$

b) $E(\hat\beta) = E\left(\frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}\right) = \frac{E\left(\sum_{i=1}^n x_i^2 y_i\right)}{\sum_{i=1}^n x_i^4} = \frac{\sum_{i=1}^n x_i^2 E(y_i)}{\sum_{i=1}^n x_i^4} = \frac{\sum_{i=1}^n x_i^2 E(\beta x_i^2)}{\sum_{i=1}^n x_i^4} = \frac{\beta \sum_{i=1}^n x_i^4}{\sum_{i=1}^n x_i^4} = \beta$

$\therefore$ $\hat\beta$ is an unbiased estimator for β

c) $\hat y = \hat\beta x_i^2 = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}\left(\sqrt{\overline{x^4}}\right)^2 = \frac{\sum_{i=1}^n x_i^2 y_i}{n} = \overline{x^2 y}$

The regression passes through $\left(\sqrt{\overline{x^4}}, \overline{x^2 y}\right)$

$\hat y = \hat\beta x_i^2 = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}(\bar x)^2 = \frac{\sum_{i=1}^n x_i^2 y_i \sum_{j=1}^n x_j^2}{n^2 \sum_{i=1}^n x_i^4} = \frac{\overline{x^2 y}\cdot\overline{x^2}}{\sum_{i=1}^n x_i^4}$

The regression does not pass through $(\bar x, \bar y)$

d) $y_i = \beta x_i^2 + e_i \sim N(\beta x_i^2, \sigma^2)$ $\qquad L(\beta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{\sum_{i=1}^n \left(y_i - \beta x_i^2\right)^2}{-2\sigma^2}}$ $\qquad l(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i^2)^2$

$$\left.\frac{\partial l}{\partial \beta}\right|_{\tilde\beta, \tilde\sigma^2} = 0$$
$$0 = \frac{1}{\tilde\sigma^2}\sum_{i=1}^n \left(y_i - \tilde\beta x_i^2\right)x_i^2$$
$$= \sum_{i=1}^n x_i^2 y_i - \tilde\beta \sum_{i=1}^n x_i^4$$
$$\tilde\beta = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}$$

$$\left.\frac{\partial l}{\partial \sigma^2}\right|_{\tilde\beta, \tilde\sigma^2} = 0$$
$$0 = -\frac{n}{2\tilde\sigma^2} + \frac{1}{2\tilde\sigma^4}\sum_{i=1}^n \left(y_i - \tilde\beta x_i^2\right)^2$$
$$n\tilde\sigma^2 = \sum_{i=1}^n \left(y_i - \tilde\beta x_i^2\right)^2$$
$$\tilde\sigma^2 = \frac{1}{n}\sum_{i=1}^n \left(y_i - \tilde\beta x_i^2\right)^2$$

Since $\frac{\partial^2 l}{\partial \beta} = -\frac{1}{\sigma^2}\sum_{i=1}^n \left(y_i - \beta x_i^2\right)x_i^4 < 0$, β is maximised; Since $\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{4\sum_{i=1}^n \left(y_i - \tilde\beta x_i^2\right)^2}{\sigma^6} < 0$, σ^2 is maximised

e) $\sum_{i=1}^n x_i^2 y_i = 72 \quad \sum_{i=1}^n x_i^4 = 34 \quad \sum_{i=1}^n y_i^2 = 156 \quad \sum_{i=1}^n x_i^2 = 10 \quad \sum_{i=1}^n y_i = 22$

$$\hat\beta = \frac{72}{34} \approx 2.1176$$
$$\hat\sigma^2 = \frac{1}{n-1}\left(\sum y_i^2 - 2\hat\beta \sum x_i^2 y_i + \hat\beta^2 \sum x_i^4\right) = \frac{1}{4}\left[156 - 2\left(\frac{72}{34}\right)(72) - \left(\frac{72}{34}\right)^2(34)\right] \approx 0.8824$$
$$\sum_{i=1}^n \hat e_i = \sum_{i=1}^n \left(y_i - \hat\beta x_i^2\right) = \sum_{i=1}^n y_i - \hat\beta \sum_{i=1}^n x_i^2 = 22 - \frac{72}{34}(10) \approx 0.8235$$

$\therefore$ The sum of residuals does not equal to zero

Question 2

$$\sum_{i=1}^n \hat e_i = \sum_{i=1}^n (y_i - \hat y_i) = \sum_{i=1}^n (y_i - \bar y) + \hat\beta_1 \sum_{i=1}^n (\bar x - x_i) = 0$$
$$\sum_{i=1}^n x_i \hat e_i = \sum_{i=1}^n x_i y_i - n\bar x \bar y - \hat\beta_1\left(\sum_{i=1}^n x_i^2 - n\bar x^2\right) = SXY - \frac{SXY}{SXX}SXX = 0$$

$$\hat\rho(x, \hat e) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar x)\left(\hat e_i - \bar{\hat e}\right) = \frac{1}{n-1}\left(\sum_{i=1}^n x_i \hat e_i - \bar{\hat e}\sum_{i=1}^n x_i - \bar x \sum_{i=1}^n \hat e_i + n\bar x \bar{\hat e}\right) = -\frac{1}{n-1}\bar{\hat e}\sum_{i=1}^n (x_i - \bar x) = 0$$

Question 4

a) $SXY = 3373.75 - 11(73.14545)(3.95455) \approx 191.9193$

$SXX = 60961.94 - 11(73.14545)^2 \approx 2109.1146$

$SYY = 202.25 - 11(3.95455)^2 \approx 30.2269$

b) $\hat{\beta}_1 = \frac{191.9193}{2109.1146} \approx 0.091$

$\hat{\beta}_0 = 3.95455 - 0.091(73.14545) \approx -2.7017$

$\hat{\sigma}^2 = \frac{1}{9}\left(30.2269 - \frac{191.9193^2}{2109.1146}\right) \approx 1.4181$

c) $\hat{V}ar(\hat{\beta}_0|X) = 1.4181\left(\frac{1}{11} + \frac{73.14545^2}{2109.1146}\right) \approx 3.7263$

$\hat{V}ar(\hat{\beta}_1|X) = 1.4181\left(\frac{1}{2109.1146}\right) \approx 0.0007$

d) $\hat{e}_{74.5,2} = 2 - (-2.7017) - 0.91(74.5) = -2.0778$

$2\hat{\sigma} = 2\sqrt{1.4181} \approx 2.3817$

$|\hat{e}_{74.5,2}| < 2\hat{\sigma}$

$\therefore$ The observation is not an outlier

e) $n = 12$ $\quad \bar{x}^* = 71.2416625$ $\quad \bar{y}^* = 3.875004167$ $\quad \sum_{i=1}^{n} x_i^{*2} = 63492.03$ $\quad \sum_{i=1}^{n} y_i^2 = 211.25$

$\sum_{i=1}^{n} x^* y^* = 3524.65$ $\quad SXY = 211.9091314$ $\quad SXX = 2587.536291$ $\quad SYY = 31.06211247$

$\hat{\beta}_1^* = \frac{211.9091314}{2587.536291} \approx 0.0819$

f) $\hat{\beta}_0^* = 3.875004167 - 0.0819(71.2416625) \approx -1.9597$

$\hat{\sigma}^{*2} = \frac{1}{10}\left(31.06211247 - \frac{211.9091314^2}{2587.536291}\right) \approx 1.3708$

Comment Summary