

## MAST20005/MAST90058: Week 6 Solutions

- Note that  $K = \sum_i (x_i - \bar{x})^2 = (n-1)s_x^2 = 594.9896$ . A 95% confidence interval for  $\beta$  is  $\hat{\beta} \pm c\hat{\sigma}/\sqrt{K}$ , where  $c$  is the 0.975 quantile of  $t_{60}$ . In this case we get  $0.75169 \pm 2.00 \times 0.6943/\sqrt{594.9896}$  or  $(0.69, 0.81)$ .
- (a) Use the double expectation formula,

$$\mu_3 = \mathbb{E}(X_3) = \mathbb{E}(\mathbb{E}(X_3 | X_1, X_2)) = \alpha + \beta_1 \mathbb{E}(X_1 - \mu_1) + \beta_2 \mathbb{E}(X_2 - \mu_2) = \alpha.$$

- (b) Note that  $\sigma_{i3} = \text{cov}(X_i, X_3) = \mathbb{E}(X_i X_3) - \mathbb{E}(X_i) \mathbb{E}(X_3) = \mathbb{E}(X_i X_3) - \mu_i \mathbb{E}(X_3) = \mathbb{E}(X_i X_3) - \mathbb{E}(\mu_i X_3) = \mathbb{E}(X_i X_3 - \mu_i X_3) = \mathbb{E}(X_3(X_i - \mu_i))$ . Using this together with the double expectation formula,

$$\begin{aligned} \sigma_{13} &= \mathbb{E}(X_3(X_1 - \mu_1)) \\ &= \mathbb{E}[\mathbb{E}(X_3(X_1 - \mu_1) | X_1, X_2)] \\ &= \mathbb{E}[\mathbb{E}(X_3 | X_1, X_2)(X_1 - \mu_1)] \\ &= \mathbb{E}[\alpha(X_1 - \mu_1) + \beta_1(X_1 - \mu_1)^2 + \beta_2(X_2 - \mu_2)(X_1 - \mu_1)] \\ &= \beta_1 \sigma_1^2 + \beta_2 \sigma_{12} \end{aligned}$$

and similarly,

$$\sigma_{23} = \beta_1 \sigma_{12} + \beta_2 \sigma_2^2.$$

Simultaneously solving these two equations gives the result.

- (a) Firstly, adding and subtracting the parameter estimate terms:

$$\sum_{i=1}^n [Y_i - \alpha_0 - \beta(x_i - \bar{x})]^2 = \sum_{i=1}^n [(\hat{\alpha}_0 - \alpha_0) + (\hat{\beta} - \beta)(x_i - \bar{x}) + Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})]^2.$$

In expanding this, consider the cross-terms and recall that  $\hat{\alpha}_0 = \bar{Y}$  and  $\hat{\beta} = \sum(x_i - \bar{x})(Y_i - \bar{Y}) / \sum(x_i - \bar{x})^2$ .

$$\begin{aligned} \sum_{i=1}^n (\hat{\alpha}_0 - \alpha_0)(\hat{\beta} - \beta)(x_i - \bar{x}) &= (\hat{\alpha}_0 - \alpha_0)(\hat{\beta} - \beta) \sum_{i=1}^n (x_i - \bar{x}) = 0 \\ \sum_{i=1}^n (\hat{\alpha}_0 - \alpha_0)\{Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})\} &= (\hat{\alpha}_0 - \alpha_0) \left( \sum_{i=1}^n (Y_i - \bar{Y}) - \hat{\beta} \sum_{i=1}^n (x_i - \bar{x}) \right) = 0 \\ \sum_{i=1}^n (\hat{\beta} - \beta)(x_i - \bar{x})\{Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})\} &= (\hat{\beta} - \beta) \left( \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) - \hat{\beta} \sum_{i=1}^n (x_i - \bar{x})^2 \right) \\ &= 0 \end{aligned}$$

Note that we substituted for  $\hat{\beta}$  in the last equation. Therefore the cross-terms disappear and we are left with,

$$\sum_{i=1}^n [Y_i - \alpha_0 - \beta(x_i - \bar{x})]^2 = n(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\beta} - \beta)^2 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n [Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})]^2$$

as required.

(b) We know that  $(\hat{\alpha}_0 - \alpha_0)/\sqrt{\hat{\sigma}^2/n} \sim t_{n-2}$ . Hence,

$$\Pr \left( -c < \frac{\hat{\alpha}_0 - \alpha_0}{\sqrt{\hat{\sigma}^2/n}} < c \right) = 1 - \gamma$$

where  $c$  is the  $1 - \gamma/2$  quantile of  $t_{n-2}$ . Rearrangement then gives the desired confidence interval.

(c) We know  $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi_{n-2}^2$ . Therefore,

$$\Pr \left( F^{-1}(\gamma/2) < \frac{(n-2)\hat{\sigma}^2}{\sigma^2} < F^{-1}(1 - \gamma/2) \right) = 1 - \gamma$$

which gives

$$\Pr \left( \frac{(n-2)\hat{\sigma}^2}{F^{-1}(1 - \gamma/2)} < \sigma^2 < \frac{(n-2)\hat{\sigma}^2}{F^{-1}(\gamma/2)} \right) = 1 - \gamma$$

as required.

4. A linear model needs to be linear in the coefficients  $\beta_1$  and  $\beta_2$ , so this is not a linear model.
5. Firstly,

$$\begin{aligned} \frac{\partial h}{\partial \beta_1} &= -2 \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2) \\ \frac{\partial h}{\partial \beta_2} &= -2 \sum_{i=1}^n x_i (y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2) \\ \frac{\partial h}{\partial \beta_3} &= -2 \sum_{i=1}^n x_i^2 (y_i - \beta_1 - \beta_2 x_i - \beta_3 x_i^2). \end{aligned}$$

Setting these to zero and simplifying gives the normal equations.

6. (a) Firstly,

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i - \bar{y} \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})y_i - \bar{y} \times 0 = \sum_{i=1}^n (x_i - \bar{x})y_i.$$

A similar argument shows that,

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i(y_i - \bar{y}).$$

We also have,

$$\sum_{i=1}^n x_i(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \bar{y} = \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}.$$

That completes the proof.

(b)

$$\begin{aligned}
d^2 &= \sum_{i=1}^n (y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2 = \sum_{i=1}^n (y_i - \bar{y} - \hat{\beta}(x_i - \bar{x}))^2 \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 - 2\hat{\beta} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 - 2\hat{\beta} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\beta} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.
\end{aligned}$$

7. (a)  $\hat{\beta} = \frac{741.1}{578.8} = 1.28$ ,  
 $\hat{\alpha} = 313.2/13 - \hat{\beta} \times 230/13 = 1.44$ ,  
 $\hat{\sigma} = \sqrt{(1000.2 - 741.1^2/578.8)/11} = 2.16$
- (b)  $\text{se}(\hat{\alpha}) = \hat{\sigma} \sqrt{1/13 + (230/13)^2/578.8} = 1.70$ ,  
 $\text{se}(\hat{\beta}) = \hat{\sigma} / \sqrt{578.8} = 0.0898$
- (c) 95% CI for  $\mu(18)$ :  $\hat{\mu}(18) \pm c \times \text{se}(\hat{\mu}(18))$   
 $\hat{\mu}(18) = \hat{\alpha} + \hat{\beta} \times 18 = 24.48$   
 $\text{se}(\hat{\mu}(18)) = 2.16 \times \sqrt{1/13 + (18 - 230/13)^2/578.8} = 0.278$   
 $c = 2.201$  (0.975 quantile of  $t_{11}$ )  
This gives the following interval:  $23.2 < \mu(18) < 25.8$
- (d) 95% PI for  $Y(18)$ :

$$24.48 \pm 2.201 \times 2.16 \times \sqrt{1 + \frac{1}{13} + \frac{(18 - 230/13)^2}{578.8}}$$

This gives the following interval:  $19.6 < Y(18) < 29.4$