# University of Technology Sydney
## School of Mathematical and Physical Sciences

Mathematical Statistics (37262) –
Class 7 Preparation Work
SOLUTIONS

1.

i) $\bar{x} = \dfrac{-5-4-3-2-1}{5} = -3$ and $\bar{y} = \dfrac{-0.8-1.1-1.4+1.8+3.0}{5} = 0.3$.

| | | | | | |
|---|---|---|---|---|---|
| $y$ | -0.8 | -1.1 | -1.4 | 1.8 | 3.0 |
| $x$ | -5 | -4 | -3 | -2 | -1 |
| $y-\bar{y}$ | -1.1 | -1.4 | -1.7 | 1.5 | 2.7 |
| $x-\bar{x}$ | -2 | -1 | 0 | 1 | 2 |
| $(y-\bar{y})(x-\bar{x})$ | 2.2 | 1.4 | 0 | 1.5 | 5.4 |
| $(x-\bar{x})^2$ | 4 | 1 | 0 | 1 | 4 |

We now have $\hat{\beta} = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \dfrac{10.5}{10} = 1.05$

and $\hat{\alpha} = \bar{y} - \beta\bar{x} = -0.3 - 1.05(-3) = 3.45.$

ii) We define $\boldsymbol{X} = \begin{pmatrix} 1 & -5 \\ 1 & -4 \\ 1 & -3 \\ 1 & -2 \\ 1 & -1 \end{pmatrix}$, $\boldsymbol{Y} = \begin{pmatrix} -0.8 \\ -1.1 \\ -1.4 \\ 1.8 \\ 3.0 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

and $\boldsymbol{V}_n = \begin{pmatrix} 1 & 0.5 & 0.25 & 0.125 & 0.0625 \\ 0.5 & 1 & 0.5 & 0.25 & 0.125 \\ 0.25 & 0.5 & 1 & 0.5 & 0.25 \\ 0.125 & 0.25 & 0.5 & 1 & 0.5 \\ 0.0625 & 0.125 & 0.25 & 0.5 & 1 \end{pmatrix}$.

We therefore have

$$\boldsymbol{V}_n^{-1} = \frac{4}{3}\begin{pmatrix} 1 & -0.5 & 0 & 0 & 0 \\ -0.5 & 1.25 & -0.5 & 0 & 0 \\ 0 & -0.5 & 1.25 & -0.5 & 0 \\ 0 & 0 & -0.5 & 1.25 & -0.5 \\ 0 & 0 & 0 & -0.5 & 1 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{V}_n^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^t\boldsymbol{V}_n^{-1}\boldsymbol{Y}) \approx \begin{pmatrix} 3.494 \\ 0.9885 \end{pmatrix}$$

Note that, although this second model "fits less well" in (as in greater sum of squared residuals compared to the initial model), it may well be a better model if the initial assumption of independent residuals was incorrect.

2. i)

As we have ten observations for each city, we have

$$\sum_{i=1}^{30} x_{Ai} = \sum_{i=1}^{30} x_{B_i} = \sum_{i=1}^{30} x_{C_i} = 10$$

The sample means across each city are

$\overline{y}_A = 173.2$, $\overline{y}_B = 174.2$ and $\overline{y}_C = 170.9$. The overall sample mean is therefore

$$\overline{y} = \frac{5183}{30} \approx 172.767$$

We therefore obtain a model of

$$y_i = 172.767 + 0.433 x_A + 1.433 x_B - 1.867 x_C + \varepsilon_i$$

ii)

$$\begin{aligned} SST &= (179-172.767)^2 + (185-172.767)^2 + ... + (179-172.767)^2 \\ &\approx 3251.367 \end{aligned}$$

and

$$\begin{aligned} SSE &= (179-173.2)^2 + (185-173.2)^2 + ... + (179-170.9)^2 \\ &\approx 3194.1 \end{aligned}$$

so $SSR = SST - SSE \approx 3251.367 - 3194.1 \approx 57.267$.

Our $F$-statistic for testing the null hypothesis that $\beta_A = \beta_B = \beta_C = 0$ is then

$F = \dfrac{\frac{SSR}{2}}{\frac{SSE}{21}} \approx 0.242$. $P(F_{2,27} > 0.242) \approx 0.787$ hence we do not reject the null hypothesis. There is no reason to believe that the population heights differ between the three cities.

(Note the 2 degrees of freedom in the numerator, since we have 4 parameters to estimate but 2 constraints – the mean and also the fact that $x_A + x_B + x_C = 1$.

This becomes more clear when we consider that we could consider the model as $y_i = \alpha + \beta_A x_{Ai} + \beta_B x_{Bi} + \beta_C x_{Ci} + \varepsilon_i$ or

$y_i = (\alpha + \beta_A) + (\beta_B - \beta_A) x_{Bi} + (\beta_C - \beta_A) x_{Ci} + \varepsilon_i$, which is more clearly three parameters – $(\alpha + \beta_A), (\beta_B - \beta_A)$ and $(\beta_C - \beta_A)$.)