University of Technology Sydney
School of Mathematical and Physical Sciences

Mathematical Statistics (37262) –
Class 7 Preparation Work

1. Consider the dataset:

y	-0.8	-1.1	-1.4	1.8	3.0
x	-5	-4	-3	-2	-1

i) Fit a simple linear regression $y_i = \alpha + \beta x_i + \varepsilon_i$ to this dataset, under the assumption that the residuals are all independent and identically distributed realisations of a normal variable.

We now consider this as an autoregressive AR(1) model such that the residuals are all correlated with $Cov(\varepsilon_i, \varepsilon_i) = \sigma^2 0.5^{|i-j|}$ for some σ^2.

ii) Refit the linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$ with this assumed covariance structure.

Hint: You may assume without proof that the least squares estimate of the model $\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_n)$ is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{V}_n^{-1} \boldsymbol{X})^{-1} (\boldsymbol{X}^t \boldsymbol{V}_n^{-1} \boldsymbol{Y})$

$$\frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 \\ 0 & -p & 1+\rho^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho \\ 0 & 0 & \ldots & \rho & 1 \end{pmatrix}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \rho^2 & p & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho \\ \rho^{n-1} & \rho^{n-2} & \ldots & \rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

2. Consider the dataset giving the heights (in cm) of adults recorded in three different cities.

City A	179	185	186	162	159	184	171	169	172	165
City B	182	195	173	165	165	161	178	187	162	174
City C	161	177	187	169	163	181	148	171	173	179

i) By first coding the data with indicator variables x_A, x_B and x_c such that, for example, $x_A = 1$ if the observation is from City A and 0 otherwise, fit the linear regression model $y_i = \alpha + \beta_A x_{Ai} + \beta_B x_{Bi} + \beta_C x_{Ci} + \varepsilon_i$ where y describes the observed height and the residuals ε_i are assumed to be independent and identically distributed realisations of a normal variable.

ii) By calculating an appropriate test statistic, would you conclude that there is sufficient evidence to believe that the population heights in the three cities differed? (Use 5% significance level.)