University of Technology Sydney
School of Mathematical and Physical Sciences

Mathematical Statistics (37262) –
Tutorial 7

1. Consider the dataset:

y	-1.6	12.3	28.2	40.0	44.3	47.5	40.5	37.7	25.0	15.1	-0.9
x	-5	-4	-3	-2	-1	0	1	2	3	4	5

Initially, this dataset is modelled with a simple linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$ where each $\varepsilon_i \sim N(0, \sigma^2)$ and each is independent of all others.

i) Calculate the sample mean of x and the sample mean of y.

ii) Show that setting the sum of the residuals to zero to estimate the two model parameters gives $\hat{\alpha} = \bar{y} - \beta\bar{x}$ where $\bar{x}$ and $\bar{y}$ are the sample means from part i) and $\hat{\alpha}$ and $\hat{\beta}$ are, respectively, the estimates of α and β.

iii) Show that using ordinary least squares to estimate the two model parameters gives $\hat{\beta} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

iv) Using ordinary least squares, fit the regression model $y_i = \alpha + \beta x_i + \varepsilon_i$ for the dataset.

v) Calculate the R-squared value for this model. Comment on the appropriateness of this model for these datapoints.

Now, the dataset is modelled with the regression model $y_i = \alpha + \beta x_i^2 + \varepsilon_i$.

vi) Using ordinary least squares, fit the regression model $y_i = \alpha + \beta x_i^2 + \varepsilon_i$ for the dataset.

vii) Calculate the R-squared value for this model Comment on the appropriateness of this model for these datapoints.

2. Consider the problem of fitting a simple linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$ where, for example, x_i represents the *i*th observation of the predictor variable x.

Consider the dataset:

y	4.3	4.375	4.6	4.75	5.5	7
x	10	8	5	4	2	1

i) Explain why, for this dataset, it looks reasonable to fit a model of the form $y_i = \alpha + \beta\left(\frac{1}{x_i}\right) + \varepsilon_i$ instead of one of the form $y_i = \alpha + \beta x_i + \varepsilon_i$.

ii) Hence fit the model $y_i = \alpha + \beta\left(\frac{1}{x_i}\right) + \varepsilon_i$ by ordinary least squares.

iii) What is the R-squared statistic for this model fit? Justify your answer.

Note: You should not need to perform any calculation for part iii).