

QUEEN MARY UNIVERSITY OF LONDON

MTH5120

Statistical Modelling I

Exercise Sheet 3

1. (a) Show that the regression sum of squares

$$SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

can be written as

$$SS_R = \frac{S_{xy}^2}{S_{xx}}.$$

- (b) Deduce that the residual sum of squares can be written as

$$SS_E = S_{yy} - \frac{S_{xy}^2}{S_{xx}}.$$

2. For the regression model

$$Y_i = \beta x_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

i.e. a straight line through the origin with the usual assumptions about the errors. In [Exercise Sheet 2](#), you found the least squares estimator of β and its mean and variance.

- (a) Derive the form of a $100(1 - \alpha)\%$ confidence interval for β .
(b) If we define the residual $e_i = Y_i - \hat{\beta}x_i$, show that

$$E[e_i] = 0$$

and that

$$\text{Var}[e_i] = \sigma^2 \left[1 - \frac{x_i^2}{\sum x_i^2} \right]$$

3. A data frame containing the impact of three advertising medias (Youtube, Facebook and newspaper) on sales. Data are the advertising budget in thousands of dollars along with the sales. The advertising experiment has been repeated 200 times.

The data are available in R in the package `datarium`. First of all, you should install the package

```
> install.packages("datarium") # Packages need to be installed only once
```

Then you should upload the data by using the following commands

```
> library(datarium)
> data("marketing", package = "datarium")
> attach(marketing)
```

Run a simple linear regression model regarding the impact of YouTube on the sales. Thus $Y = \text{sales}$ and $X = \text{Youtube}$. Obviously you can play around by including Facebook or the newspapers.

- (a) Plot the data and run a linear regression
- (b) What could you say on the linear regression estimators?
- (c) Carry out a Normality test on the standardized residuals of the model.