

# QUEEN MARY UNIVERSITY OF LONDON

MTH5120  
Exercise Sheet 1

Statistical Modelling I

1. A sample of ten claims and corresponding payments on settlement for household policies is taken from the business of an insurance company. Data were collected from 10 claims and payments above (in units of 100 pounds).

X: Claim	Y: Payment
2.10	2.18
2.40	2.06
2.50	2.54
3.20	2.61
3.60	3.67
3.80	3.25
4.10	4.02
4.20	3.71
4.50	4.38
5.00	4.45

Payments may vary for different claims. Claim,  $X$ , is known exactly and is not random, but we may assume that  $Y$  is random, so that repeated observations of  $Y$  for the same values of  $X$  will vary.

We find  $\sum x_i = 35.4$ ,  $\sum y_i = 32.87$ ,  $\sum x_i^2 = 133.76$ ,  $\sum y_i^2 = 115.2025$  and  $\sum x_i y_i = 123.81$

Other types of function could also describe the relationship well, for example a quadratic polynomial with a very small second order coefficient. However, it is better to use the simplest model which describes the relationship well. This is called *the principle of parsimony*.

- (a) Use R to check the values given for  $\sum x_i$ ,  $\sum y_i$ ,  $\sum x_i y_i$ ,  $\sum x_i^2$  and  $\sum y_i^2$ .
  - (b) Find the mean of  $x$  and  $y$  and  $S_{yy}$ ,  $S_{xx}$  and  $S_{xy}$  and check that they agree with the values given in the lecture.
2. (a) Use the command  

```
plot(x, y, main="Plot of Y versus X")
```

to produce a scatterplot of the data. The text after `main=` is the title of your plot. The plot appears in the lower right pane. To save it you should Export it as an image to your directory. Does the relationship between  $y$  and  $x$  seem to be linear?  
(b) What is the estimates of the intercept and of the slope? You can use `lm(y ~ x)`  
(c) Add a fitted line to the scatter plot of the data by using `abline` of the model

(d) What is predict mean value of  $y$  when  $x = 3$  by using `predict(model1, newdata=data.frame(x=3))`?

3. (a) Derive the least squares estimates of  $\alpha$  and  $\beta$  for the centred form of the simple linear regression model given by

$$y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i \quad i = 1, 2, \dots, n.$$

- (b) Check that the estimates do give a minimum in the same way as we saw for the standard form of the simple linear regression model.