QUEEN MARY UNIVERSITY OF LONDON

MTH5120 Statistical Modelling I

Solutions to Exercise Sheet 1

---

1. (a) First of all we need to define the values of x and y from the lectures:

```
> x <- c(2.1,2.4,2.5,3.2, 3.6, 3.8, 4.1, 4.2, 4.5, 5.0)
> y <- c(2.18, 2.06, 2.54, 2.61, 3.67, 3.25, 4.02, 3.71, 4.38, 4.45)
```

Then we compute the $\sum x_i$ and $\sum y_i$ in R

```
> sum(x)
[1] 35.4
> sum(y)
[1] 32.87
```

Then we compute the $\sum x_i y_i$, $\sum x_i^2$ and $\sum y_i^2$ in R

```
> sum(x*y)
[1] 123.81
> sum(x^2)
[1] 133.76
> sum(y^2)
[1] 115.2025
```

(b) First of all we compute the mean of x and y

```
> mean(x)
[1] 3.54
> mean(y)
[1] 3.287
```

Secondly for computing S_{yy}, S_{xx} and S_{xy}, we need to the following commands

```
> sum(x^2)-(sum(x)^2)/10
[1] 8.444
> sum(y^2)-(sum(y)^2)/10
[1] 7.15881
> sum(x*y)-(sum(x)*sum(y))/10
[1] 7.4502
```

Analogously, one can also compute the variance and the covariance of x and y

```
> var(y)*9
[1] 7.15881
> var(x)*9
[1] 8.444
> cov(x,y)*9
[1] 7.4502
```

2. (a) By using the command

```
plot(x,y,main="Settlement payments for claims")
```

we can see the data in the following figure

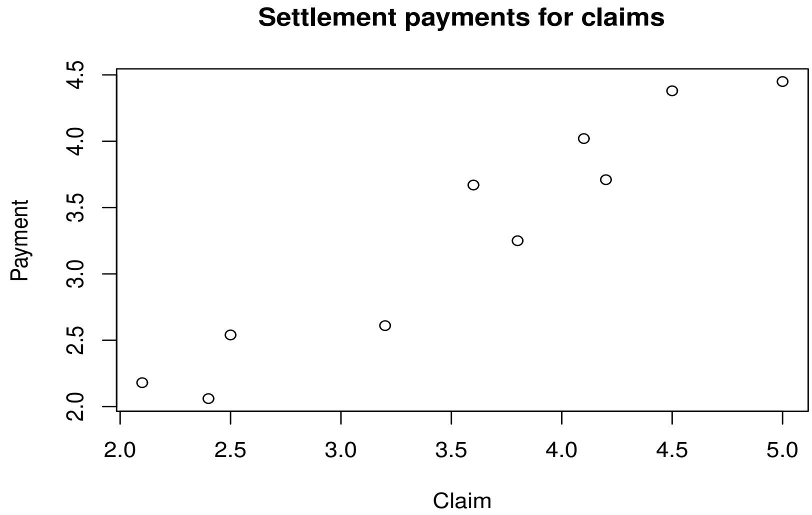

Figure 1.1: Plot of the payments against the claims.

From Figure 1.1, we have a linear relationship between the payment and the claims.

(b) We use the command lm for estimating the intercept and the slope

```
> payment <- lm(y~x)
Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
     0.1636       0.8823
```

By using the command summary we have a better representation of the estimation:

```
> summary(payment)
Call:
lm(formula = y ~ x)
Residuals:
     Min        1Q    Median        3Q       Max
-0.37702 -0.20571  0.01918  0.22183  0.33006
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.16363    0.34048   0.481    0.644
x            0.88231    0.09309   9.478 1.27e-05 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2705 on 8 degrees of freedom
Multiple R-squared:  0.9182,Adjusted R-squared:  0.908
F-statistic: 89.82 on 1 and 8 DF,  p-value: 1.265e-05
```

Thus the estimates of the intercept is 0.16363 and of the slope is 0.88231.

(c) To Figure 1.1, we can add the fitted line of the model previously estimated as stated

```
> plot(x,y,main="Settlement payments for claims")
> abline(payment)
```

and it can be represented in the following Figure.

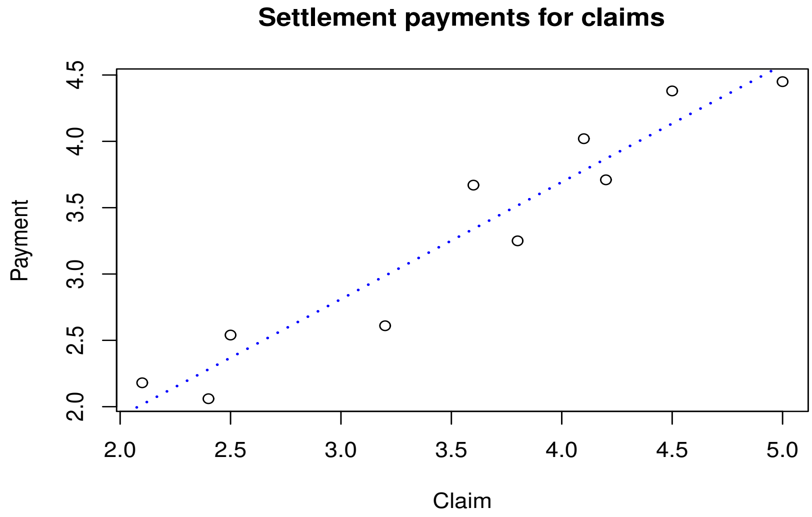

Figure 1.2: Plot of the payments against the claims.

(d) We predict a new value of y such that $x = 3$ by using

```
> predict(payment, newdata=data.frame(x=3))
       1
2.810554
```

Thus the predicted value is 2.81 when $x = 3$.

3. (a) The ***LS estimators*** of the model parameters α and β minimize the sum of squares of errors, that is

$$S(\alpha, \beta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} [y_i - (\alpha + \beta(x_i - \bar{x}))]^2.$$

S is a function of the parameters and so to find its minimum we differentiate it with respect to α and β, then equate the derivatives to zero.

$$\begin{cases} \frac{\partial S}{\partial \alpha} = -2\sum_{i=1}^{n}[y_i - (\alpha + \beta(x_i - \bar{x}))] = 0 \\ \frac{\partial S}{\partial \beta} = -2\sum_{i=1}^{n}[y_i - (\alpha + \beta(x_i - \bar{x}))](x_i - \bar{x}) = 0 \end{cases}$$

This set of equations can be written as

$$\begin{cases} n\widehat{\alpha} + \widehat{\beta}\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} y_i \\ \widehat{\alpha}\sum_{i=1}^{n}(x_i - \bar{x}) + \widehat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \end{cases}$$

Since $\sum_{1=1}^{n}(x_i - \bar{x}) = 0$, the solutions to these equations are

$$\begin{aligned} \widehat{\alpha} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\ &= \bar{y} \end{aligned}$$

and, from the second equation

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad = \frac{S_{xy}}{S_{xx}}.$$

To check that $S(\alpha, \beta)$ attains a minimum at $(\widehat{\alpha}, \widehat{\beta})$ we calculate second derivatives and evaluate the determinant

$$\begin{vmatrix} \frac{\partial^2 S}{\partial \alpha^2} & \frac{\partial S}{\partial \alpha \partial \beta} \\ \\ \frac{\partial S}{\partial \beta \partial \alpha} & \frac{\partial^2 S}{\partial \beta^2} \end{vmatrix} = \begin{vmatrix} 2n & 0 \\ \\ 0 & 2\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{vmatrix} = 4n\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$

for all α, β.

(b) Also, $\frac{\partial^2 S}{\partial \alpha^2} > 0$ for all α, β. This means that the function $S(\alpha, \beta)$ attains a minimum at $(\widehat{\alpha}, \widehat{\beta})$.