

MATH1712 Probability and Statistics II

Homework 4

<http://www1.maths.leeds.ac.uk/~voss/2017/MATH1712/>

Jochen Voss, J.Voss@leeds.ac.uk

2017/18, semester 2

*This homework sheet does not correspond to a tutorial class. Solutions must be submitted online before the deadline of **Monday, 26th February, 4pm.***

Submit your answers here: <https://goo.gl/forms/lqxbWxcWIE2gfbjm2>

The topic of this homework is doing linear regression in R. Help is available in the file “A Short Introduction to R” on the module web site. The instructions below refer to a number $\kappa_2 \in \{1, \dots, 100\}$. This number represents the last two digits of your student ID, plus one. For example, if your student ID is 200123456, then $\kappa_2 = 57$.

Start by loading the following dataset into R, where κ_2 is the number described above:

[http://www1.maths.leeds.ac.uk/~voss/2017/MATH1712/ex04/ \$\kappa_2\$.csv](http://www1.maths.leeds.ac.uk/~voss/2017/MATH1712/ex04/κ_2.csv)

(You need to substitute the numerical value for κ_2 .)

Exercise 13. Load the data into R and use `lm()` to fit a linear model of the form $y_i = \alpha + \beta x_i + \varepsilon_i$ to the data. From the `lm()` output, determine the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$.

Exercise 14. Using the R command `fitted()`, determine the “fitted value” $\hat{y}_1 = \hat{\alpha} + \hat{\beta}x_1$ for the first observation.

Exercise 15. Using the R function `resid()`, determine the value of the residual $\hat{\varepsilon}_2 = y_2 - \hat{y}_2$ for the second observation.

Exercise 16. Produce a scatter plot of the data. You will see that there is one obvious outlier (*i.e.* a value which is very different from the others). Find out which sample is the outlier, *i.e.* find the index i such that the observation (x_i, y_i) is the outlier. (Hint: The command `which()` may be useful here.)