

## Lab 1

Remember: to get help on any command you can type `?command`.

1. Produce a plot of the line  $y = 2 + 3x$ , for any range of  $x$  you choose.
2. Simulate data pairs  $(x_1, y_1), \dots, (x_n, y_n)$ , where  $n = 50$ , using the following R commands:

```
x = 1:50
y = rnorm(50, 2+3*x, 5)
```

Explain in your own words what the R commands are doing. Plot a scatterplot of  $x$  against  $y$ , and superpose an appropriate line on top of the points (if you understand what is going on then there is an “obvious” guess for a line that should fit the points well).

3. Compute a vector of the residuals for the data and line you used in 2 above. Plot a histogram of the residuals (command `hist`), plot a qqplot (command `qqnorm`) of the residuals.
4. What distribution do the residuals have? What are the theoretical expectation and standard deviation ( $\sigma$ ) of the residuals? Compare the empirical mean and standard deviation of the residuals with their theoretical values.
5. Plot a scatterplot of  $x$  against the residuals, and overlay dotted horizontal lines on your scatterplot at the positions  $2\sigma$  and  $-2\sigma$ . What proportion of your residuals lie within 2 standard deviation of 0? What proportion lie within 1 standard deviation? How do these numbers compare with what you would expect from theory?
6. Repeat 2-5 above, but modify the R commands in 2 above so that the residuals have a uniform distribution on  $[-5, 5]$ .