STATG006: Exercise Sheet #4

The exercises in this sheet focus on linear regression. Most of the questions are from James et al., "An Introduction to Statistical Learning" (ISLR). This book is freely available as a PDF file, see link in our Moodle page. As before (see Sheet #3), questions marked with the indication "(COMPUTER IMPLEMENTATION)*" require programming. In particular, ISLR poses questions which explicitly require* R. *Feel free to use an equivalent software package for linear regression, but all solutions provided in Moodle will be based around* R.

1. Exercise 1 of Chapter 3, ISLR.
2. Exercise 3 of Chapter 3, ISLR.
3. Exercise 4 of Chapter 3, ISLR.
4. Exercise 5 of Chapter 3, ISLR.
5. The estimated correlation coefficient (or just **sample correlation**) of two variables X and Y is defined as

$$Cor(X,Y) \equiv \frac{\sum_{i=1}^{n}(x^{(i)}-\bar{x})(y^{(i)}-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x^{(i)}-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}}$$

   for a dataset of bivariate measurements $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\}$. It is a measure of linear association.

   (a) What would this correlation be if each $y^{(i)}$ was a constant multiple of the respective $x^{(i)}$, that is, $y^{(i)} = ax^{(i)}$ for some constant a? What if $y^{(i)} = ax^{(i)} + b$ for some other constant b? Conclude that correlation coefficients are a linear measure of association between -1 and 1, with the extremes corresponding to deterministic linear dependencies.

   (b) Exercise 7 of Chapter 3, ISLR.
6. (COMPUTER IMPLEMENTATION) Exercise 8 of Chapter 3, ISLR.
7. (COMPUTER IMPLEMENTATION) Exercise 9 of Chapter 3, ISLR.
8. (COMPUTER IMPLEMENTATION) Exercise 10 of Chapter 3, ISLR.