

Exercise 7. Linear regression

1. Read the data in the file `xy.txt` into `R`. Compute the sample correlation between x_1 and y_1 then repeat for the other three pairs. Record the values in the table below. Also write down what these values tell you about the association between the four pairs of x and y variables. Plot y_1 against x_1 then repeat for the other three pairs. Do these plots change your ideas about the associations? What implications does this have for the interpretation of correlations?

Pair	1	2	3	4
Correlation				

2. Write down a mathematical representation for the simple linear regression of each y -variable on the corresponding x -variable. Make sure that you know which is the response variable and which is the explanatory variable. Using a calculator instead of \mathbf{R} , estimate the slope and intercept parameters in each of the four cases using the formulae in the lecture notes.

Dataset	$\hat{\beta}_0$	$\hat{\beta}_1$	R^2	p
1				
2				
3				
4				

3. Now perform the linear regressions with **R** and check that the parameter estimates agree with your calculations. What are the values of the coefficient of determination? What does this tell you about the linear models? What is the p -value for testing whether or not the slope is zero? What do you conclude about the explanatory power of the x variable in each case?
4. Compute the residuals from the fitted models and assess the model fits by making diagnostic plots. Are any of the model assumptions inappropriate? Do these plots change your conclusion about the explanatory power of the x variable in each case?