Exercise 1:

You received a dataset with 9 observations and two features:

	1	2	3	4	5	6	7	8	9	$\sum_{i=1}^{n}$
y	-7.79	-5.37	-4.08	-1.97	0.02	2.05	1.93	2.16	2.13	-10.92
x1	-1.00	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75	1.00	0
x2	0.95	0.57	0.29	-0.03	0.02	0.08	0.23	0.54	0.98	3.63

The last column corresponds to the sum of values of each row.

a) Compute the Pearson correlation of x_1 and x_2. The formula is:

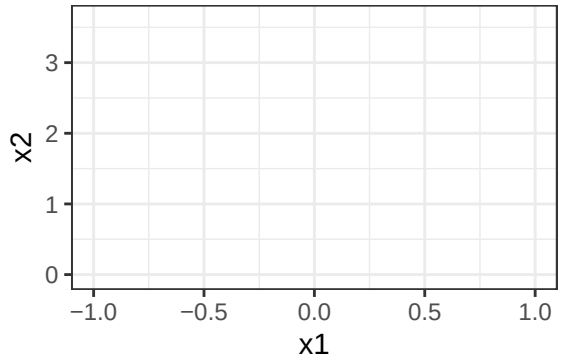
$$\rho(x_1, x_2) = \frac{\sum_{i=1}^{n}(x_1^{(i)} - \overline{x}_1)(x_2^{(i)} - \overline{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1^{(i)} - \overline{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2^{(i)} - \overline{x}_2)^2}}$$

To speed up things, the individual differences to the means ($x_1^{(i)} - \overline{x_1}$, $x_2^{(i)} - \overline{x_2}$), are given in the following table.

	1	2	3	4	5	6	7	8	9
$x_1^{(i)} - \overline{x_1}$	-1.00	-0.75	-0.50	-0.25	0.00	0.25	0.50	0.75	1.00
$x_2^{(i)} - \overline{x_2}$	0.55	0.17	-0.11	-0.43	-0.38	-0.32	-0.17	0.14	0.58

Interprete the results. Based on $\rho(x_1, x_2)$, are x_1 and x_2 correlated?

b) Add points (x_1, x_2) to the following figure:



Based on your drawing, do you consider the Pearson correlation coefficient a reliable measure to detect dependencies for the above use case?

Exercise 2:

Show, that the following holds:

$$R^2 = \rho^2.$$

Recap:

$$\rho = \frac{\sum_{i=1}^{n} (x^{(i)} - \bar{x}) \cdot (y^{(i)} - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x^{(i)} - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y^{(i)} - \bar{y})^2}},$$

$$R^2 = 1 - \frac{SSE_{LM}}{SSE_c},$$

where

$$SSE_{LM} = \sum_{i=1}^{n} (y^{(i)} - \hat{f}_{LM}(x^{(i)}))^2,$$

$$SSE_c = \sum_{i=1}^{n} (y^{(i)} - \bar{y})^2$$

are the sum of squares due to regression and the total sum of squares, respectively.

Exercise 3:

Consider the following function:

$$f(\mathbf{x}) = 2x_1 + 3x_2 - x_1|x_2|.$$

Mathematically check whether interactions are present.