Lab VIII

Linear Regression

1. We aim to determine whether there is a correlation between the AM (Applied Mathematics) exam scores and ALGAED (Linear Algebra and Geometry) scores obtained by first-year ACS students in the 2022–2023 academic year. To this end, the scores of 10 students in these two subjects were observed and are presented in the table below:

AM	8	3	7	6	7	5	4	9	6	4
ALGAED	8	4	5	6	7	5	5	8	6	3

   (a) Calculate the correlation coefficient.

   (b) Determine a linear model describing the dependency of ALGAED scores on AM scores; then represent both the score pairs and the obtained regression line graphically in the same coordinate system.

   (c) Using this model, estimate with a confidence interval the mean ALGAED score if the AM score is 7 ($\alpha = 0.05$).

2. (Machine Learning, Stanford Univ.) The file `ex1data2.txt` contains observations regarding the real estate market in Portland, Oregon, as follows: a sample of size $n = 47$ houses for which the following data is tracked: area (x_1, in ft^2), number of bedrooms ($x_2 \in \mathbb{N}$), and sale price (y, in thousands of dollars).

   (a) Graph the data $(x_{1i}, y_i)_{i=1,n}$ and calculate the correlation coefficient for the given sample.

   (b) Observe that a linear approximation model is suitable for describing the relationship between house area and price. Determine this model.

   (c) Test whether the price of houses depends on their area ($\beta_1 = 0$), at a 95% significance level.

   (d) Determine a prediction interval with a 99% significance level for the price of a house with an area of $x_0 = 2100\text{ft}^2$.

3. Using the dataset `https://www.statlearning.com/s/Advertising.csv`, asses a linear model between TV advertising expenses and sales (see G. James, D. Witten, T. Hastie, R. Tibshirani. *An Introduction to Statistical Learning* and `https://www.statlearning.com/resources-python` for notebooks with Python code). Include statistical tests for the regression line parameters, 95% confidence intervals for them, the coefficient of determination, and analysis of variance.