This is the twelfth homework assignment. Students should tick in TUWEL problems they have solved and upload their detailed solutions by 20:00 on Monday January 15, 2024.

1. **Correlation 1**

   Let X and Z be two uncorrelated random variables with $\mathbb{V}ar(X) = \mathbb{V}ar(Z) = 1$, and let

   $$Y = k \cdot X + \sqrt{1 - k^2} \cdot Z, \quad \text{for } k \in [-1, 1].$$

   Compute the correlation between X and Y.

   Note: Y depends linearly on X and linearly on Z.

2. **Correlation 2**

   Let X and Z be random variables such that $\mathbb{E}(X) = \mu_x$, $\mathbb{V}ar(X) = \sigma_x^2$, $\mathbb{E}(Z) = 0$ and $\mathbb{V}ar(Z) = 1$. Let Y be a random variable with known expectation $\mathbb{E}(Y) = \mu_y$ and variance $\mathbb{V}ar(Y) = \sigma_y^2 > 0$ such that the correlation $\rho = \mathbb{C}orr(X, Y) \in [-1, 1]$ is also given.

   Reperesent Y in the form
   $$Y = \beta_0 + \beta_1 X + \sigma Z$$

   and determine the unknown coefficients b_0, b_1 and σ.

3. **Correlation 3**

   Let $X \sim N(3, 1)$ and $Y \sim N(2, 4)$. Plot 150 realizations of (X, Y) (y_i plotted against x_i, $i = 1, \ldots, 150$) for each of the values of their correlation $\mathbb{C}orr(X, Y) \in \{0, 0.8, -0.5, -1\}$. To each plot add lines
   $$\{(x, y) \in \mathbb{R}^2 \,|\, y = \beta_0 + \beta_1 x\}$$

   with β_0 and β_1 obtained according to Problem 2. Compute the empirical correlation r of the realizations and add it to your plots.

4. **Correlation 4**

   Let $X \sim N(2, 1)$ and $Y = (X - 2)^2$.

   (a) Generate 200 realizations $(x_i)_{i=1,\ldots,200}$ of the random variable X and plot all $y_i = (x_i - 2)^2$ against x_i, $i = 1, \ldots, 200$.

   (b) Compute the empirical correlation r of the realizetions $(x_i)_i$ and $(y_i)_i$.

   (c) Would you say that a lack of correlation implies independence? What does the correlation measure? Would you say that the line is meaningful? Justify your answers.

5. **Regression**
   A lecture was evaluated. In the file `Evaluation.Rdata` you find data of $n = 25$ students. Acquired were first the points achieved in the associated exercises (between 0 and 200 possible), and second the result of the exam (in %). Can the result of the exam be explained by the points achieved in the exercises?

   (a) Plot the result of the exam (y_i) against the exercise points (x_i). Do you observe a relationship?

   (b) Compute the intercept b_0 and the slope b_1 of the regression line (without `lm()`) and plot the regression line. Comment on the meaning of the slope.

   (c) Now perform the analysis using the command `lm()`. Fit the data to the model using the command `lm()`. Plot the data points and the regression line and discuss the plausibility of the model assumptions. Which result in the exam would you predict for a students that achieved 140 points in the exercises? Mark the prediction in the plot.