Practice Tasks for Weeks 8 & 9

On the *Income* worksheet of the file *PracticeData89.xlsx*, the following data are shown for employees of an IT company: monthly income (USD), education level, work experience (years), and the number of training courses completed in the past 5 years. Complete the following tasks:

1. Examine the nature of the relationship between income and education graphically.
2. Calculate and interpret the within-group standard deviation with respect to income and education.
3. Evaluate the strength of the relationship in both the sample and the population.
4. Build a regression model using income as the dependent variable, with the number of trainings and work experience as explanatory variables. Assess the explanatory power of the model.
5. Interpret the intercept and the coefficient of experience.
6. Calculate and interpret the elasticity of the number of trainings at both the mean and median levels.
7. Which variables can be considered to have no effect in the population with 90% confidence? Justify your choice!

On the *Apartment* worksheet of *PracticeData89.xlsx*, data for 500 Dubai apartments are shown, including: price (in thousand AED), area (in square meters), number of bathrooms, and air conditioning status (0: no, 1: yes). Complete the following tasks!

1. Using quintiles, test whether apartment prices can be considered exponentially distributed at conventional significance levels based on the sample data. Clearly state the null and alternative hypotheses, perform the necessary calculations, check the conditions of the test, and provide a written conclusion. Pay attention to precise wording and justification.
2. Test whether the distribution of the number of bathrooms is the same for apartments with and without air conditioning at conventional significance levels. Clearly state the null and alternative hypotheses, perform the necessary calculations, check the assumptions of the test (and if necessary, logically combine categories), and provide a written conclusion. Be careful with precise wording and justification.
3. Identify where and by how much the sample frequency deviates most from the expected frequencies under the assumption of identical number of bathrooms distributions.
4. Construct a regression model in which apartment price is explained by area and number of bathrooms. Interpret the coefficient of area and its 99% confidence interval.
5. Evaluate the explanatory power of the model for unobserved apartments. State the null and alternative hypotheses, perform the necessary calculations, and provide a written conclusion. Use precise terminology and justification.
6. Determine the direct, indirect, and total effect of the number of bathrooms. Interpret these values and their relationships in context.

In Testistan, many 10-story panel buildings have been built in recent decades. The *Flats* worksheet of *PracticeData89.xlsx* contains the following sample data from the country's stock of panel apartments:

- Price: apartment value (in million Testistanian pounds)
- Settlement: the type of settlement where the apartment is located (capital, town, large village, small village)
- Size: apartment size in square meters
- Rooms: number of rooms in the apartment
- Floor: the floor on which the apartment is located (between 0 and 10)

Most panel apartments in Testistan (60%) were built in the capital. The remaining are distributed across cities (33%), towns (5%), and municipalities (2%).

1. We wish to examine whether the sample is representative in terms of settlement type. Can this test be performed on the sample? Justify your answer. If not, adjust the data so the test becomes applicable.
2. Perform the representativeness test and answer the question. Briefly justify your decision with numerical evidence. (Since the significance level is not given, a general conclusion is expected.)
3. Based on the data, conduct a regression analysis where the apartment price is explained by the other variables, excluding settlement type. Which is the most important and the least important explanatory variable in the model, and why?
4. Determine the direct, indirect, and total effect of the Rooms variable. Interpret these values and their relationships in context.
5. Determine and interpret the following!
    a. coefficient of determination
    b. 95% confidence interval of the Floor variable
    c. elasticity of size for apartment number 42