Question: Suppose you have data from an experiment where a firm has introduced a bonus for randomly selected units. You have a data set that includes employees' job satisfaction, a dummy that equals 1 for units treated with the bonus and 0 otherwise, and a dummy that equals 1 for female employees and 0 otherwise. You regress job satisfaction on the bonus dummy, the female dummy and an interaction term between the bonus dummy and the female dummy. Suppose that you find the following estimated coefficients.
Bonus dummy: 0.8
Female dummy: 0.4
Interaction term: 0.3
What is the estimated effect of the bonus on the job satisfaction of female employees?
A) 0.8
B) 0.3
C) 1.1
D) 1.5

Question: In an OLS regression analysis, which of the following best describes the coefficient of determination (R^2)?
A) It measures the total variation in the dependent variable.
B) It indicates the strength and direction of a linear relationship between two variables.
C) It represents the proportion of the variance in the dependent variable that is predictable from the independent variable.
D) It is the same as the slope of the regression line.

Question: What is the purpose of including an interaction term in a regression model?
A) To test whether one independent variable mediates the effect of another independent variable on the dependent variable.
B) To test whether the effect of one independent variable on the outcome depends on the level of another independent variable.
C) To transform the dependent variable.
D) To reduce multicollinearity.

Question: You have a company-level data set (stored in the DataFrame `employee_data`) in which each row is one employee. The column `dummy_exit` equals 1 for employees who have left the company and 0 otherwise. The dummy does not have any missing values. Which of the following does **not** return the proportion of employees who have left the company?
A) `employee_data['dummy_exit'].sum() / len(employee_data)`
B) `employee_data['dummy_exit'].mean()`
C) `employee_data['dummy_exit'].value_counts(normalize=False)`
D) `employee_data['dummy_exit'].value_counts(normalize=True)`

Question: When should the null hypothesis H_0 be rejected?
A) When the absolute value of the t-statistic is greater than the critical value corresponding to the significance level (α).
B) When the absolute value of the t-statistic is less than the critical value corresponding to the significance level (α).
C) When the absolute value of the t-statistic is between the critical values at the significance level (α).
D) When the absolute value of the t-statistic is equal to the significance level (α).

Question: Which of the following is an example of potential outcomes?
A) Estimating the average income of individuals with different education levels in a population.

B) Examining the relationship between education level and income across different individuals.
C) Considering the income of the same individual in two different scenarios: One where they receive a job training program and one where they do not.
D) Measuring the income of a group of people before and after receiving a job training program.

Question: What does supervised learning mean?
A) Supervised learning is a method where the model is evaluated without prior knowledge of the correct answers.
B) Supervised learning involves clustering similar data points together based on their features.
C) Supervised learning is a type of machine learning where the model is trained on labeled data, meaning the input data is paired with the correct output.
D) Supervised learning is a technique that allows models to learn from data without any labeled outcomes.