Statistical Methods in Python - Answer Key (Modules 1 to 10)

Module 1 - Answer Key

1. Descriptive statistics summarize data; inferential statistics make generalizations from samples.

2. a. Interval, b. Ordinal, c. Nominal

3. Between 40 and 60 (mean ± 1 standard deviation)

4. Binomial = fixed number of binary trials; Poisson = rare events over time/space.

5. a - iii, b - i, c - ii

Module 2 - Answer Key

1. H0: No effect/difference; H1: There is an effect/difference.

2. The probability of obtaining the observed result (or more extreme) under the null hypothesis.

3. Type I: False positive, Type II: False negative.

4. When comparing the means of two independent samples.

5. It gives a range in which the true population parameter is likely to fall.

Module 3 - Answer Key

1. -1 to 1

2. Strong negative linear relationship.

3. Expected change in Y for one unit increase in X.

4. Proportion of variance explained by the model.

5. To assess linearity and homoscedasticity.

Module 4 - Answer Key

1. When predictors are highly correlated; detected using VIF.

2. The expected value of Y when all predictors are 0.

3. Adjusted R^2 accounts for number of predictors.

4. Linearity, independence, homoscedasticity, normality, no multicollinearity.

5. To compare impact of predictors on different scales.

Module 5 - Answer Key

1. Ridge uses L2 penalty; Lasso uses L1 (can shrink coefficients to 0).

2. Binary classification problems.

3. True Positive Rate vs False Positive Rate.

4. Area under the ROC curve; closer to 1 is better.

5. Prevents overfitting by shrinking large coefficients.

Module 6 - Answer Key

1. Tree that splits data based on feature conditions.

2. Overfitting and instability.

3. Ensemble of decision trees for improved accuracy.

4. By averaging results across many diverse trees.

5. How much each feature contributed to predictions.

Module 7 - Answer Key

1. Whether group means are significantly different.

2. All group means are equal.

3. When there are two categorical independent variables.

4. A ratio of between-group to within-group variance.

5. To determine which specific group means differ.

Module 8 - Answer Key

1. To estimate uncertainty of a statistic from limited data.

2. Shuffle group labels and compute test statistics repeatedly.

3. Repeated sampling to approximate probabilities.

4. Resample, calculate the statistic, and use percentiles.

5. When assumptions for parametric tests are violated.

Module 9 - Answer Key

1. Problem statement, EDA, modeling, evaluation, conclusions.

2. To uncover patterns, trends, and anomalies before modeling.

3. Visuals, summaries, and clean narratives.

4. Accuracy, R², RMSE, ROC AUC, etc.

5. Helps reinforce learning and identify areas for improvement.

Module 10 - Answer Key

1. Problem statement, EDA, modeling, interpretation, reflection.

2. T-tests, ANOVA, regression, classification, etc.

3. What question was asked, challenges faced, insights gained.

4. It enhances clarity, reproducibility, and peer understanding.

5. Use markdown, comments, and consistent structure.