

Statistical Methods in Python - Quiz Booklet

Module 1: Foundations of Statistical Thinking

1. What is the difference between descriptive and inferential statistics?
2. Identify the data type for each:
 - a. Temperature in Celsius
 - b. Customer satisfaction rating
 - c. Country of origin
3. A dataset has a mean of 50 and standard deviation of 10. What range captures 68% of the data in a normal distribution?
4. What is the main difference between the binomial and Poisson distribution?
5. Match the following:
 - a. Median
 - b. Mode
 - c. Variance
 - i. Most frequent value
 - ii. Average squared deviation
 - iii. Middle value

Module 2: Hypothesis Testing

1. Define null and alternative hypotheses.
2. What does a p-value represent in a hypothesis test?
3. What is the difference between a Type I and Type II error?
4. When would you use a two-sample t-test?
5. What does a confidence interval tell you about a population parameter?

Module 3: Correlation and Simple Regression

1. What is the range of the Pearson correlation coefficient?
2. Describe the meaning of a correlation coefficient of -0.85.
3. What does the slope represent in a linear regression?
4. What does the R-squared value indicate?
5. Why do we examine residual plots?

Module 4: Multiple Linear Regression

1. What is multicollinearity and how can it be detected?
2. What does the intercept mean in a regression equation?
3. How is adjusted R-squared different from R-squared?
4. What assumptions must be checked in MLR?
5. Why might you use standardized coefficients?

Module 5: Advanced Regression Techniques

1. What is the main difference between Ridge and Lasso regression?
2. What type of problems is logistic regression used for?
3. What does a ROC curve show?
4. Define the AUC metric.
5. How does regularization help prevent overfitting?

Module 6: Ensemble Learning and Decision Trees

1. What is a decision tree and how does it work?
2. Name one disadvantage of decision trees.
3. What is a Random Forest?
4. How do Random Forests reduce overfitting?
5. What does feature importance mean in a tree-based model?

Module 7: ANOVA and Beyond

1. What does ANOVA test?
2. What is the null hypothesis in a One-Way ANOVA?
3. When should you use a Two-Way ANOVA?
4. What is an F-statistic?
5. What is the purpose of a post-hoc test?

Module 8: Resampling and Simulation

1. What is the purpose of bootstrap resampling?
2. What are the steps of a permutation test?
3. Define Monte Carlo simulation.
4. How do you estimate a confidence interval using bootstrap?
5. When is resampling preferred over parametric methods?

Module 9: Capstone Projects

1. What are the components of a full data analysis project?
2. Why is exploratory data analysis (EDA) important?
3. What are some tools for presenting findings?
4. How do you evaluate model performance?
5. Why is reflection important after completing a project?

Module 10: Final Assessment & Certification

1. What should be included in your final project notebook?
2. List three key statistical techniques you might apply.
3. What questions should you answer in your reflection section?
4. What is the value of documenting your process clearly?
5. How can you ensure your analysis is reproducible?