

STATISTICS WORKSHEET

Choice questions:

1. - a) True

2. - a) Central Limit Theorem

3. - b) Modeling bounded count data

4. - c) The square of a standard normal random variable follows what is called chi-squared distribution

5.- c) Poisson

6. - b) False

7. - b) Hypothesis

8. - a) 0

9. - d) None of the mentioned

:

Subjective Answers:

10. - The normal distribution is a way to describe how data is spread out. Imagine a bell curve—most data points cluster around the center, where the peak is, and fewer points are found as you move away from the center. This creates a symmetrical shape. In a normal distribution, the average (mean), the middle value (median), and the most frequent value (mode) are all the same.

11. - Dealing with missing data is really important. Here are some common methods:

- Removing Data: If there are only a few missing values, you can just remove those rows or columns.

- Mean/Median/Mode Imputation: Replace missing values with the mean, median, or mode of the column. This is simple but might not always be the best approach.

- Regression Imputation: Use other variables to predict the missing values.

- K-Nearest Neighbors (KNN) Imputation: Fill in missing values based on the nearest neighbors.

- Multiple Imputation: Create several different datasets with different imputed values and combine the results to account for uncertainty.
- Using Algorithms that Handle Missing Data: Some machine learning algorithms can work with missing data directly (e.g., XGBoost).

12. - A/B testing is like a competition between two versions of something to see which one performs better. For example, if you're trying to decide between two website designs, you show one version (A) to half your users and the other version (B) to the other half. By comparing how users interact with each version, you can see which design leads to better results.

13. - Using the mean to fill in missing data is quick and easy, but it can sometimes be too simplistic. It can make your data less varied and might introduce bias if the data isn't missing randomly. It's okay for small amounts of missing data, but for more accurate results, consider more advanced methods like regression or multiple imputation.

14. - Linear regression is a method to understand the relationship between two things. For example, you might want to see how studying time affects test scores. By plotting data points on a graph and drawing the best-fit line, linear regression helps you see the trend. The equation of the line tells you how changes in one variable (like study time) affect the other variable (like test scores).

15. - Statistics is divided into several branches:

- Sure, here are the branches of statistics without bullet points:

Descriptive Statistics: Summarizes data, such as using averages or percentages to describe a group of people.

Inferential Statistics: Makes predictions or inferences about a larger group based on a sample.

Probability Theory: Deals with the likelihood of events happening.

Bayesian Statistics: Updates predictions as new data comes in, using Bayes' theorem.

Multivariate Statistics: Looks at data involving multiple variables to find relationships between them.

Non-parametric Statistics: Analyzes data without assuming it fits a normal distribution.