

STATISTICS WORKSHEET 4

Q1to Q15 are descriptive types. Answer in brief.

1. What is central limit theorem and why is it important?

The central limit theorem is a statistical theorem that states that, given a sufficiently large sample size from a population with any distribution, the mean of the sample will be approximately normally distributed.

This theorem is important because it allows statisticians to make predictions and inferences about a population based on the characteristics of a sample. For example, if you have a sample of 100 values from a population with an unknown distribution, you can use the central limit theorem to estimate the mean and standard deviation of the population based on the mean and standard deviation of the sample.

2. What is sampling? How many sampling methods do you know?

Sampling is the process of selecting a subset of a population for the purpose of studying and making inferences about the population. Sampling is used in statistical analysis to study and make predictions about a population based on characteristics of a sample.

There are several different sampling methods, including:

Simple random sampling:

Stratified sampling:

Cluster sampling:

Systematic sampling

Convenience sampling:

Quota sampling

Snowball sampling:

Non-probability sampling:

3. What is the difference between type I and type II error?

Type I and type II errors are errors that can occur when making statistical inferences about a population based on sample data.

A type I error, also known as a false positive, occurs when you reject the null hypothesis (H_0) when it is actually true. For example, if you are testing a new drug and the null hypothesis is that the drug has no effect, a type I error would occur if you conclude that the drug is effective when it is actually not.

A type II error, also known as a false negative, occurs when you fail to reject the null hypothesis when it is actually false. For example, if you are testing a new drug and the null hypothesis is that the drug has no effect, a type II error would occur if you conclude that the drug is not effective when it is actually effective.

4. What do you understand by the term Normal distribution?

Normal distribution, also known as the Gaussian distribution, is a statistical distribution that is symmetrical and bell-shaped. It is characterized by its mean, median, and mode, which are all equal, and by its standard deviation, which measures the spread of the distribution.

5. What is correlation and covariance in statistics?

In statistics, correlation and covariance are measures of the relationship between two variables.

Correlation is a statistical measure that describes the strength and direction of the relationship between two variables. It is a value between -1 and 1, where -1 indicates a strong negative relationship, 0 indicates no relationship, and 1 indicates a strong positive relationship.

Covariance is a measure of the joint variability of two variables. It is a measure of how much the two variables vary together. Like correlation, covariance can be positive, negative, or zero, depending on the relationship between the variables.

Both correlation and covariance are used to understand the relationship between two variables and to make predictions about one variable based on the other. They are often used in statistical analysis to identify patterns and trends in data.

6. Differentiate between univariate ,Biavariate,and multivariate analysis.

Univariate analysis is a statistical method that involves the analysis of one variable at a time. It is used to describe the characteristics of a single variable, such as its mean, median, mode, range, and standard deviation.

Bivariate analysis is a statistical method that involves the analysis of two variables at a time. It is used to investigate the relationship between two variables and to understand how one variable influences the other.

Multivariate analysis is a statistical method that involves the analysis of three or more variables at a time. It is used to investigate the relationships between multiple variables and to understand how the variables influence each other.

7. What do you understand by sensitivity and how would you calculate it?

Sensitivity is a statistical measure that describes the ability of a test to correctly identify positive cases. It is a measure of the proportion of true positive cases that are correctly identified by the test.

Sensitivity is calculated as the number of true positive cases divided by the total number of positive cases.

P.T.O

8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?

Hypothesis testing is a statistical procedure that is used to test a hypothesis about a population parameter based on sample data. The purpose of hypothesis testing is to determine whether the sample data provides sufficient evidence to support or reject the hypothesis.

In hypothesis testing, the null hypothesis (H0) is a statement of no effect or no relationship. It represents the default assumption that there is no difference between the population parameter and the hypothesized value.

The alternative hypothesis (H1) is a statement of the opposite of the null hypothesis. It represents the hypothesis that is being tested against the null hypothesis

In a two-tail test, the null and alternative hypotheses are symmetrical around the hypothesized value, and the test is designed to detect any deviation from the hypothesized value. For example, in the drug example above, the null hypothesis (H0) would be that the drug has no effect on blood pressure, while the alternative hypothesis (H1) would be that the drug either increases or decreases blood pressure.

In a two-tail test, the null hypothesis is rejected if the sample data falls in either of the two tails of the distribution, indicating that the population parameter is significantly different from the hypothesized value. If the sample data falls in the center of the distribution, the null hypothesis is not rejected.

9. What is quantitative data and qualitative data?

Quantitative data is data that is numerical and can be measured or counted. It is often used to describe the characteristics of a sample or population, such as its mean, median, mode, range, and standard deviation. Quantitative data can be either continuous (e.g., height, weight, age) or discrete (e.g., number of students in a class, number of visits to a doctor).

Qualitative data is data that describes characteristics or attributes of a sample or population, but cannot be measured or counted. It is often used to describe the quality or characteristics of a sample or population, such as their attitudes, opinions, beliefs, or behaviors. Qualitative data can be either nominal (e.g., gender, nationality) or ordinal (e.g., levels of satisfaction, levels of importance).

Both quantitative and qualitative data are important in statistical analysis and can be used together to provide a more complete understanding of a sample or population. Quantitative data can be used to describe the characteristics of a sample or population and to make statistical inferences about the population, while qualitative data can be used to provide context and depth to the analysis

10. How to calculate range and interquartile range?

To calculate the range of a sample or population, you need to find the difference between the highest and lowest values. The range is a measure of the spread of the data and is calculated as follows:

Range = Maximum value - Minimum value

The interquartile range (IQR) is a measure of the spread of the middle 50% of the data. It is calculated as the difference between the third quartile (Q3) and the first quartile (Q1).

11.What do you understand by bell curve distribution ?

A bell curve, also known as a normal distribution or Gaussian distribution, is a statistical distribution that is symmetrical and bell-shaped. It is characterized by its mean, median, and mode, which are all equal, and by its standard deviation, which measures the spread of the distribution

12.Mention one method to find outliers.

One method for identifying outliers in a sample or population is to use the interquartile range (IQR) method.

13.What is p-value in hypothesis testing?

In hypothesis testing, the p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming that the null hypothesis is true. The p-value is used to evaluate the strength of the evidence against the null hypothesis.

If the p-value is small, it indicates that the observed test statistic is unlikely to have occurred by chance if the null hypothesis is true. In this case, the null hypothesis is rejected in favor of the alternative hypothesis.

If the p-value is large, it indicates that the observed test statistic is likely to have occurred by chance if the null hypothesis is true. In this case, the null hypothesis is not rejected

14.What is the Binomial Probability Formula?

The binomial probability formula is a statistical formula used to calculate the probability of a particular outcome in a binomial experiment. A binomial experiment is a statistical experiment with two possible outcomes, such as success or failure, heads or tails, or pass or fail.

15. Explain ANOVA and it's applications.

ANOVA (Analysis of Variance) is a statistical method used to compare the mean of two or more groups. It is used to determine whether there is a significant difference in the means of the groups, and to identify which groups are significantly different from each other. ANOVA is based on the assumption that the data from each group is normally distributed and has equal variances. It uses a test statistic known as the F-value to determine the statistical significance of the differences between the group means. There are several types of ANOVA, including one-way ANOVA, two-way ANOVA, and repeated measures ANOVA. One-way ANOVA is used to compare the means of two or more groups on a single factor or independent variable. Two-way ANOVA is used to compare the means of two or more groups on two factors or independent variables.