

STATISTICS WORKSHEET-4

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

ANS:- The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal. The CLT is a statistical theory that states that - if you take a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from that population will be roughly equal to the population mean.

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

ANS:- Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate the characteristics of the whole population. There are 4 types of Sampling

- Simple Random Sampling
- Stratified Random Sampling
- Cluster Random Sampling
- Systematic Random Sampling.

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

ANS:- In statistics, a Type I error means rejecting the null hypothesis when it's actually true, while a Type II error means failing to reject the null hypothesis when it's actually false.

Ex. Type I error (false positive): the test result says you have coronavirus, but you actually don't. Type II error (false negative): the test result says you don't have coronavirus, but you actually do.

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

ANS:- A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

The mean, median and mode are exactly the same in Normal Distribution

5. What is correlation and covariance in statistics?

ANS:- Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency.

Correlation is a statistical measure that indicates how strongly two variables are related.

Covariance only measures how two variables change in relationship to each other, the correlation coefficient is a measure of the strength of this relationship. While covariance values can be any number in both the negative and positive directions, correlation coefficient values have a range of -1 to +1

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

ANS:- Univariate analysis looks at one variable, Bivariate analysis looks at two variables and their relationship. Multivariate analysis looks at more than two variables and their relationship.

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

ANS:- Sensitivity analysis is the study of how the uncertainty in the output of a mathematical model or system can be divided and allocated to different sources of uncertainty in its inputs.

Assuming this function is $n = f(A, \lambda)$, Then the amplitude sensitivity S at a given $\lambda = (dn/n) / (dA/A)$, one can calculate S at different λ s and plot S versus λ . This is the amplitude sensitivity λ curve.

8. What is hypothesis testing? What is H_0 and H_1 ? What is H_0 and H_1 for two-tail test?

ANS:- The procedure to decide whether the sample data are agreeable or consistent with the null hypothesis is called statistical hypothesis testing or simply hypothesis testing or test of significance.

In hypothesis testing there are two mutually exclusive hypotheses; the Null Hypothesis (H_0) and the Alternative Hypothesis (H_1)

9. What is quantitative data and qualitative data?

ANS:- Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often).

Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

10. How to calculate range and interquartile range?

ANS:- The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q_1) and quartile 3 (Q_3). The IQR is the difference between Q_3 and Q_1

11. What do you understand by bell curve distribution ?

ANS:- A bell curve is a graph depicting the normal distribution, which has a shape reminiscent of a bell. The top of the curve shows the mean, mode, and median of the data collected. Its standard deviation depicts the bell curve's relative width around the mean.

12. Mention one method to find outliers.

ANS:- You can choose from four main ways to detect outliers

- Sorting your values from low to high and checking minimum and maximum values
- Visualizing your data with a box plot and looking for outliers
- Using the interquartile range to create fences for your data
- Using statistical procedures to identify extreme values

13. What is p-value in hypothesis testing?

ANS:- The p value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis

14. What is the Binomial Probability Formula?

ANS:- Binomial probability refers to the probability of exactly x successes on n repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment). If the probability of success on an individual trial is p , then the binomial probability is $nCx \cdot p^x \cdot (1-p)^{n-x}$.

15. Explain ANOVA and it's applications.

ANS:- Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups
ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources