

Statistics Worksheet-4

1. Central Limit Theorem

The central limit theorem (CLT) states that the distribution of sample approximates a normal distribution (also known as a “bell curve”) as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population distribution shape.

Central Limit Theorem exhibits a phenomenon where the average of the sample means and standard deviations equal the population mean and standard deviation, which is extremely useful in accurately predicting the characteristics of populations.

2. Sampling

Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of analysis being performed.

Types of Sampling

1. Random Sampling
2. Systematic Sampling
3. Stratified Sampling
4. Cluster Sampling

4. Normal Distribution

The normal distribution is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distribution are similarly unlikely.

5. Covariance

Covariance measures the directional relationship between the returns on two assets. A positive covariance means that asset returns move together while a negative covariance means they move inversely. Covariance is calculated by analysing at-return surprises (standard deviations from the expected return) or by multiplying the correlation between the two variables by the standard deviation of each variable.

Correlation

The correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0. A calculated number greater than 1.0 or less than -1.0 means that there was an error in the correlation measurement

3. Difference between Type1 and Type2 error

Type 1 error, in statistical hypothesis testing, is the error caused by rejecting a null hypothesis when it is true.	Type II error is the error that occurs when the null hypothesis is accepted when it is not true.	
Also termed	Type I error is equivalent to false positive.	Type II error is equivalent to a false negative.
Meaning	It is a false rejection of a true hypothesis.	It is the false acceptance of an incorrect hypothesis.
Symbol	Type I error is denoted by α .	Type II error is denoted by β .
Probability	The probability of type I error is equal to the level of significance.	The probability of type II error is equal to one minus the power of the test.
Reduced	It can be reduced by decreasing the level of significance.	It can be reduced by increasing the level of significance.
Cause	It is caused by luck or chance.	It is caused by a smaller sample size or a less powerful test.
What is it?	Type I error is similar to a false hit.	Type II error is similar to a miss.

Hypothesis	Type I error is associated with rejecting the null hypothesis.	Type II error is associated with rejecting the alternative hypothesis.
When does it happen?	It happens when the acceptance levels are set too lenient.	It happens when the acceptance levels are set too stringent.

6.Univariate : Univariate statistics summarize only one variable at a time

Bivariate : Bivariate statistics compare two variables.

Multivariate Analysis: Multivariate statistics compare more than two variables

7.Sensitivity

Sensitivity is a measure of the proportion of actual positive cases that got predicted as positive (or true positive).

8. Hypothesis Testing

Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data. The test provides evidence concerning the plausibility of the hypothesis, given the data. Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analysed.

H0

The null hypothesis (H0), stated as the null, is a statement about a population parameter, such as the population mean, that is assumed to be true.

H1

The alternative hypothesis, H1 or Ha, is a statistical proposition stating that there is a significant difference between a hypothesized value of a population parameter and its estimated value. It has no meaning when the decision is that the null hypothesis is true.

9.Quantitative Data

Quantitative data is defined as the value of data in the form of counts or numbers where each data-set has a unique numerical value associated with it. This data is any quantifiable information that can be used for mathematical calculations and statistical analysis.

Qualitative Data

Qualitative data describes qualities or characteristics. It is collected using questionnaires, interviews, or observation, and frequently appears in narrative form.

10. $\text{Range} = \text{max} - \text{min}$

$$\text{IQR} = Q3 - Q1$$

11. Bell Curve Distribution

The term "bell curve" is used to describe a graphical depiction of a normal probability distribution, whose underlying standard deviations from the mean create the curved bell shape.

12. Plotting Scatterplot, Boxplot,

13. P Value in Hypothesis Testing

The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. ... A smaller p-value means that there is stronger evidence in favor of the alternative hypothesis.

14. Binomial Probability Theorem

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. ANOVA

Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.