

## **Worksheet-Set 4: Statistics Assignment**

### **01. What is central limit theorem and why is it important?**

The central limit theorem states that the distribution of the mean of a large number of independent, identically distributed (i.i.d.) samples from a population will be approximately normal, regardless of the shape of the distribution of the population from which the samples are drawn. This theorem is important because it allows statisticians to make predictions and draw conclusions about a population based on sample data, even if the population is not normally distributed.

### **02. What is sampling? How many sampling methods do you know?**

Sampling is the process of selecting a subset of observations from a larger population in order to make inferences about the population. There are many different sampling methods, including simple random sampling, stratified sampling, cluster sampling, and systematic sampling.

### **03. What is the difference between type I and type II error?**

In statistical hypothesis testing, a type I error is a false positive, meaning that the null hypothesis ( $H_0$ ) is rejected when it is actually true. A type II error is a false negative, meaning that the null hypothesis is not rejected when it is actually false.

### **04. What do you understand by the term Normal distribution?**

A normal distribution is a type of continuous probability distribution that is bell-shaped and symmetrical. It is characterized by its mean, median, and standard deviation.

### **05. What is correlation and covariance in statistics?**

Correlation is a statistical measure of the relationship between two variables, while covariance is a measure of the degree to which two variables vary together.

### **06. Differentiate between univariate, Bivariate, and multivariate analysis.**

Univariate analysis involves the study of one variable at a time, bivariate analysis involves the study of two variables, and multivariate analysis involves the study of three or more variables.

### **07. What do you understand by sensitivity and how would you calculate it?**

Sensitivity is a measure of the ability of a test to correctly identify positive cases. It is calculated as the number of true positive results divided by the total number of actual positive cases.

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

Hypothesis testing is a statistical procedure used to determine whether there is sufficient evidence in a sample of data to infer that a certain condition or relationship holds true for the larger population. The null hypothesis (H0) is a statement of no relationship or no difference between groups, while the alternative hypothesis (H1) is a statement of some relationship or difference. For a two-tail test, H0 and H1 are located at both ends of the distribution and the test statistic can be greater than or less than the critical value.

## 09. What is quantitative data and qualitative data?

Quantitative data is data that can be measured and expressed in numerical form, while qualitative data is data that describes attributes or characteristics.

## 10. How to calculate range and interquartile range?

The range of a set of data is the difference between the largest and smallest values, while the interquartile range is the difference between the upper and lower quartiles (the points that divide the data into four equal groups).

## 11. What do you understand by bell curve distribution?

A bell curve distribution, also known as a normal distribution, is a type of continuous probability distribution that is symmetrical and bell-shaped.

## 12. Mention one method to find outliers.

One method to find outliers in a dataset is to use the interquartile range (IQR). Observations that are more than 1.5 times the IQR above the upper quartile or below the lower quartile are considered to be outliers.

## 13. What is p-value in hypothesis testing?

The p-value in hypothesis testing is the probability of obtaining a test statistic at least as extreme as the one observed, given that the null hypothesis is true.

## 14. What is the Binomial Probability Formula?

The binomial probability formula is used to calculate the probability of a specific number of successes in a fixed number of independent trials, where each trial has only two possible outcomes (success or failure).

It is given by the equation:  $P(x) = \frac{n!}{x!(n-x)!} * p^x * (1-p)^{(n-x)}$ .

where n is the total number of trials and x is the number of successes.

## 15. Explain ANOVA and its applications.

ANOVA (Analysis of Variance) is a statistical procedure used to test whether there are significant differences between the means of two or more groups. It can be used to compare the means of two or more independent groups or the means of a single group across multiple levels of a categorical variable.

There are three main types of ANOVA:

One-way ANOVA: This is used to compare the means of two or more independent groups.

Two-way ANOVA: This is used to compare the means of two or more independent groups while controlling for the effects of another variable (called a confounding variable).

Repeated-measures ANOVA: This is used to compare the means of a single group across multiple levels of a categorical variable.