**ASSIGNMENT - 4**

# STATISTICS WORKSHEET

**Q1. What is central limit theorem and why is it important?**
**Ans.** the Central Limit is a statistical theory because it allows us to safely assume that the sampling distribution of the mean will be normal in most cases. If we 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.

**Q2. What is sampling? How many sampling methods do you know?**
**Ans.** In statistics, sampling is a method when researchers determine a representative segment of a larger population that is then used to conduct a study. Sampling generally comes in two forms — probability sampling and non-probability sampling.

**Q3. What is the difference between type I and type II error?**
**Ans.** Type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; while Type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

**Q4. What do you understand by the term Normal distribution?**
**Ans.** A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range. And the first property of normal distribution is the mean, mode and median are all equal and The curve is symmetric at the center. the theorem says that the sampling distribution of the mean will always be normally distributed, along the sample size is large enough.

**Q5. What is correlation and covariance in statistics?**
**Ans.** correlation to denote association between two quantitative variables. We also assume that the association is linear, that one variable increases or decreases a fixed amount for a unit increase or decrease in the other. And Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency.

**Q6. Differentiate between univariate, Bivariate and multivariate analysis.**
**Ans.** Univariate analysis is the analysis of only one variable at a time. Bivariate analysis compares two variables. and Multivariate analysis compare more than two variables

**Q7. What do you understand by sensitivity and how would you calculate it?**
**Ans.** Sensitivity analysis is used to identify how much variations in the input values for a given variable impact the results for a mathematical model. And it can identify the best data to be collected for analyses to evaluate a project's return on investment (ROI). The sensitivity is calculated by dividing the percentage change in output by the percentage change in input

**Q8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?**
**Ans.** A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.
In hypothesis testing there are two mutually exclusive hypotheses, the Null Hypothesis (H0) and the Alternative Hypothesis (H1). One of these is the claim to be tested and based on the sampling results which infers a similar measurement in the population, the claim will either be supported or not.
**Null hypothesis (H0):** The null hypothesis here is what currently stated to be true about the population. In our case it will be the average height of students in the batch is 100. H0 : $\mu = 100$.
**Alternate hypothesis (H1):** The alternate hypothesis is always what is being claimed

**Q9. What is quantitative data and qualitative data?**
**Ans.** Quantitative data are measures of values or counts and are expressed as numbers. Some examples are length, mass, temperature, and time while Qualitative Data Involves a descriptive judgment using concept words instead of numbers.

**Q10. How to calculate range and interquartile range?**
**Ans.** The range is calculated by subtracting the lowest value from the highest value while Inter Quartile Range (IQR) is the range between the first and third quartile namely Q1 and Q3. (IQR = Q3 – Q1)

**Q11. What do you understand by bell curve distribution?**
**Ans.** A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side.

**Q12. Mention one method to find outliers.**
**Ans.** Inter Quartile Range (IQR) is the range between the first and third quartile namely Q1 and Q3. (IQR = Q3 – Q1) the data point fall below Q1 – 1.5 or above Q3 + 1.5 IQR are outlier.

**Q13. 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.

**Q14. What is the Binomial Probability Formula?**

**Ans.** The formula for binomial Probability is:

$$P(x: n, p) = nCx\ px\ (q)n-x$$

Where

p is the probability of success,

q is the probability of failure and

n= number of trials

**Q15. Explain ANOVA and it's applications.**

**Ans.** ANOVA is a statistical formula used to compare variances across the means of different groups. And it is use to determine if there is any difference between the means of different groups. It is helpful for testing three or more variables and 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.