

# **STATISTICS WORKSHEET-4**

## **Internship-28**

**Submitted By- Anshul Dubey**

**Q1to Q15 are descriptive types. Answer in brief.**

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

**Ans-**The central limit theorem states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough. Importance: It all has to do with the distribution of our population. This theorem allows you to simplify problems in statistics by allowing you to work with a distribution that is approximately normal.

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

**Ans-**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, but it may include simple random sampling or systematic sampling . There are two types of sampling methods: Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group. Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

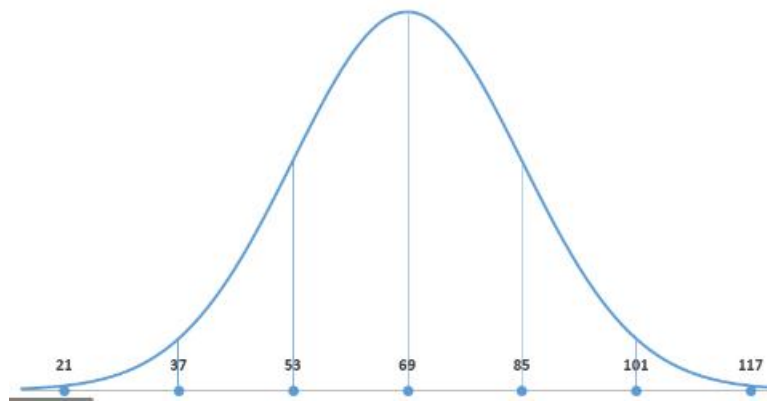
3. What is the difference between type I and type II error?

**Ans-** A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

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

**Ans-**Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

For example:



5. What is correlation and covariance in statistics?

**Ans-**Correlation on the other hand measures both the strength and direction of the linear relationship between two variables.

Covariance indicates the direction of the linear relationship between variables.

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

**Ans-**Univariate statistics summarize only one variable at a time.

Bivariate statistics compare two variables. Multivariate statistics compare more than two variables.

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

**Ans-**Sensitivity (True Positive rate) measures the proportion of positives that are correctly identified (i.e. the proportion of those who have some condition (affected) who are correctly identified as having the condition).

**Sensitivity = number of true positives / (number of true positives + number of false negatives)**

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

**Ans-**Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.

**Alternative Hypothesis:  $H_1$ :** The hypothesis that we are interested in proving. **Null hypothesis:  $H_0$ :** The complement of the alternative hypothesis.

The default null hypothesis for a 2-sample t-test is that the two groups are equal. You can see in the equation that when the two groups are equal, the difference (and the entire ratio) also equals zero.

9. What is quantitative data and qualitative data?

**Ans-**Quantitative data is information about quantities, and therefore numbers, and qualitative data is descriptive, and regards phenomenon which can be observed but not measured, such as language.

10. How to calculate range and interquartile range?

**Ans-**  $\text{Range} = \text{Max} - \text{Min}$

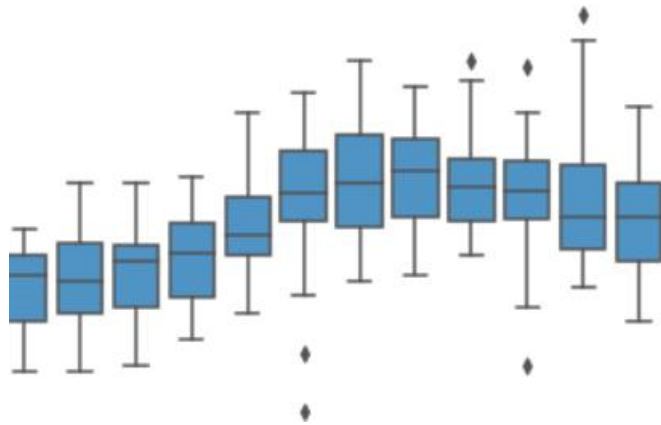
$\text{IQR} = Q_3 - Q_1 \dots \dots \dots (3^{\text{rd}} \text{ quantile} - 1^{\text{st}} \text{ quantile})$

11. What do you understand by bell curve distribution ?

**Ans-**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. A standard deviation is a measurement used to quantify the variability of data dispersion, in a set of given values around the mean.

12. Mention one method to find outliers.

**Ans-**Boxplots display asterisks or other symbols on the graph to indicate explicitly when datasets contain outliers. These graphs use the interquartile method with fences to find outliers. All data points beyond the IQR limit are considered outliers.



13. What is p-value in hypothesis testing?

**Ans-**The P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis ( $H_0$ ) of a study question is true – the definition of 'extreme' depends on how the hypothesis is being tested.

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:

$${}^nC_x \cdot p^x \cdot (1-p)^{(n-x)}$$

Here  ${}^nC_x$  indicates the number of different combinations of  $x$  objects selected from a set of  $n$  objects.

15. Explain ANOVA and its applications

**Ans-**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.

Eg, A manufacturing plant would most likely use an ANOVA test to determine the best materials to use to build a product for a customer. The company may need to test which metal is the sturdiest to buy from. If the cost of three different types of metals is significantly different in price, the company may be looking for ways to save money but still provide a quality product.

**Thankyou .**