

## **STATISTICS WORKSHEET-4**

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

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

Answer: As long as the sample size is sufficient, the central limit theorem states that the sampling distribution of the mean will always be normally distributed. The sampling distribution of the mean will have a normal distribution whether the population has a normal, Poisson, binomial, or any other type of distribution.

Statistics benefits from the Central Limit Theorem because it makes it safe to assume that the sampling distribution of the mean will be typically normal. As we will see in the next section, this means that we can benefit from statistical methods that presume a normal distribution.

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

Answer: Choosing a group (a sample) from a population from which we will gather data for our research is known as sampling. Sampling is a crucial component of any research study because it has a significant impact on the findings. Therefore, it is important to choose the sampling technique carefully in order to obtain accurate findings or results that can accurately predict the population.

### **Types of Sampling**

**Simple Random Sampling:** The method of simple random sampling is one of the best probability sampling strategies that aids in time and resource saving. It is a trustworthy method of gathering data when each person in a population is chosen at random, just by chance. The likelihood of each person being selected to be a member of a sample is the same.

**Cluster Sampling** Using the cluster sampling technique, researchers separate the total population into groups or clusters that each represent a certain demographic. Based on demographic factors like age, sex, geography, etc., clusters are found and included in a sample. This makes it very easy for a survey developer to draw useful conclusions from the responses.

**Systematic Sampling:** Researchers use this technique to select a population's sample members at predetermined intervals. It necessitates choosing a sample size and starting point that can be repeated on a regular basis. This sort of sampling strategy is the least time-consuming because it uses a specified range.

**Stratified random sampling:** This technique involves breaking the population up into smaller groups that don't overlap but still accurately represent the whole population. These groups can be set up for sampling, and a sample can be taken from each group independently.

3. What is the difference between typeI and typeII error?

Answer:

BASIS FOR COMPARISON	TYPE I ERROR	TYPE II ERROR
Meaning	Type I error refers to non-acceptance of hypothesis which ought to be accepted.	Type II error is the acceptance of hypothesis which ought to be rejected.
Equivalent to	False positive	False negative
What is it?	It is incorrect rejection of true null hypothesis.	It is incorrect acceptance of false null hypothesis.
Represents	A false hit	A miss
Probability of committing error	Equals the level of significance.	Equals the power of test.
Indicated by	Greek letter ' $\alpha$ '	Greek letter ' $\beta$ '

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

Answer: A probability distribution that is symmetric about the mean is the normal distribution, sometimes referred to as the Gaussian distribution. It demonstrates that data that are close to the mean occur more frequently than data that are far from the mean. The normal distribution appears as a "bell curve" on a graph.

5. What is correlation and covariance in statistics?

Answer: A statistical measure known as correlation expresses how closely two variables are related linearly (meaning they change together at a constant rate). It's a typical technique for describing straightforward connections without explicitly stating cause and consequence.

A statistical tool called covariance can be used to compare the motions of two random variables. Two equities are said to have a positive covariance when they move in unison; when they move conversely, the covariance is negative.

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

Answer: Only one variable is analysed at a time in univariate statistics.

Two variables are compared in bivariate statistics.

Multiple variables are compared using multivariate statistics.

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

Answer: "Sensitivity analysis is the study of how the uncertainty in a model's output (whether numerical or not) can be distributed among various sources of uncertainty in the model input."

**Sensitivity:** The sensitivity of a test is its ability to determine the patient cases correctly. To estimate it, we should calculate the proportion of true positive in patient cases. Mathematically, this can be stated as:  $\text{Sensitivity} = \frac{TP}{TP+FN}$

TP: True Positive ,  
FN: False Negative

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

Answer: In statistics, the process of hypothesis testing involves putting an analyst's presumption about a population parameter to the test. The type of data used and the purpose of the study will determine the methodology the analyst uses. Using sample data, hypothesis testing is done to determine whether a claim is plausible.

The Null Hypothesis (H0) and the Alternative Hypothesis are two hypotheses that are mutually exclusive in hypothesis testing (H1). One of these is the claim that will be put to the test, and depending on the findings of the sample (which suggest a comparable measurement in the population), the claim will either be supported or unsupported. Two-tailed Test **H0 :  $\mu = k$  H1 :  $\mu \neq k$**  P-value =  $2P(z > |t|)$  If P-value  $\leq \alpha$ , we reject H0. If P-value  $> \alpha$ , we do not reject H0.

9. What is quantitative data and qualitative data?

Answer: Quantitative: Any information that may be expressed as a number, or as quantitative data, is referred to. It is quantitative in nature if it can be measured, tallied, or given a numerical value. Consider it as a yardstick.

Qualitative: Information that is described in qualitative data cannot be tallied or measured. It alludes to the terminology or names applied to specific traits or attributes.

10. How to calculate range and interquartile range?

Answer: 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 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.

11. What do you understand by bell curve distribution ?

Answer: 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.

12. Mention one method to find outliers.

Answer: Inter Quartile Range Method

13. What is p-value in hypothesis testing?

Answer: The p-value is a figure that, when generated from a statistical test, indicates how likely it is that, if the null hypothesis were true, you would have discovered a certain collection of observations. In order to determine whether to reject the null hypothesis, P-values are utilised in hypothesis testing.

14. What is the Binomial Probability Formula?

Answer: 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

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

15. Explain ANOVA and its applications.

Answer: A statistical procedure called Analysis of Variance (ANOVA) is used to examine variations between the means (or averages) of several groups. It is used in a variety of situations to discover whether there are any differences between the means of various groups.

One Way ANOVA – It is also known as one factor ANOVA. Here, we are using one criterion variable (or called as a factor) and analyze the difference between more than two sample groups. Suppose in glass industry, we want to compare the variation of three batches (glass) for their average weight (factor).

Two Way ANOVA – Here, we are using two independent variables (factors) and analyze the difference between more than two sample groups. Similarly, we want to compare the variation of three batches of glass w.r.t weight and hardness (two factors).

Application: The ANOVA test can be used to compare many suppliers and choose the best one. When we have more than two sample groups, we utilise ANOVA (Analysis of Variance) to see if there are any statistically significant differences between the means of two or more independent sample groups. In other words, it compares the means of various samples to determine the impact of one or more factors. We use the one sample and two sample tests when there are fewer than or equal to two sample groups.

---