**NAME OF THE PROJECT**

**Statistics  
Submitted by:**

**SAUNAK MUKHERJEE**



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

The central limit theorem states that if we have a population with mean μ and standard deviation σ and we have to take sufficiently large random samples from the population with replacement, then the distribution of the sample mean is asymptotically normal.

We can calculate the mean of the sample means for the random samples we choose from the population:

μX¯=μ

As well as the standard deviation of sample means:

σX¯=σn

According to the central limit theorem, the form of the sampling distribution will approach normalcy as the sample size is sufficiently large (usually n>30). regardless of the population distribution.

**Importance of Central Limit Theorem:**

This is useful since the researcher never knows which mean in the sampling distribution which is corresponds to the population mean, but by taking numerous random samples from a population, the sample means will cluster together, allowing the researcher to obtain a very accurate estimate of the population mean.

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

Sampling is a process used in statistical analysis in that is predetermined number of observations are taken from a larger [population](https://www.investopedia.com/terms/p/population.asp).

There are two types of sampling methods that I know.

3. What is the difference between type1 and typeII error?

Type I error which is reject the null hypothesis and Type II error which is faIling to reject null hypothesis. Type I error is false positive and Type II error is false negative.

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

A normal distribution is an arrangement of a data set in that is most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

**5. What is correlation and covariance in statistics?**

Correlation is a statistical measure which is expresses the extent to which two variables are linearly related.

Covariance is a statistical tool which is used to determine the relationship between the movements of two random variables.

**6. Differentiate between univariate ,Biavariate,and multivariate analysis.**

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?**

Sensitivity Index = 100 - Percentage of Change required in the current value of the variable to alter the decision.

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

Hypothesis testing is an act in statistics there is 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 the hypothesis by using sample data.

H0 is null hyphothesis

H1 is Alternative hyphothesis

Two-tailed Test H0 : µ = k H1 : µ 6= k P-value = 2P(z > |t|) If P-value  α, we reject H0. If P-value > α, we do not reject H0. Note: For each formula to find z-scores, if you can assume that x has a normal distribution, then any sample size n will work. If you cannot assume this, use a sample size n 30.

**9. What is quantitative data and qualitative data?**

Quantitative data is data which can be counted or measured in numerical values. The two main types of quantitative data are discrete data and continuous data.

Qualitative data is the descriptive and conceptual that is findings collected through questionnaires, interviews, or observation.

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

**IQR = Q3 – Q1.**

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

 bell curve which is a graph depicting the normal distribution, which has a shape reminiscent of a bell.

**12. Mention one method to find outliers**

Using the distance and density of data points for outlier detection.

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

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

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

 P(r) = nCr · pr (1 − p)n−r.

**15. Explain ANOVA and it’s applications.**

 It is a type of statistical test used to determine if there is a statistically significant difference between two or more categorical groups by testing for differences of means using variance

a linear modeling method for evaluating the relationship among fields.

 A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.