

STATISTICS WORKSHEET-4

Q1 to Q15 are descriptive types. Answer in brief.

1. What is central limit theorem and why is it important?
2. What is sampling? How many sampling methods do you know?
3. What is the difference between type I and type II error?
4. What do you understand by the term Normal distribution?
5. What is correlation and covariance in statistics?
6. Differentiate between univariate, bivariate, and multivariate analysis.
7. What do you understand by sensitivity and how would you calculate it?
8. What is hypothesis testing? What is H_0 and H_1 ? What is H_0 and H_1 for two-tail test?
9. What is quantitative data and qualitative data?
10. How to calculate range and interquartile range?
11. What do you understand by bell curve distribution?
12. Mention one method to find outliers.
13. What is p-value in hypothesis testing?
14. What is the Binomial Probability Formula?
15. Explain ANOVA and its applications.

Answer

- 1) The CLT is a statistical theory that states that - if you 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.

It is Important

The CLT has several applications. Look at the places where you can use it.

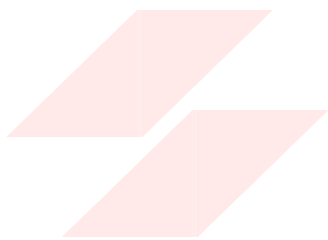
- Political/election polling is a great example of how you can use CLT. These polls are used to estimate the number of people who support a specific candidate. You may have seen these results with confidence intervals on news channels. The CLT aids in this calculation.
 - You use the CLT in various census fields to calculate various population details, such as family income, electricity consumption, individual salaries, and so on.
- 2) Probability sampling methods include **simple random sampling, systematic sampling, stratified sampling, and cluster sampling**. What is non-probability sampling? In non-probability sampling, the sample is selected based on non-random criteria, and not every member of the population has a chance of being included.
 - 3) 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) A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution
 - 5) Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related
 - 6) . Univariate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables
 - 7) The technique used to determine how independent variable values will impact a particular dependent variable under a given set of assumptions is defined as **sensitive analysis**. Its usage will depend on one or more input variables within the specific boundaries, such as the effect that changes in interest rates will have on a bond's price.
 - 8) In a jury trial the hypotheses are: H_0 : defendant is innocent; • H_1 : defendant is guilty. H_0 (innocent) is rejected if H_1 (guilty) is supported by evidence beyond "reasonable doubt." Failure to reject H_0 (prove guilty) does not imply innocence, only that the evidence is insufficient to reject it
 - 9) Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.
 - 10) 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) 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) Sorting method

You can **sort quantitative variables** from low to high and scan for extremely low or extremely high values. Flag any extreme values that you find.

This is a simple way to check whether you need to investigate certain data points before using more sophisticated methods.

- 13)The p-value is defined as the probability of obtaining the result at least as extreme as the observed result of a statistical hypothesis test, assuming that the null hypothesis is true
- 14) the binomial probability distribution is $P(r) = {}^nC_r \cdot p^r (1 - p)^{n-r}$.
- 15) ANOVA is **helpful for testing three or more variables**. It is 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



FLIP ROBO

