

ANSWERSHEET

STATISTICS

ANSWER: 01 Theorem states that the sampling distribution of the mean approaches a normal distribution, as the sample size increases. This fact holds especially true for sample sizes over 30. This theorem tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach normality as the sample size (N) increases.

ANSWER: 02 Sampling is a method that allows us to get information about the population based on the statistics from a subset of the population (sample), without having to investigate every individual. Some sampling methods are Simple random sampling, Systematic sampling, Stratified sampling and clustered sampling

ANSWER: 03 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.

Type I Error	Type II Error
Type I error tends to assert something that is not really present, i.e. it is a false hit.	type II error fails in identifying something, that is present, i.e. it is a miss.
The probability of committing type I error is the same as the level of significance.	The likelihood of committing type II error is same as the power of the test.
Greek letter ' α ' indicates type I error.	type II error which is denoted by Greek letter ' β '.

ANSWER: 04 Normal distribution is the continuous probability distribution with a probability density function that gives you a symmetrical bell curve. A normal distribution has a probability distribution that is centered around the mean. This means that the distribution has more data around the mean. The data distribution decreases as you move away from the center.

ANSWER: 05 Covariance that only measures the direction of the relationship between two variables, correlation also measures the relationship's strength. Thus, correlation quantifies the relationship between the variables and signifies how strong or weak the relationship is. Covariance is an extension of variance and determines the direction of the relationship between two variables. In other words, covariance indicates whether the two variables are directly proportional or inversely proportional to one other.

ANSWER: 06

Univariate Analysis	Bivariate Analysis	Multivariate Analysis
Univariate analysis is the most basic form of statistical data analysis technique. When the data contains only one variable and doesn't deal with a causes or effect relationships then a Univariate technique is used.	This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables.	Multivariate analysis is a more complex form of statistical analysis technique and used when there are more than two variables in the data set.

ANSWER: 07 Sensitivity Analysis determines how different values of an independent variable affect a particular dependent variable under a given set of assumptions. It can be calculate as

$$\text{Sensitivity} = \frac{\text{Number of True positives}}{\text{Number of true positives} + \text{number of false negatives}}$$

ANSWER: 08 Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test. It is used to estimate the relationship between 2 statistical variables. The Null Hypothesis is the assumption that the event will not occur. A null hypothesis has no bearing on the study's outcome unless it is rejected. H_0 is the symbol for it, and it is pronounced H-naught. The Alternate Hypothesis is the logical opposite of the null hypothesis. The acceptance of the alternative hypothesis follows the rejection of the null hypothesis. H_1 is the symbol for it. In two tails, the test sample is checked to be greater or less than a range of values in a Two-Tailed test, implying that the critical distribution area is two-sided. If the sample falls within this range, the alternate hypothesis will be accepted, and the null hypothesis will be rejected.

ANSWERSHEET

ANSWER: 09 Quantitative data is, quite simply, information that can be quantified. It can be counted or measured, and given a numerical value. Qualitative data is defined as the data that approximates and characterizes. Qualitative data can be observed and recorded. This data type is non-numerical in nature.

ANSWER: 10 The range is calculated by subtracting the lowest value from the highest value. The interquartile range can be calculated $IQR = Q3 - Q1$

ANSWER: 11 A bell curve is a common type of distribution for a variable, also known as the normal distribution. The term "bell curve" originates from the fact that the graph used to depict a normal distribution consists of a symmetrical bell-shaped curve.

ANSWER: 12 Interquartile range method, z scores

ANSWER: 13 The p value, or probability value, tells you how likely it is that your data could have occurred under the null hypothesis. It does this by calculating the likelihood of your test statistic, which is the number calculated by a statistical test using your data.

ANSWER: 14 The binomial distribution formula is for any random variable X, given by; $P(x:n,p) = {}^nC_x p^x (1-p)^{n-x}$
where

- n = the number of experiments
- x = 0, 1, 2, 3, 4, ...
- p = Probability of success in a single experiment
- q = Probability of failure in a single experiment (= 1 – p)

ANSWER: 15 Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups.