

STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

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

ANSWER: The central limit theorem states that if you have a population with a mean and a standard deviation, and you get a sufficiently large random sample from the population by permutation, the distribution of the sample mean is asymptotically normal.

You can compute the mean of sample means for a random sample selected from a population.

$$\mu_{\bar{X}} = \mu$$

As well as the standard deviation of sample means:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

According to the central limit theorem, if the sample size is large enough (usually), the shape of the sample distribution approaches normality. Regardless of population distribution.

Importance of the central limit theorem: This is useful because the researcher cannot know which mean of the sampling distribution corresponds to the population mean, but by drawing a large number of samples randomly from the population, the sample means can be aggregated and aggregated. , allowing researchers to obtain very accurate estimates of the population mean.

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

ANSWER: Sampling is the technique of selecting individual members or subsets of a population to draw statistical inferences and estimate characteristics of the population as a whole. Various sampling techniques are commonly used by market research researchers, so it is not necessary to study the entire population to glean actionable insights.

It also forms the basis of any study design as it is a time-saving and cost-effective method. Sampling techniques can be used in research survey software for optimal derivation.

Types of sampling: sampling methods

Sampling in market action research is of two types – probability sampling and non-probability sampling. Let's take a closer look at these two methods of sampling.

1. Probability sampling: Probability sampling is a sampling technique where a researcher sets a selection of a few criteria and chooses members of a population randomly. All the members have an equal opportunity to be a part of the sample with this selection parameter.
2. Non-probability sampling: In non-probability sampling, the researcher chooses members for research at random. This sampling method is not a fixed or predefined selection process. This makes it difficult for all elements of a population to have equal opportunities to be included in a sample.

3. What is the difference between type1 and typell 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 ' α '	Greek letter ' β '

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

ANSWER: The normal distribution, also called the Gaussian distribution, is a probability distribution that is symmetrical about the mean, indicating that data close to the mean are more common than data far from the mean.

Properties of the normal distribution

The normal distribution has several key features and properties that define it.

First, the mean (mean), median (midpoint), and mode (most common observation) are equal to each other. Furthermore, all of these values represent the peak or highest point of the distribution, which descends symmetrically around the mean and whose width is defined by the standard deviation.

5. What is correlation and covariance in statistics?

ANSWER:

Covariance is a measure of how two random variables vary with each other. Correlation, on the other hand, measures the strength of that relationship. correlation

The value is bound to the upper bound at +1 and the lower bound at -1. So it's a specific area.

However, the covariance range is unlimited. Can be any positive or negative value (theoretically, the range is $-\infty$ to $+\infty$). A correlation of 0.5 is greater than 0.3, and we can see that the first group of numbers (with a correlation of 0.5) is more interdependent than the second group (with a correlation of 0.3). Interpreting such results is difficult due to covariance calculations.

A change in scale affects the covariance. For example, multiplying the values of two variables by similar or different constants affects the calculated covariance of these two numbers. However, applying the same correlation mechanism, multiplying by a constant does not change the previous result. This is because changing the scale does not affect the correlation.

In contrast to covariance, correlation is a unitless measure of the interdependence of two variables.

This makes it easy to compare calculated correlation values between any two her variables, regardless of units or dimensions.

You can compute the covariance

Two variables only. Correlation, on the other hand, can be calculated for some set of numbers. Correlation vs. covariance is another factor that analysts like.

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

ANSWER:

Univariate analysis

Univariate analysis is the most basic form of statistical data analysis technique. If the data contain only one variable and do not correspond to causality, univariate analysis techniques are used.

Bivariate analysis

Bivariate analysis is slightly more analytical than univariate analysis. Bivariate analysis is an appropriate type of analysis technique when a dataset contains two variables and a researcher wants to compare two of his datasets.

Multivariate analysis

Multivariate analysis is a more complex form of statistical analysis technique and is used when there are more than two variables in the data set.

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

ANSWER: Sensitivity analysis is an analytical technique that works on the basis of what-if analysis. B. How independent factors influence dependent factors and are used to predict outcomes when the analysis is performed under certain conditions. It is commonly used by investors to test, predict and evaluate outcomes given the conditions that affect potential investments.

Calculation Of Sensitivity Analysis:

1. Firstly, the analyst is required to design the basic formula, which will act as the output formula. For instance, say NPV formula can be taken as the output formula.
 2. Next, the analyst needs to identify which are the variables that are required to be sensitized as they are key to the output formula. In the NPV formula in excel, the cost of capital and the initial investment can be the independent variables.
 3. Next, determine the probable range of the independent variables.
 4. Next, open an excel sheet and then put the range of one of the independent variable along the rows and the other set along with the columns.
- Range of 1st independent variable.
 - Range of 2nd independent variable.

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

ANSWER: A hypothesis is a statement about a given problem. Hypothesis testing is a statistical technique used to make statistical decisions based on experimental data. A hypothesis test is basically an assumption you make about a population parameter. Evaluate two mutually exclusive statements about the population to determine which statement is best supported by the sample data.

Null Hypothesis (H0):

In statistics, the null hypothesis is the commonly held statement or standard position that there is no relationship between two measured cases or between groups.

In other words, it is a basic assumption or based on knowledge of the problem.

example:

In-house production = 50 pieces/day, etc.

Alternative hypothesis (H1):

An alternative hypothesis is a hypothesis used in hypothesis testing that contradicts the null hypothesis.

example :

In-house production does not support 50 pieces/day.

9. What is quantitative data and qualitative data?

ANSWER: Quantitative data:

Data collected based on numeric variables are quantitative data. Quantitative data is more objective and meaningful. Measure a value and represent it numerically. Data collection is based on "how much" the quantity is. Data in quantitative analysis are represented numerically so that they can be counted or measured. Data is extracted from experiments, surveys, market reports, matrices, etc.

Qualitative data:

Data collected based on categorical variables are qualitative data. Qualitative data are more descriptive and conceptual in nature. Measure data based on data types, collections, or categories. Data collection is based on what quality is given. Qualitative data are classified into different groups based on their characteristics. Data obtained from this type of analysis or research are used for theorization, perception, and development of hypothetical theories. This data is collected from texts, documents, transcripts, audio and video recordings, etc.

10. How to calculate range and interquartile range?

ANSWER:

The procedure to calculate the interquartile range is given as follows:

- Arrange the given set of numbers into increasing or decreasing order.
- Then count the given values. If it is odd, then the center value is median otherwise obtain the mean value for two center values. This is known as Q_2 value. If there are even number of values, the median will be the average of the middle two values.
- Median equally cuts the given values into two equal parts. They are described as Q_1 and Q_3 parts.
- The median of data values below the median represents Q_1 .
- The median of data values above the median value represents Q_3 .
- Finally, we can subtract the median values of Q_1 and Q_3 .
- The resulting value is the interquartile range.

11. What do you understand by bell curve distribution ?

ANSWER: A bell curve is a common type of distribution for a variable, also called a normal distribution. The term "bell curve" comes from the fact that the graph used to represent the normal distribution consists of symmetrical bell curves.

The highest point of the curve, or bell apex, represents the most likely event in the set of data (in this case its mean, mode, and median), while all other possible events are centered around the mean. It is symmetrically distributed. Create a downward slope curve on each side of the peak. The width of the bell curve is represented by its standard deviation.

12. Mention one method to find outliers.

ANSWER: Univariate method

One of the easiest ways to detect outliers is with a boxplot. A boxplot is a graphical representation used to represent the distribution of data. Boxplots use the median, lower quartile, and upper quartile.

Tukey's method defines outliers as values in the data set that are far from the center point, which is the median. The maximum allowable distance from the center of the data is called the cleaning parameter. If the cleaning parameter is very large, the test will be less sensitive to outliers. Conversely, if it is too small, many values will be recognized as outliers.

13. What is p-value in hypothesis testing?

ANSWER: The P-value method is used in hypothesis testing to check the significance of a given null hypothesis. The decision to reject or support it is based on the specified significance level or threshold. This method computes the P-value, which is the test statistic. This statistic gives the probability of finding a value as far apart as the population mean (the sample mean).

P in P values represent probabilities. Based on this probability and significance level, you either reject or do not reject the null hypothesis.

In general, the lower the p-value, the more likely you are to reject the null hypothesis and vice versa.

We also use a Z-table to do this process.

14. What is the Binomial Probability Formula?

ANSWER: The binomial probability of success can be calculated using the binomial probability formula. Binomial probability distribution and normal probability distribution are two probability distribution types. The binomial distribution is a type of distribution in statistics that has two possible outcomes. For example, if you flip a coin and there are only two possible outcomes: head or tail. Similarly, taking the test can have two possible outcomes: pass or fail. A binomial probability distribution is an experiment with the following properties:

- There is a fixed number of trials denoted by n .
- All studies are independent.
- Each attempt results in either "success" or "failure".
- The probability of success remains constant and is denoted by p .

15. Explain ANOVA and its applications.

ANSWER: A common approach to finding reliable treatments is to analyze the number of days it takes a patient to heal. A statistical technique can be used that can compare these three treatment samples and show how these samples differ from each other. Such a technique that compares samples based on their means is called ANOVA. will be

Analysis of variance (ANOVA) is a statistical technique used to test whether the means of two or more groups are significantly different. ANOVA tests the influence of one or more factors by comparing the means of different samples. ANOVA can be used to prove/disprove whether all drug treatments are equally effective.

Another measure used to compare samples is called the t-test. With only two samples, the t-test and ANOVA give the same result. However, using the t-test is unreliable when there are more than two samples. Running multiple t-tests to compare more than two samples has a greater effect on the error rate of the results.