

## Question 12 : Why Inferential Statistics?

Inferential statistics is a branch of statistics that involves using data from a sample to make inferences or draw conclusions about a larger population. Unlike descriptive statistics, which summarize and describe the features of a dataset, inferential statistics aim to make predictions or generalizations beyond the immediate data.

- **Explain the difference between Correlation and Causation with an example.**

Correlation indicates a statistical connection between two variables. When a correlation exists between two variables, it signifies that when one variable alters, the other variable generally changes in a particular manner. This does not necessarily mean that one variable is the cause of the change in the other. Correlations can either be positive (where both variables rise or fall together) or negative (where one variable increases while the other decreases).

Causation indicates that one event is the outcome of another event's occurrence; in other words, there exists a cause-and-effect link between the two variables. To establish causation, it takes more than merely recognizing a correlation; it necessitates showing that alterations in one variable directly lead to changes in a second variable.

### **For example:**

For the two variables "hours worked" and "income earned" there is a relationship between the two if the increase in hours worked is associated with an increase in income earned. If we consider the two variables "price" and "purchasing power", as the price of goods increases a person's ability to buy these goods decreases (assuming a constant income).

### **Also eg of difference between two relationships:**

smoking causes an increase in the risk of developing lung cancer), or it can correlate with another (e.g. smoking is correlated with alcoholism, but it does not cause alcoholism).

### Question 13: Population vs. Sample:

- **Why do we need sampling? Provide a real-world example.**

Population :The entire set of individuals or items of interest. The term population denotes the complete collection of individuals, items, or data points that are relevant to a particular research study.

Sample : Sample is a subset of the population that is selected for the actual study. It is used to make inferences about the population.

Sampling allows to :-

- Save time and resources
- Increase feasibility
- Maintain accuracy

Researchers randomly select 10 schools from the districts from total of 50 schools. Then, from each selected school, they randomly choose 50 students to participate in the study.

### Question 14 : Hypothesis Testing Concepts:

- **Define Null Hypothesis, Alternate Hypothesis, Significance Level ( $\alpha$ ), and P-value.**

#### **Null Hypothesis**

A statement that there is no effect or no difference, and it is the hypothesis that is tested.

## **Alternate Hypothesis**

A statement that there is an effect or a difference.

## **Significance Level ( $\alpha$ )**

Typically used values are 0.05 (e.g. e-commerce ) and 0.01 (e.g in fields like medicine)

## **P-value**

The probability of obtaining the observed results, or more extreme results, if the null hypothesis is true. A low p-value (typically  $< 0.05$ ) indicates that the null hypothesis can be rejected.

## **20. Summary and Insights:**

- **Summarize the key takeaways from the analysis performed above and describe how descriptive and inferential statistics can be used in real-world data analysis.**

From the analysis performed above we can conclude that Descriptive statistics helps us summarize and explain a sample, giving a clear view of the data. From this sample, inferential statistics lets us make wider conclusions or predictions about a whole population based on what we learn from the sample.

For example the use of exam scores and census data. If a teacher wants to analyze the exam scores of their students to understand the distribution of scores and identify any outliers or patterns. Using descriptive statistics they can calculate the mean, median, and standard deviation of the scores to see how well the students performed overall and how much variation there was in the scores.

A political survey taker wants to know how likely a particular candidate is to win an election. They conduct a random sample survey of voters and ask them who they plan to vote for. They use inferential statistics to estimate the candidate's level of support in the entire population of likely voters. The use of these statistical

methods, hypothesis testing, can help them estimate the probability of the candidate winning and predict the outcome of the election.