12.CORRELATION:

Correlation means there’s a relationship or pattern between two variables — when one changes, the other tends to change too. However, this doesn't mean one is causing the other to change.

CAUSATION:

Causation means that one variable directly affects another — a change in one variable causes a change in the other.

Examples:

Correlation:

Ice cream sales and drowning incidents are positively correlated — both tend to rise during summer. But eating ice cream doesn’t cause drownings. The real cause is the warm weather — people swim more and also buy more ice cream.

Causation:

If you increase the number of hours you study, your exam scores might improve. Here, the increased study time directly causes better results.

13. Why do we need sampling?

Sampling is essential because it’s often impossible, expensive, or time-consuming to collect data from an entire population. Instead of studying every single individual or item, we collect data from a subset (sample) that represents the whole population, allowing us to draw conclusions or make predictions.

Real-world example:

Imagine you're working on a project to analyze student satisfaction with the college transport system.

Population: All students in the college (say, 10,000 students).

Challenge: It’s unrealistic to survey every single student — it would take too much time and effort.

Solution: You randomly select 500 students from different years and departments — this is your sample.

Goal: Based on the sample's feedback, you estimate the overall satisfaction level for the entire student body.

14. **Null Hypothesis (H₀):**  
 The null hypothesis is a statement that there is **no effect or no difference** — it assumes that any observed result is due to chance. It represents the default or status quo position.  
Example: "There is no difference in test scores between students who studied online and those who studied in-person."

**Alternate Hypothesis (H₁ or Ha):**  
 The alternate hypothesis is the opposite of the null hypothesis — it suggests that there **is an effect or a difference**. It’s what you want to prove or support with evidence.  
Example: "Students who studied online have different test scores compared to those who studied in-person."

**Significance Level (α):**  
 The significance level, denoted by **α**, is the threshold for determining whether a result is statistically significant. It represents the probability of **rejecting the null hypothesis when it is actually true** (Type I error).  
Common values are:

* **0.05 (5%)** — meaning you are willing to accept a 5% chance of a false positive.
* **0.01 (1%)** — stricter criteria, used for more sensitive tests.

**P-value:**  
 The p-value is the **probability of obtaining results at least as extreme as the observed data, assuming the null hypothesis is true**.

* If **p ≤ α**: You reject the null hypothesis (the result is statistically significant).
* If **p > α**: You fail to reject the null hypothesis (insufficient evidence to support the alternate hypothesis).

20. **Descriptive Statistics:**  
 Used to **summarize and visualize** data, giving insights into its basic features without making predictions or conclusions.  
 **Examples:**

* Mean, median, mode, standard deviation.
* Visualizations like histograms or box plots.  
  **Real-world use:** Analyzing customer age distribution in an e-commerce platform.

**Inferential Statistics:**  
 Used to **draw conclusions or make predictions** about a population based on sample data.  
 **Examples:**

* Hypothesis testing, confidence intervals, regression analysis.  
  **Real-world use:** Testing if a new marketing strategy significantly increases sales compared to the old one.