

# Phase 1:- Statistical Foundation for Data Science

## Lec 1:- Introduction to Statistics.

\* What is statistics?

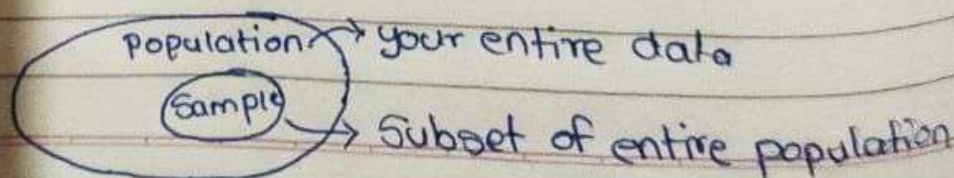
- - Statistics is a branch of mathematics that involves the collection, organizations, analysis, interpretation, & presentation of data.
- It's used in many field to understand patterns, trends, and make informed decisions.

\* ① Descriptive Statistics

- Description of the entire dataset
- (Measure of Central tendency - Mean, Median, Mode)
- (Measure of spread - SD, Variance, Inter-Quartile Range (IQR), Range)
- (Correlation coefficient)
- (Data Distribution)
- (Plots)

② Inferential Statistics

- Inferential word means conclusion
- Population vs ~~sta~~ sample
  - Population means your entire data
  - Sample means small part of your entire data (Subset of the entire population)



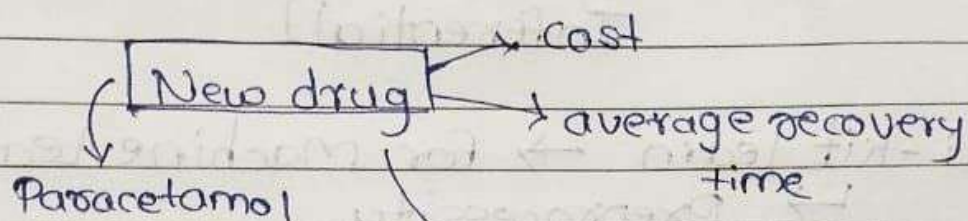


DATE  /  /  PAGE

- Statistical Testing like Hypothesis Testing, t-test, 1-test, chi-square test, ANOVA

\* Real time

→ Healthcare Domain



$H_0 \rightarrow$  new drug = paracetamol

$H_1 \rightarrow$  new drug  $\neq$  paracetamol

Reject / Accept

→ Share Market

→ Economic Budget

→ Banking & Finance

- Fraud Detection

- Descriptive Statistics & Inferential Statistic (Both)

\* Role of Statistics → Help in Decision Making

① Informed choiced

② Risk Management / Assessment

③ Performance Evaluation

④ Predictive Analysis

(Statistics is a backbone of the data-driven world)



DATE  /  /  PAGE

## \* Python Model/Framework for Statistical Mathematics :-

- SciPy → Scientific Computation in Python (Both Descriptive, Inferential)

- Sci-kit learn → for Machine Learning  
↳ Preprocessing  
↳ Data Normalization

- StatsModels

## \* Types of Data in Real World

- Qualitative Data :- Nominal & Ordinal Data  
↳ Categorical data

• - Quantitative Data :- Interval Data vs Ratio Data  
→ Interval Data :- equal interval but no true zero  
→ Ratio Data :- equal interval & a true zero point