Get the best out of Live Sessions

HOW?

e!

Check your Internet Connection

Log in 10 mins before, and check your internet connection to avoid any network issues during the LIVE session.

Speak with the Instructor

By default, you will be on mute to avoid any background noise. However, if required you will be **unmuted by instructor**.

Clear Your Doubts

Feel free to clear your doubts. Use the "**Questions**" tab on your webinar tool to interact with the instructor at any point during the class.

Let us know if you liked our content

Please share feedback after each class. It will help us to enhance your learning experience.

edureka!

edureka!

Predictive Analytics

COURSE OUTLINE
MODULE 01

1. Statistical Foundations
2. Probability
3. Inferential Statistics
4. Regression

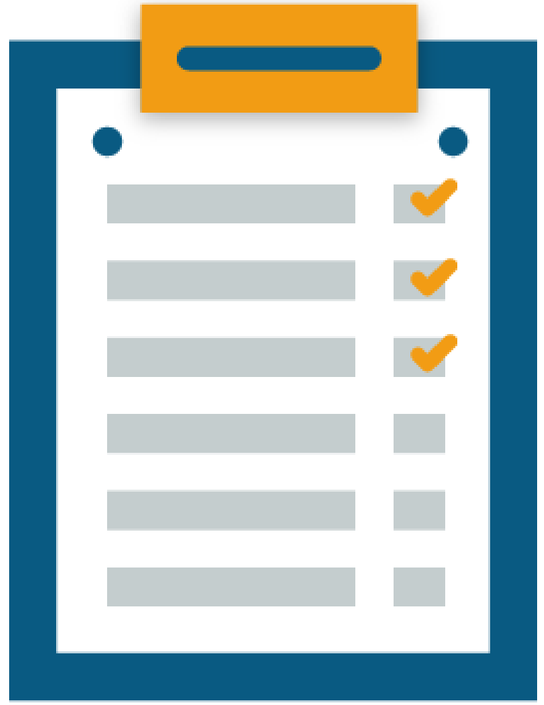
edureka!

Module 1 – Statistical Foundations Part I

Topics

Following are the topics covered in this module:

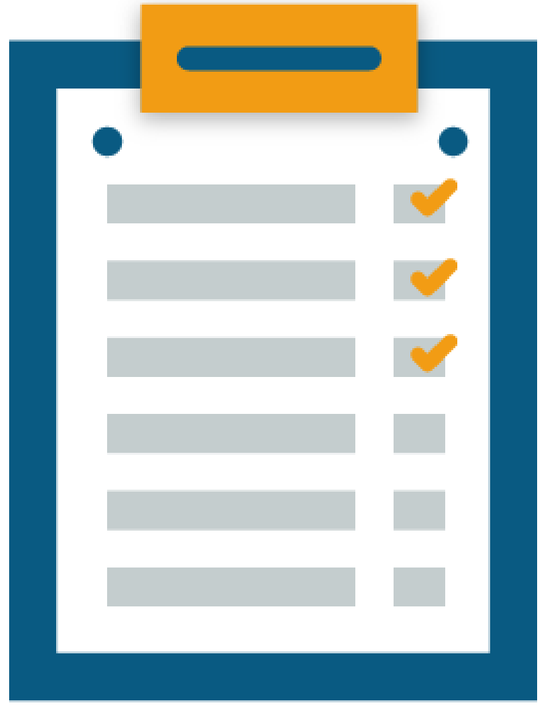
- Introduction to Statistics
  - What is Statistics?
  - Why Statistics?
  - Summary Vs. Inferential Statistics
- Measures of Central Tendency
  - Mean
  - Median
  - Mode
- Data Types
  - Categorical (Qualitative)
  - Numerical (Quantitative)
- In-class Practice Use Case:
  - Descriptive Statistics on US Superstore Dataset



Objectives

After completing this module, you should be able to:

- Examine the need for Statistical Analysis
- Summarize the given data using measures of central tendency
- Identify various types of data and learn to distinguish them



Why Statistical Analysis?

Use Case: Accenture

Statistical Analysis at Accenture

Professional Services company ***Accenture*** incurred losses in some of its business deals

- **Business Requirement:** ***Help Sales Team understand whether the deals they pursue are likely to get converted into a win or not***
- This will improve Sales Effectiveness and Resource Allocation within the company

Statistical Analysis at Accenture (Cont.)

accenture

Win Probability Predictor App by Accenture

Combined past five years of data, including customer details, sales details, and deal characteristics

Large number of variables such as details of partners, geography, etc. were analyzed using statistics and hypotheses

45000 Sales Opportunities are active in the Tool at a time, and it predicts the Win Probability with 97% accuracy

Case Study: Campus Recruitment

Consider a case study of **Riverview Engineering College** that recently concluded campus placements where students received several job offers.

What can you conclude from this Placement Data?

Check out the demo from the LMS

edureka!

Demo 1: Why Statistical Analysis is Required?

Check out the demo “Statistical Foundations I” from the LMS

edureka!

What is Statistics?

Statistics is a mathematical branch that involves **collecting**, **analyzing**, **interpreting**, and **organization** of data

Why Use Statistical Analysis?

- Statistics is used to extract useful insights from data by performing mathematical computations on it
- Math and Stats are the building blocks of Machine Learning algorithms

"Data Scientist is a person who is better at statistics than any programmer and better at programming than any statistician" – Josh Wills

ALICE's QUESTION

1. What is Statistics?

1. A technique used to graphically visualize the data
2. A technique used to summarize the available data
3. A technique used to form observations and predictions from the given data
4. All of the above

edureka!

ALICE's QUESTION

1. What is Statistics?

1. A technique used to graphically visualize the data
2. A technique used to summarize the available data
3. A technique used to form observations and predictions from the given data
4. All of the above

edureka!

ALICE's QUESTION

2. What can you deduce based on the information presented below?

Average Salary in LPA across Streams:

Commerce	Arts	Science
2.9	1.5	3

edureka!

ALICE's QUESTION

2. What can you deduce based on the information presented below?

Average Salary in LPA across Streams:

Commerce	Arts	Science
2.9	1.5	3

Students from "Science" stream get paid the maximum salary on an Average, closely followed by "Commerce" stream.

The average salary offered to "Arts" students is significantly less than the other two streams.

Statistical Analysis: Types

Descriptive	Inferential
Uses data to provide descriptions about the population	Uses data to make prediction and estimations about the population based on a sample
E.g.- Calculating mean age of people residing in New York	E.g.- Estimating mean age of population of the US from the mean age of people in New York

Why Use Descriptive Statistics?

United States Department of Agriculture (USDA) commissioned a study of adult's nutrition intake in the country on several parameters such as Calcium, Iron, Protein, Vitamin A, Vitamin C. They are interested in analyzing the overall nutritional information in the entire region. Descriptive Statistics allows us to do this

Click Here to access the Dataset

Descriptive Statistics

Descriptive Statistics is a method used to describe and understand the features of a data set by giving short summaries about the sample and measures of the data

- Deals with the presentation of numerical facts or data, in either **tabular or graphical form**

- Helps in understanding what you need to do before modeling

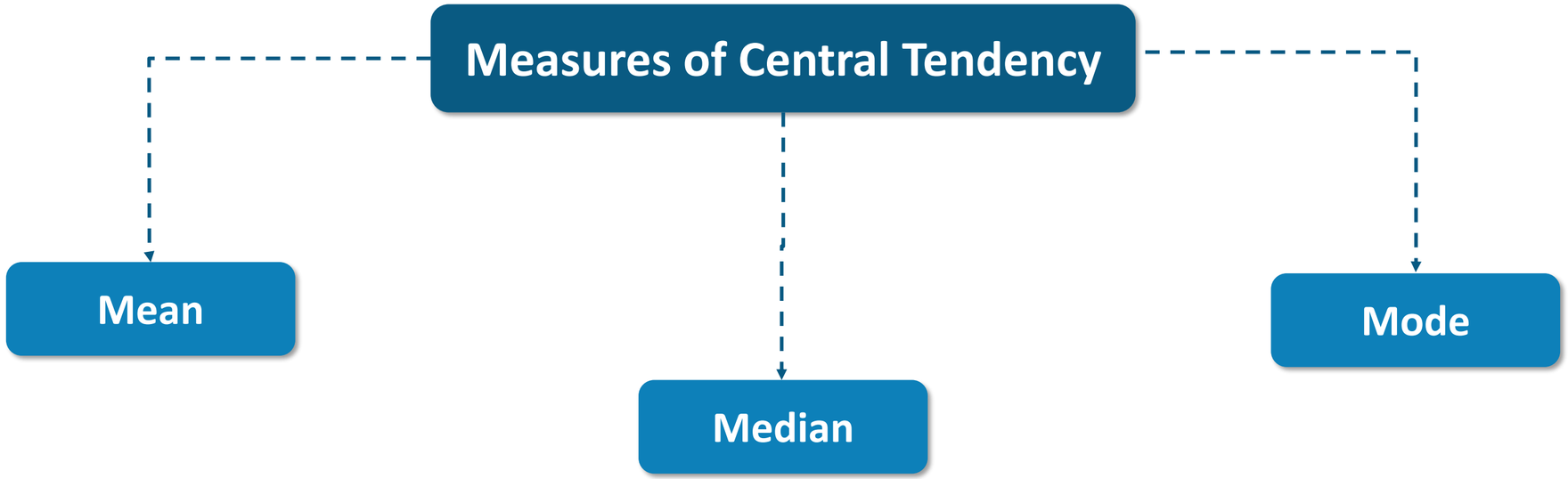
edureka!

Video: Measures of Central Tendency

edureka!

Measures of Central Tendency: Summary

- Values for most numerical variables tend to group around a specific value (generally mean or average)
- **Measures of Central Tendency** describe to what extent this pattern holds for a specific variable



edureka!

ALICE's QUESTION

3. Which of the following values provide information about central value of data?

1. Mean
2. Median
3. Mode
4. All of the above

edureka!

ALICE's QUESTION

3. Which of the following values provide information about central value of data?

1. Mean
2. Median
3. Mode
4. All of the above

edureka!

ALICE's QUESTION

4. What is the median of even numbered data points such as below:

1, 3, 5, 5, 5, 6, 7, 8, 8, 9, 10, 13

edureka!

ALICE's QUESTION

4. What is the median of even numbered data points such as below:

1, 3, 5, 5, 5, 6, 7, 8, 8, 9, 10, 13

$$\frac{6+7}{2} = 6.5$$

Arithmetic mean of two middle values

edureka!

Retail Analytics: Descriptive Statistics on US Superstore Data

edureka!

Retail Analytics

Retail Analytics is the process of providing information on inventory, sales, logistics, and customers of a retail business

- Better Marketing Decisions
- Efficient Inventory Management
- Improved Sales
- Increased Customer Satisfaction

edureka!

Descriptive Statistics on US Superstore Data

Business Scenario: A US-based Superstore was one the top retailers in the country till recent times. It has seen a decline in its revenue in the past 2 years.

Problem Statement: The company has hired you as an Analyst and wants you to perform Statistical Analysis on its data and extract some insights which might help it understand the ways of improvement.

Statistics Covered:

- Measures of Central Tendency
  - Mean
  - Median
  - Mode

Demo 2: Measures of Central Tendency

Check out the demo “Statistical Foundations I” from the LMS

edureka!

In-Class Practice I

Task 1: Create a DataFrame from the following list representing total rainfall in the month of June in Kerala (a State in India) in 9 years.

lst1 = [350, 500, 600, 700, 600, 500, 600, 300, 550]

What is the average rainfall in 9 years? (Hint: Create a DataFrame from the list and calculate mean)

Task 2: Data of another year is now available. Recalculate the mean rainfall for 10 years.

lst2 = [350, 500, 600, 700, 600, 500, 600, 300, 550, 1400]

Task 3: Calculate the median for data of 9 years and 10 years separately and compare the results.

Task 4: Calculate the most common amount of rainfall. (Hint: Calculate the mode separately for 9 and 10 years and compare.)

Additional Task: View the distribution of data using a Histogram.

Data Types

edureka!

Video: Data Types

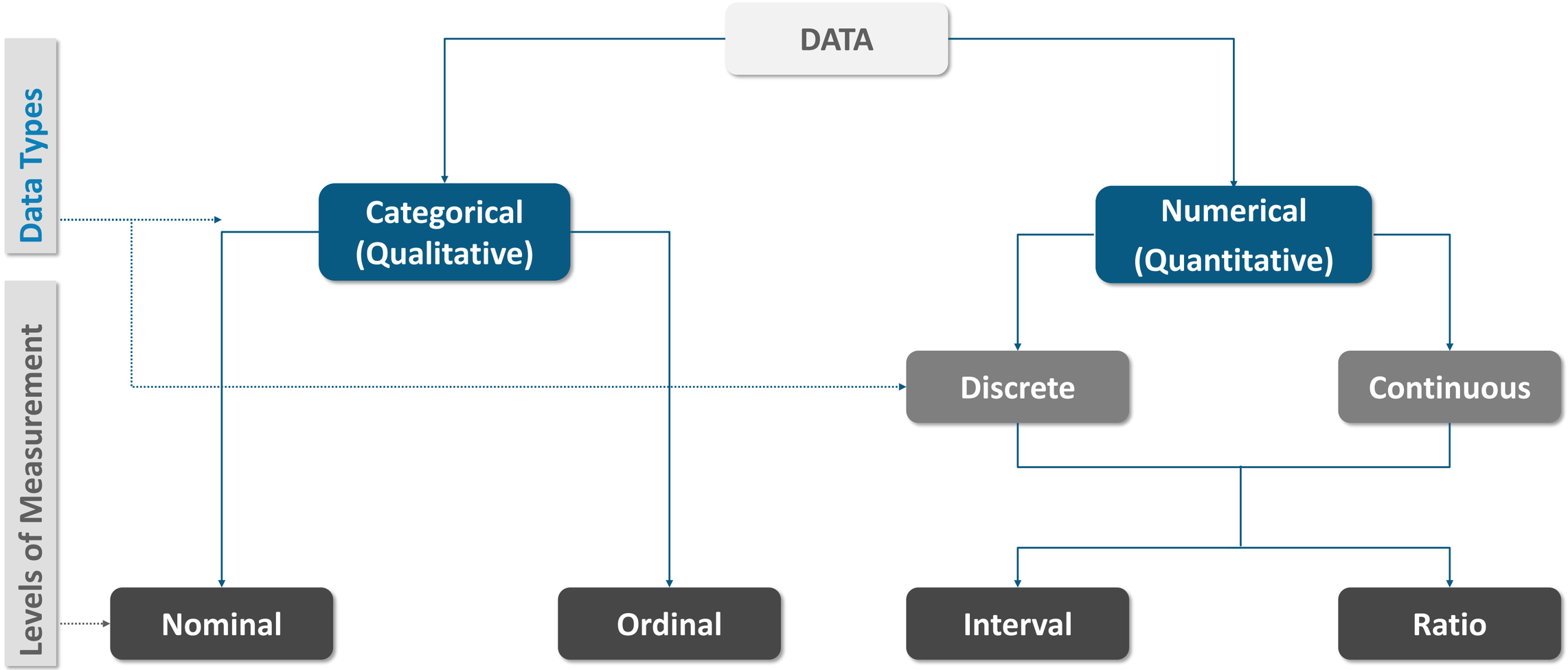
edureka!

Demo 3: Data Types

Check out the demo “Statistical Foundations I” from the LMS

edureka!

Data Types: Summary



edureka!

Mark's QUESTION

5. Which of the following columns of the US Superstore data set are Numerical?

1. Segment
2. Ship-mode
3. Sales
4. Quantity
5. Discount
6. Profit

edureka!

Mark's QUESTION

5. Which of the following columns of the US Superstore data set are Numerical?

1. Segment
2. Ship-mode
3. Sales
4. Quantity
5. Discount
6. Profit

edureka!

ALICE's QUESTION

6. An Investment Company wants to analyze fluctuations in stock prices of various stocks. These stocks belong to different industry sectors.

Is ***Stock Price*** a Discrete variable or Continuous?

edureka!

ALICE's QUESTION

6. An Investment Company wants to analyze fluctuations in stock prices of various stocks. These stocks belong to different industry sectors.

Is ***Stock Price*** a Discrete variable or Continuous?

Stock Prices may have any positive real number value.

edureka!

Mark's QUESTION

7. An Investment Company wants to analyze fluctuations in stock prices of various stocks. These stocks belong to different industry sectors.

What is the type of ***Industry Sector*** variable?

1. Nominal
2. Ordinal
3. Discrete
4. Continuous

edureka!

Mark's QUESTION

7. An Investment Company wants to analyze fluctuations in stock prices of various stocks. These stocks belong to different industry sectors.

What is the type of ***Industry Sector*** variable?

1. Nominal
2. Ordinal
3. Discrete
4. Continuous

Industry Sectors consist of different categorical values which do not have a specific order

edureka!

In-Class Practice II

Task: Calculate the Measures of Central Tendency for the *Quantity* column

(Use DataFrame Functions).

edureka!

edureka!

edureka!

Thank You

For more information please visit our website
www.edureka.co