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

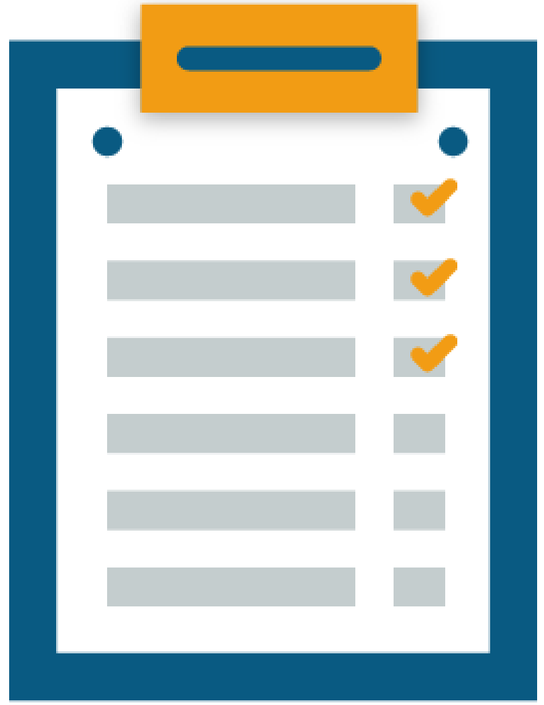
edureka!

Module 1 – Statistical Foundations Part 2

Topics

Following are the topics covered in this module:

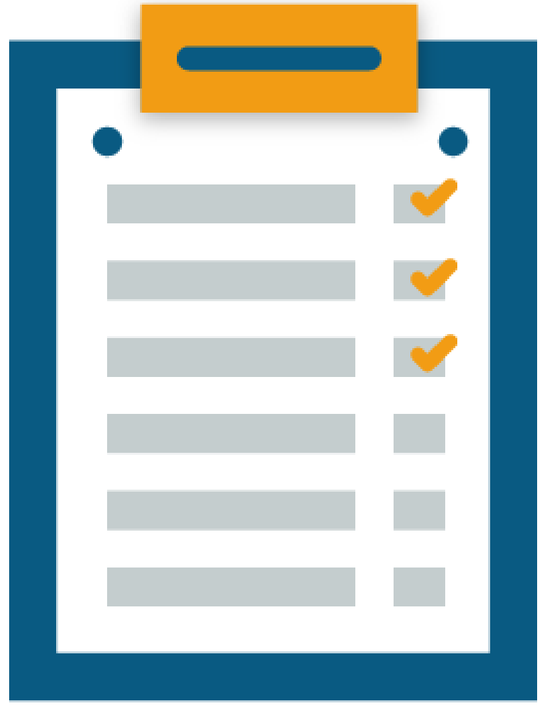
- Measures of spread
  - Why measure spread
  - Variance
  - Standard Deviation
  - Range
  - IQR
  - Outliers
- Sampling
  - What is Sampling
  - Types of Sampling
- In-class Practice Use Case:
  - Measure of variance on real world data
  - Detecting outliers
  - Obtaining sample from a larger data



Objectives

After completing this module, you should be able to:

- Examine the need for the measure of variance
- Understand the various measures for variance
- Apply the learnings to understand the data for a given scenario

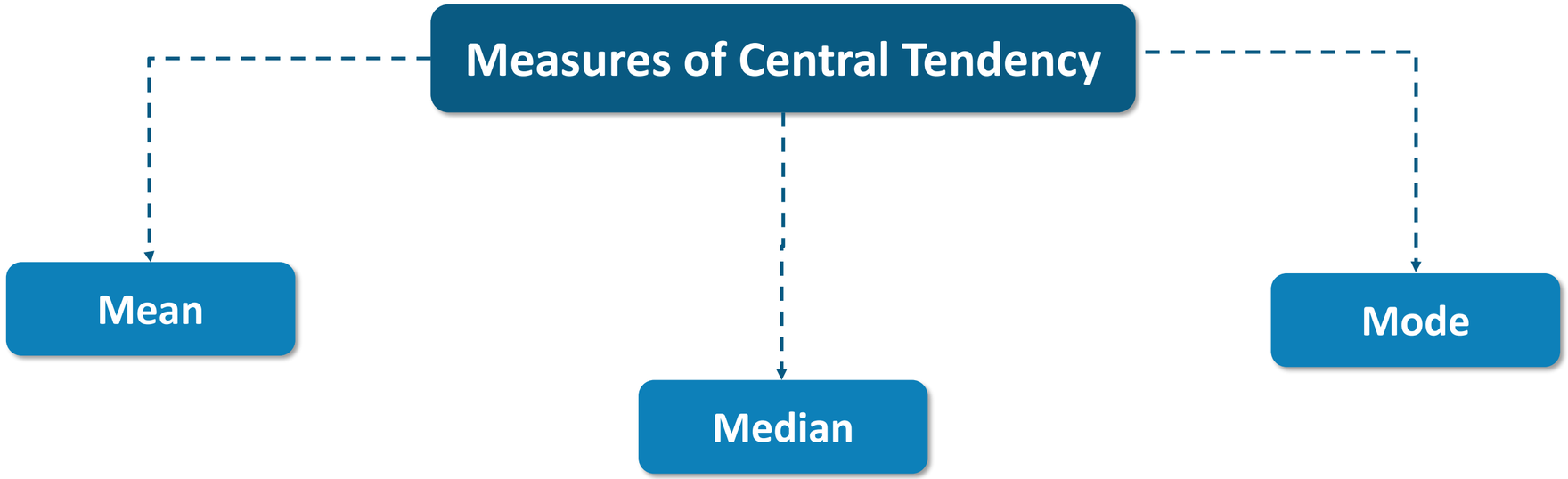


Recap of Part 1 from Day 1

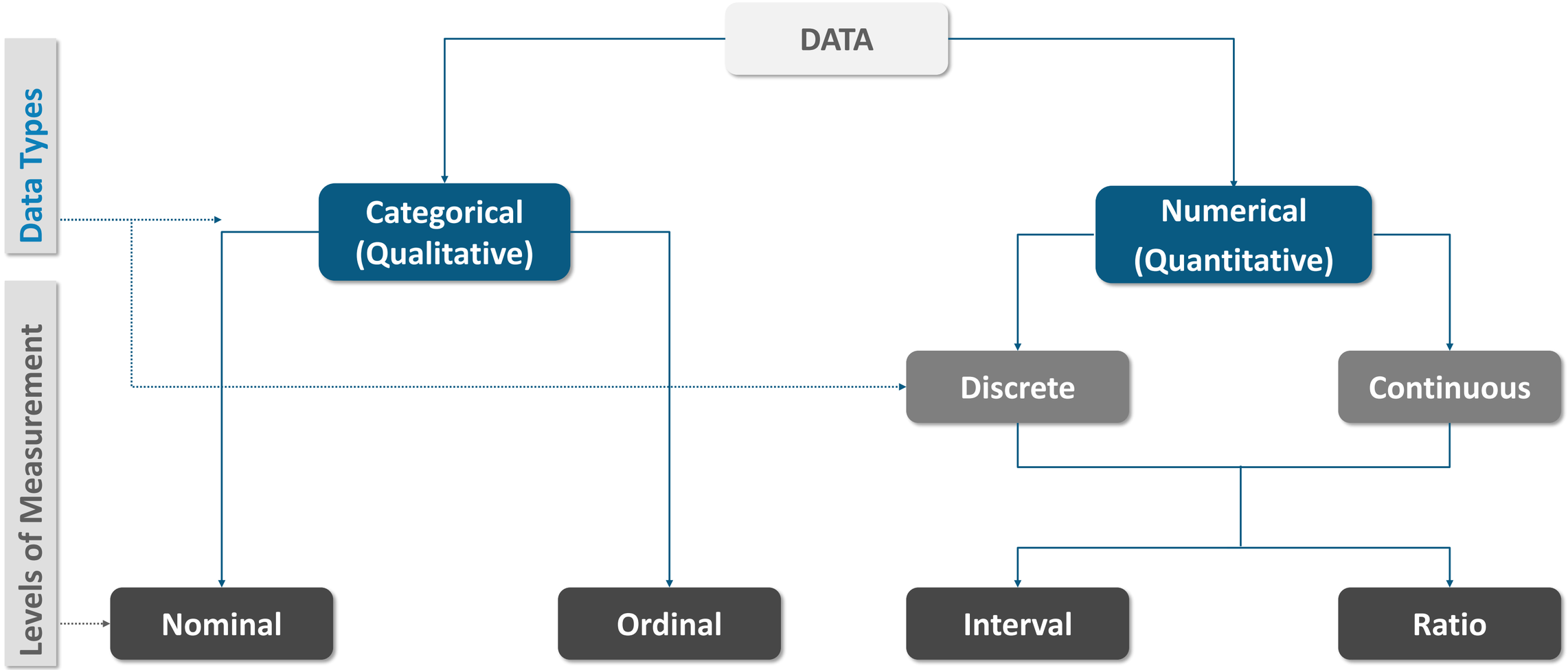
- Measures of Central Tendency
- Different Types of Data

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



Data Types: Summary

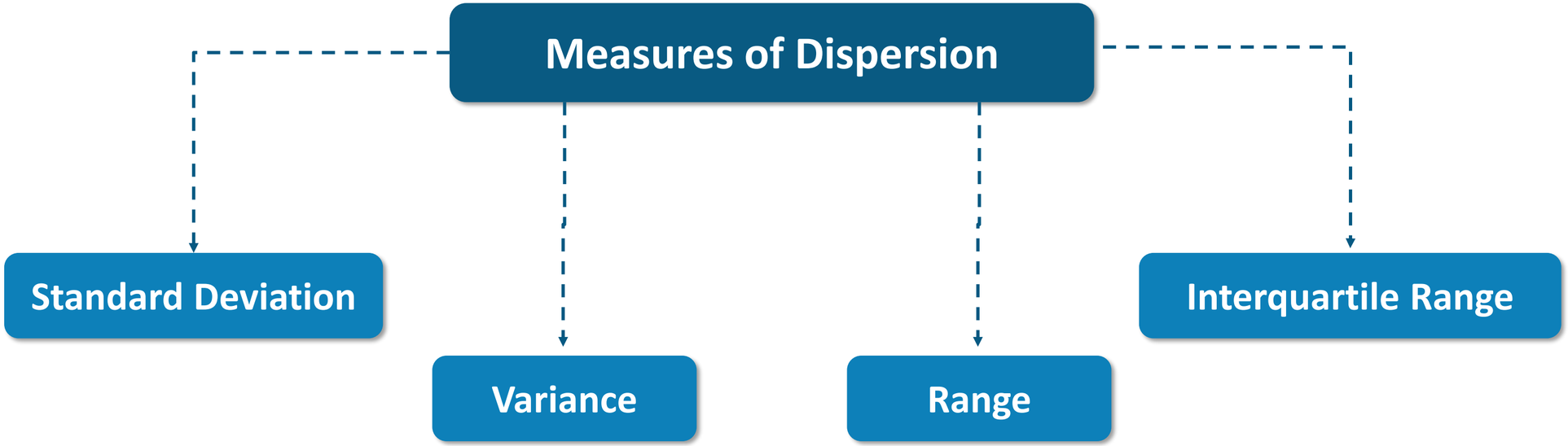


edureka!

Video: Measures of Spread

edureka!

Measures of Spread – Summary



ALICE’s QUESTION

1. What are the different measures of spread?

1. Range
2. Variance
3. Standard Deviation
4. All of the above

edureka!

ALICE's QUESTION

1. What are the different measures of spread?

1. Range
2. Variance
3. Standard Deviation
4. All of the above

edureka!

ALICE's QUESTION

2. Which of the following measures help in identifying outliers?

1. Range
2. Variance
3. Standard Deviation
4. IQR

edureka!

ALICE's QUESTION

2. Which of the following measures help in identifying outliers?

1. Range
2. Variance
3. Standard Deviation
4. IQR

edureka!

ALICE's QUESTION

3. Presence of outliers affect which of the following?

1. Mean
2. Median
3. Mode
4. Variance

edureka!

ALICE's QUESTION

3. Presence of outliers affect which of the following?

1. Mean
2. Median
3. Mode
4. Variance

edureka!

Demo 1: Measures of Spread

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

edureka!

In-Class Practice I

Task 1: Find the presence of outliers (if any) on the salary column from “Placement_Data_Full_Class.csv” using the box plots. Print the outlier values.

Task 2: Check if there are outliers in the Profit column.

edureka!

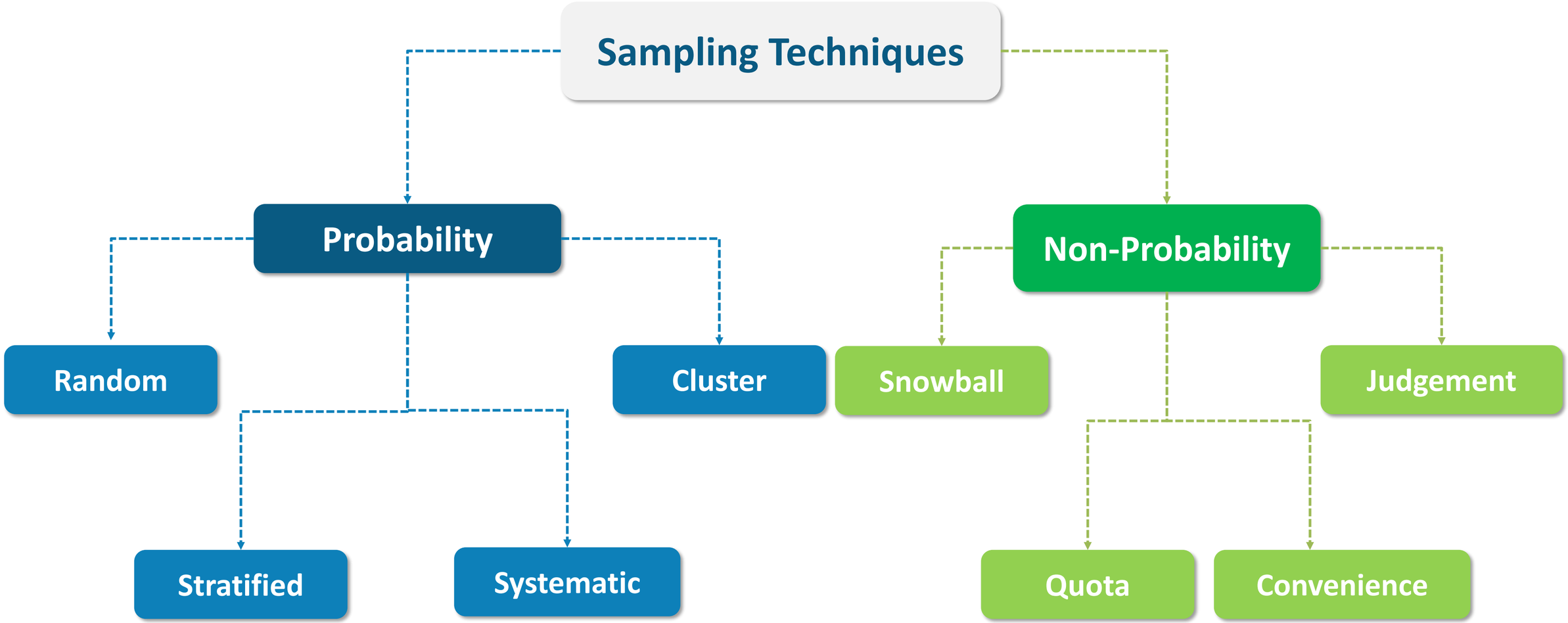
Sampling

edureka!

Video: Sampling

edureka!

Different Types of Sampling Techniques – Summary



edureka!

ALICE's QUESTION

4. Which of the following belong to Random sampling?

1. Simple
2. Stratified
3. Systematic
4. All of the above

edureka!

ALICE's QUESTION

4. Which of the following belong to Random sampling?

1. Simple
2. Stratified
3. Systematic
4. All of the above

edureka!

ALICE's QUESTION

5. Why is a sample preferred over directly analysing the population?

1. Less time consuming
2. Cost effective
3. Works well with limited resources
4. All of the above

edureka!

ALICE’s QUESTION

5. Why is a sample preferred over directly analysing the population?

1. Less time consuming
2. Cost effective
3. Works well with limited resources
4. All of the above

edureka!

In-Class Practice II

Task: Obtain a stratified random sampling on the US Superstore data using the Segment column, i.e each segment can be considered as a strata

Hint: Create multiple DataFrames by filtering based on each segment. Obtain a simple random sample from each of the DataFrames and create a new DataFrame by appending the results obtained.

edureka!

edureka!

edureka!

Thank You

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