SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

Department of Computer Science and Engineering

Course Code & Title : 23ADO201 - Data Science Fundamentals

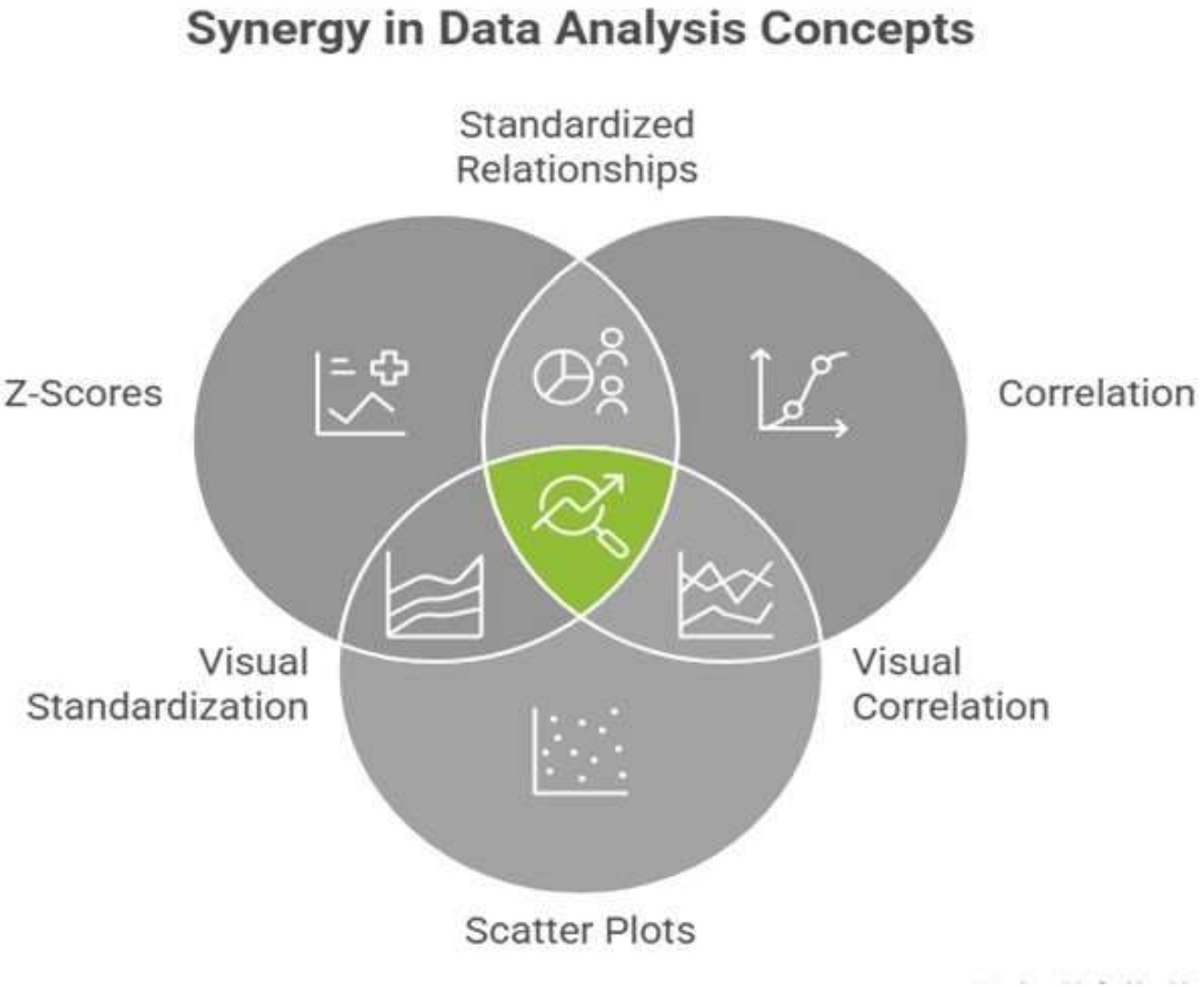
III YEAR / VI SEMESTER - EEE

Unit 2 - DESCRIPTIVE ANALYTICS

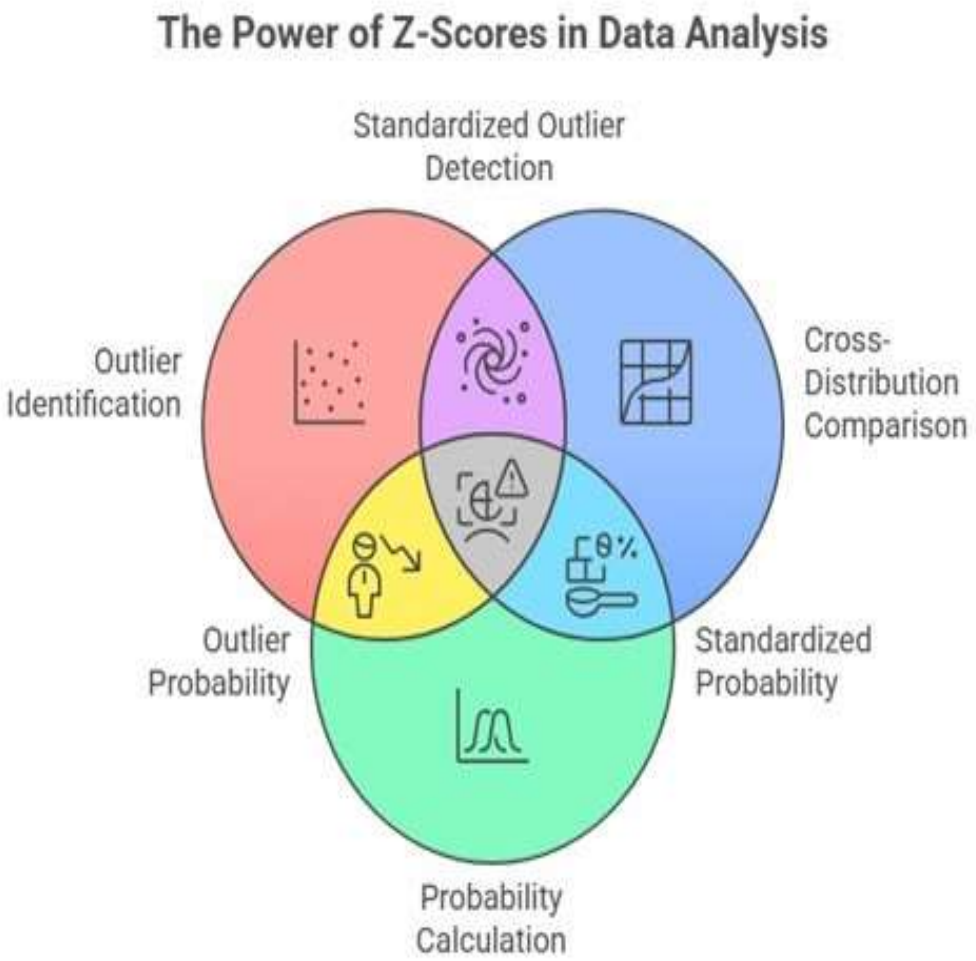
Topic : z scores , correlation and scatter plots

Mr.K.Karthikeyan AP /CSE

Empathy


Synergy in Data Analysis Concepts
Standardized Relationships
Z-Scores
Correlation
Visual Standardization
Visual Correlation
Scatter Plots

A **z-score**, also known as a standard score, indicates how many standard deviations an element is from the mean. A z-score can be calculated for any normally distributed dataset.



Formula for Z-Score

The z-score is calculated using the following formula:

$z = (x - \mu) / \sigma$

Where:

- **z is the z-score**
- **x is the observed value**
- **μ is the population mean**
- **σ is the population standard deviation**

If you have a sample instead of the entire population, you would use the sample mean ($\bar{x}$) and sample standard deviation (s) in the formula:

$z = (x - \bar{x}) / s$

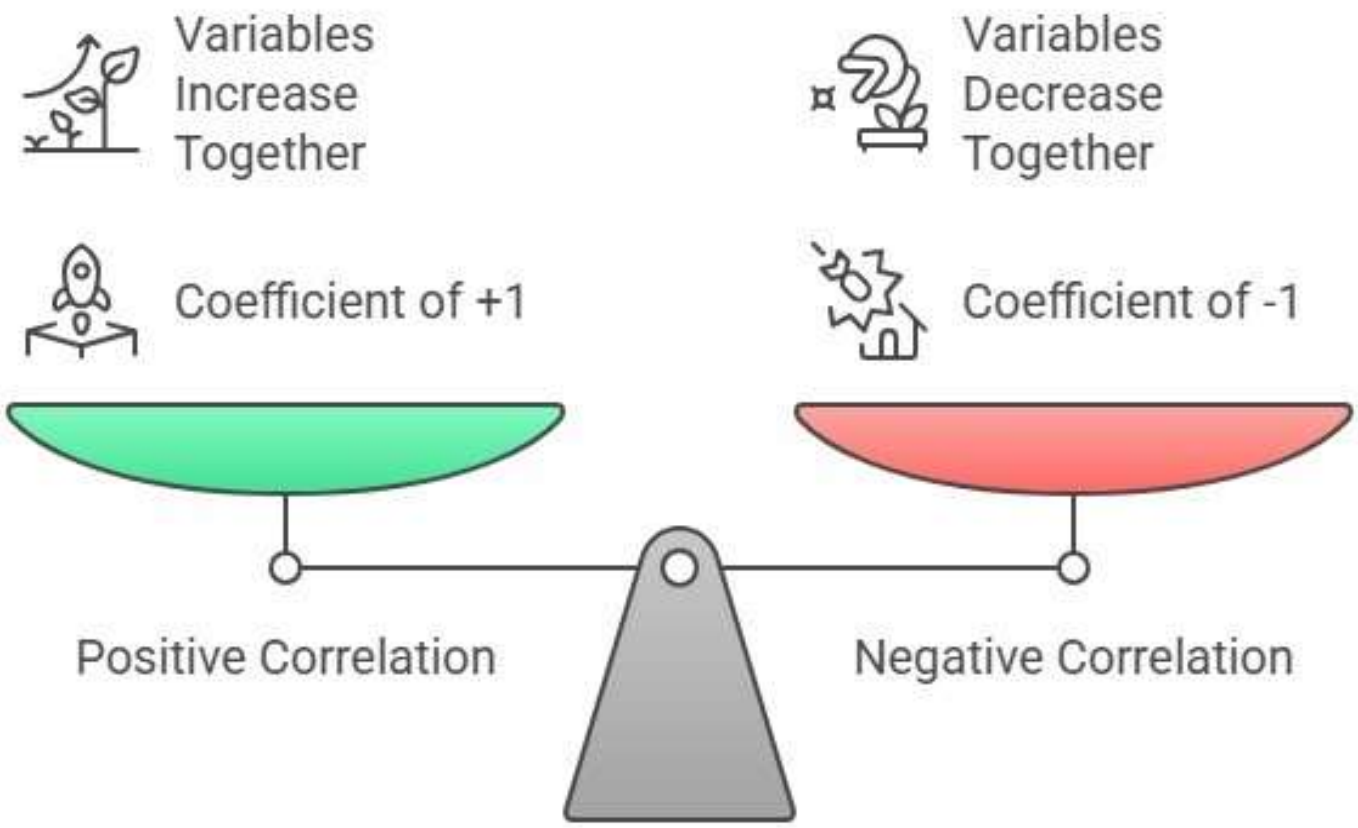
Example

Suppose we have a dataset of test scores with a mean of 75 and a standard deviation of 10. A student scored 85 on the test. What is the student's z-score?

$$z = (85 - 75) / 10 = 1$$

This means the student's score is 1 standard deviation above the mean.

Understanding Correlation Strength and Direction



Pearson Correlation Coefficient

The most common type of correlation is the Pearson correlation coefficient, which measures the linear relationship between two continuous variables. The formula for the Pearson correlation coefficient is:

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2]}$$

Where:

- r is the Pearson correlation coefficient
- x_i and y_i are the individual data points for variables x and y
- $\bar{x}$ and $\bar{y}$ are the sample means of variables x and y
- Σ denotes summation

Interpretation of Correlation Strength

While the exact interpretation can depend on the context here's a general guideline:

- **$|r| > 0.7$:** Strong correlation
- **$0.5 < |r| \leq 0.7$:** Moderate correlation
- **$0.3 < |r| \leq 0.5$:** Weak correlation
- **$|r| \leq 0.3$:** Very weak or no correlation

🎯 B. CORRELATION – ACTIVITIES

3. Correlation Guess 🤔

Time: 10 minutes

Mode: Whole class

Steps:

Show variable pairs:

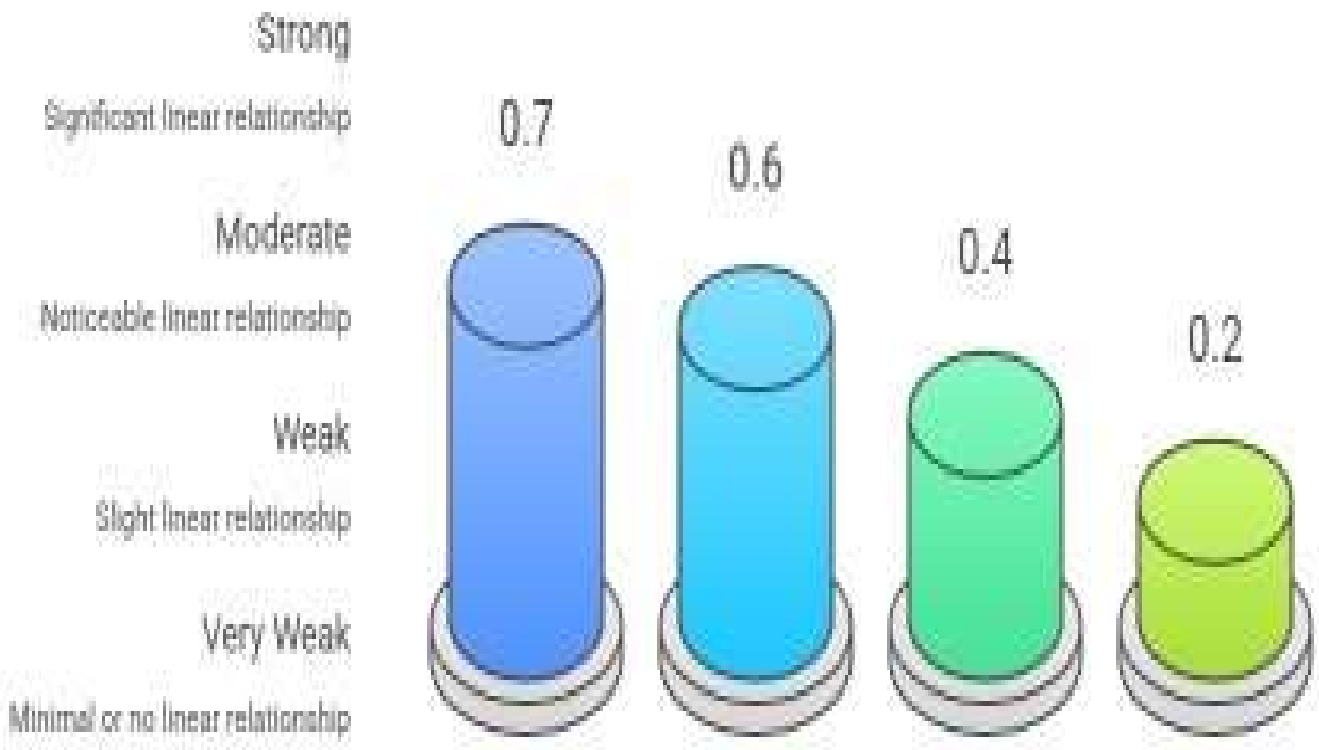
- Height vs Weight
- Study Hours vs Marks
- Shoe Size vs Intelligence 😄

Students predict:

- Positive correlation
- Negative correlation
- No correlation

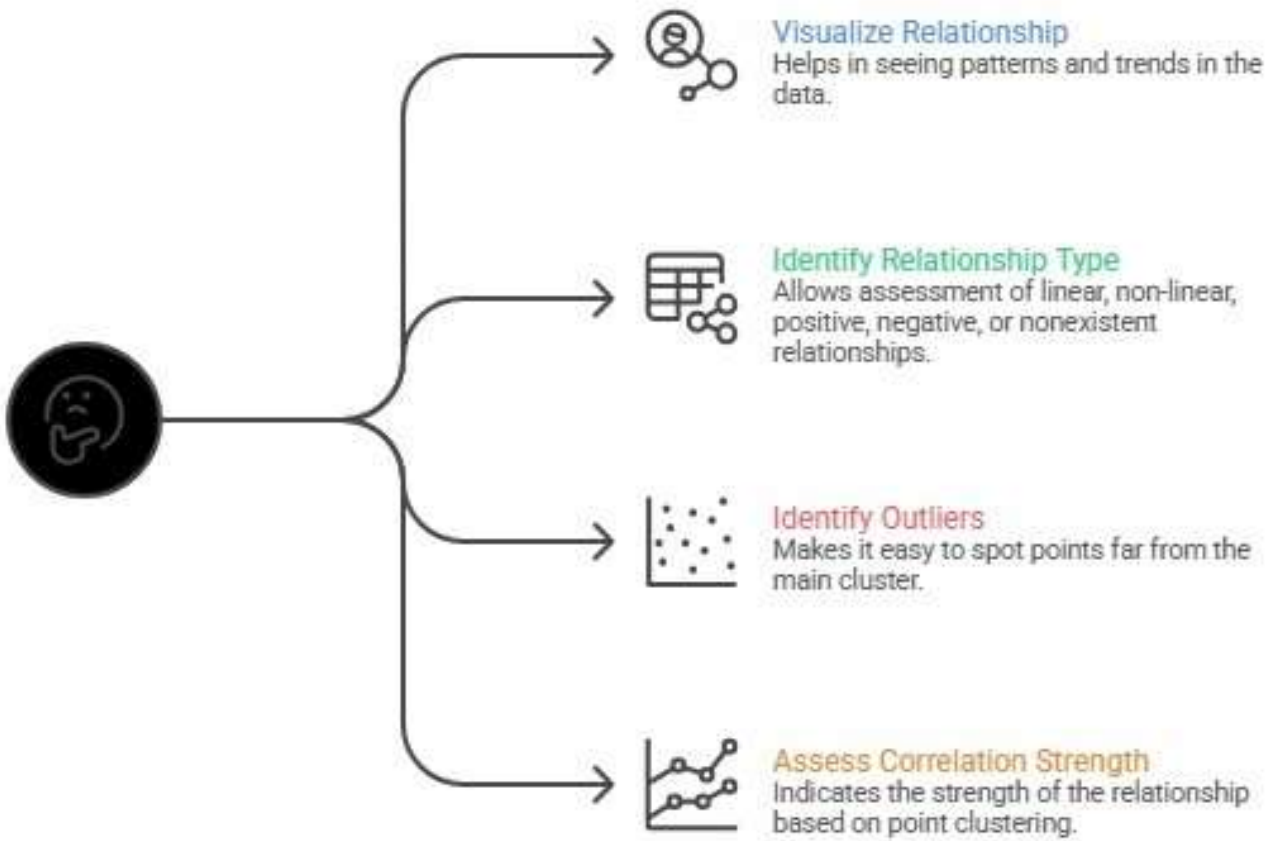
✅ **Learning Outcome:** Understanding relationship direction

Interpretation of Pearson Correlation Coefficient Strength

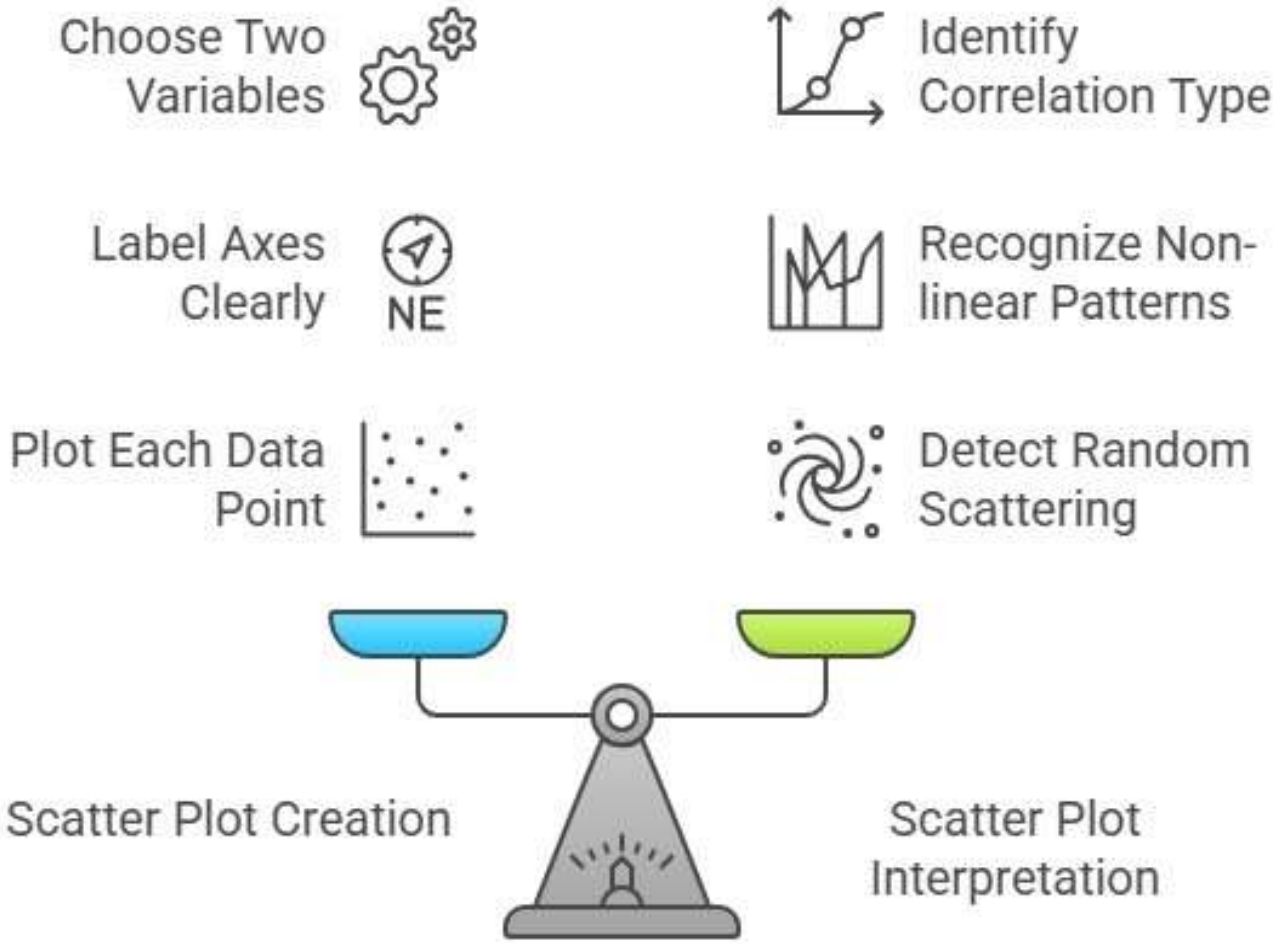


www.snsgroups.com

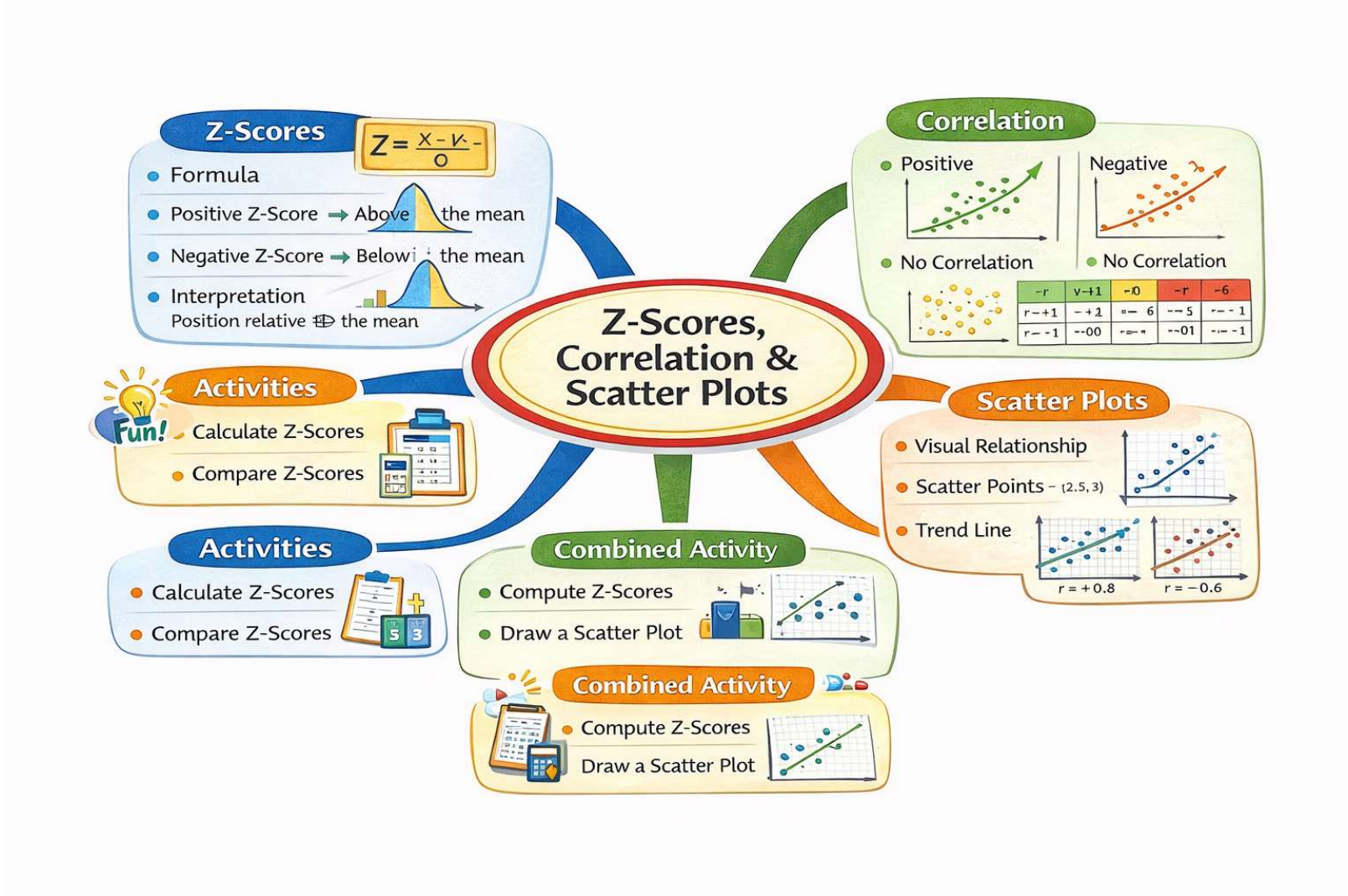
How to use scatter plots for data analysis?



Creating and Interpreting Scatter Plots



MIND MAP


Z-Scores, Correlation & Scatter Plots
Z-Scores
Formula
Positive Z-Score → Above the mean
Negative Z-Score → Below the mean
Interpretation
Position relative to the mean
Correlation
Positive
Negative
No Correlation
No Correlation
Activities
Fun!
Calculate Z-Scores
Compare Z-Scores
Scatter Plots
Visual Relationship
Scatter Points – (2.5, 3)
Trend Line
r = +0.8
r = −0.6
Activities
Calculate Z-Scores
Compare Z-Scores
Combined Activity
Compute Z-Scores
Draw a Scatter Plot
Combined Activity
Compute Z-Scores
Draw a Scatter Plot

ASSESSMENT

www.snsgroups.com

📊 Z-SCORES, CORRELATION & SCATTER PLOTS – MCQs

🔢 A. Z-SCORES

1. A Z-score represents:

A. Raw value

B. Position of a value relative to the mean

C. Total frequency

D. Variance

✅ **Answer:** B

2. The formula for Z-score is:

A. $(X - \sigma) / \mu$

B. $(X - \mu) / \sigma$

C. $(\mu - X) / \sigma$

D. X / σ

✅ **Answer:** B

REFERENCE BOOKS

1.Allen B. Downey, “Think Stats: Exploratory Data Analysis in Python”, Green Tea Press, 2014.
2.Sanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, “Fundamentals of Data Science”, CRC Press, 2022.
3.Chirag Shah, “A Hands-On Introduction to Data Science”, Cambridge University Press, 2020.
4.Vineet Raina, Srinath Krishnamurthy, “Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice”, A press, 2021.

www.snsgroups.com