

# Statistical Thinking Workshop Day 1

Nalanda Academy, Wardha

---

## Exploring Social Media Addiction and Its Impacts

These exercises are designed to introduce core statistical concepts through meaningful real-world data. By analyzing a dataset on social media usage, addiction, academic performance, and personal relationships, participants will develop hands-on skills in descriptive statistics, data visualization, correlation analysis, and basic hypothesis testing. The aim is to foster statistical thinking that can be applied across diverse social science and behavioral research scenarios.

### About the Dataset

#### Source:

Social Media Addiction vs Relationships – [Kaggle Dataset](#)

The dataset includes survey responses from students across different countries, covering demographic details, social media habits, perceived impacts on academics, mental health scores, and relationship status.

### Dataset Fields Overview

- ***Student\_ID:*** Unique identifier for each respondent
- ***Age:*** Respondent's age in years
- ***Gender:*** Respondent's gender
- ***Academic\_Level:*** Current educational stage (High School, Undergraduate, Graduate)
- ***Country:*** Country of residence
- ***Avg\_Daily\_Usage\_Hours:*** Average daily hours spent on social media
- ***Most\_Used\_Platform:*** Primary social media platform used (e.g., Instagram, TikTok)
- ***Affects\_Academic\_Performance:*** Whether social media is perceived to affect academic work (Yes/No)
- ***Sleep\_Hours\_Per\_Night:*** Average sleep duration per night in hours
- ***Mental\_Health\_Score:*** Self-rated mental health on a scale (likely 1-10)
- ***Relationship\_Status:*** Relationship status (Single, In Relationship, Complicated)

- **Conflicts\_Over\_Social\_Media:** Number of conflicts arising due to social media use
  - **Addicted\_Score:** Self-reported addiction level to social media (scale not specified)
- 

## Practice Exercises

### Section 1: Descriptive Statistics

*Goal:* Learn to summarize data.

1. Calculate mean, median, mode, standard deviation, and range for the following numeric variables:
    - Avg\_Daily\_Usage\_Hours
    - Sleep\_Hours\_Per\_Night
    - Mental\_Health\_Score
    - Addicted\_Score
  2. Create frequency tables for categorical variables:
    - Gender
    - Most\_Used\_Platform
    - Relationship\_Status
- 

### Section 2: Data Visualization

*Goal:* Understand distributions and relationships.

3. Create histograms for:
  - Addicted\_Score
  - Sleep\_Hours\_Per\_Night
4. Create bar charts showing counts of:
  - Most\_Used\_Platform

5. Create a pie chart for:
    - Relationship\_Status
- 

### Section 3: Correlation and Scatter Plots

*Goal:* Study relationships between variables.

6. Use scatter plots to analyze relationships between:
    - Avg\_Daily\_Usage\_Hours vs. Mental\_Health\_Score
    - Addicted\_Score vs. Sleep\_Hours\_Per\_Night
    - Conflicts\_Over\_Social\_Media vs. Addicted\_Score
  7. Add a trendline to each scatter plot showing the equation and the  $R^2$  value.
  8. Compute the Pearson correlation coefficient for each pair using Excel's `=CORREL(array1, array2)` function.
  9. Interpret the meaning of the trendline,  $R^2$ , and correlation coefficients for each pair.
- 

### Section 4: Pivot Tables for Group Analysis

*Goal:* Learn conditional statistics and group comparisons.

10. Create a pivot table to compare average Addicted\_Score by:
  - Gender
  - Country
11. Create a pivot table to compare average Mental\_Health\_Score across different Academic\_Level groups.

### Tools Required

- Microsoft Excel or Google Sheets (for data analysis and visualization)