Data Representation and View

Environmental Statistics

م.م. دعاء زياد الكاتب

Lecture 4: Data Representation and View

1. Introduction

Data representation and view involve organizing and displaying data in a meaningful way to make it easier to understand, analyze, and communicate. This includes **tabular** and **graphical** methods.

2. Frequency Tables

A **frequency table** is a structured way to summarize how often each value or group of values occurs in a dataset.

Example: Exam Scores of 15 Students

Data:

```
55, 60, 62, 65, 65, 67, 70, 72, 72, 72, 75, 78, 80, 82, 85
```

Step 1: Create Class Intervals

Class Interval	Frequency (f)
50–59	1
60–69	5
70–79	6
80–89	3

Step 2: Add More Columns

Class Interval	f	Relative f	Cumulative f
50–59	1	1/15 ≈ 0.07	1
60–69	5	0.33	6
70–79	6	0.40	12
80–89	3	0.20	15

Where **Relative f** is the **proportion** or **percentage** of the total number of data points that falls into a particular category or interval.

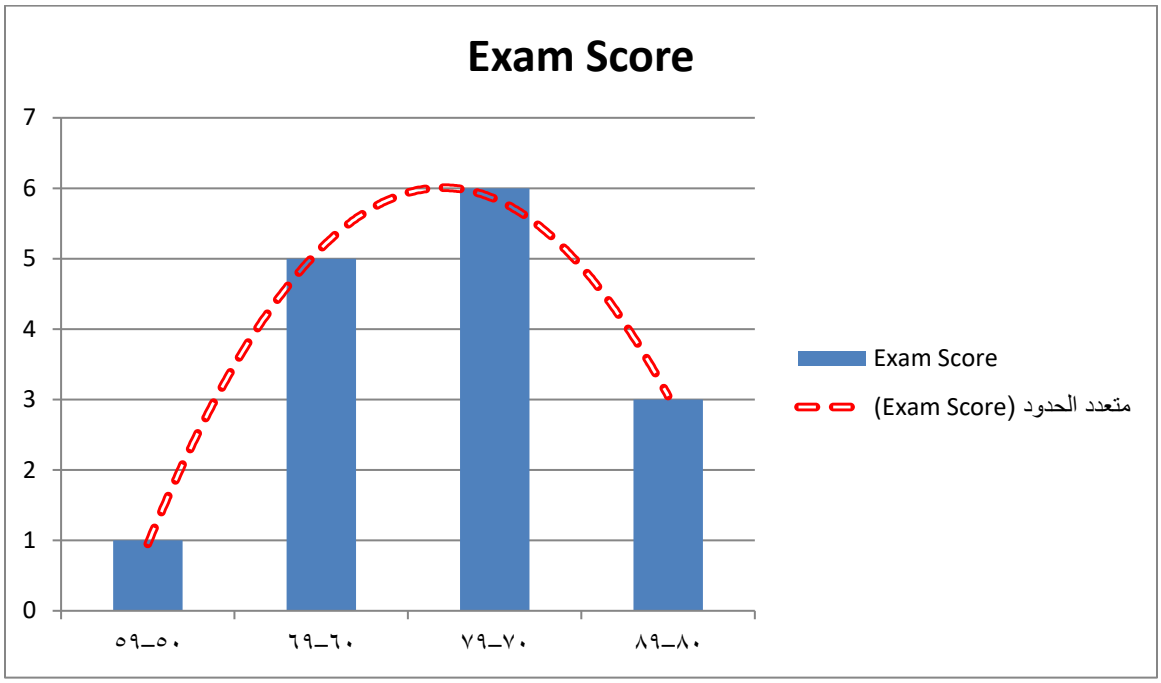
$$Relative\ F = \frac{f}{F_{total}}$$

And **Cumulative f is running total** of frequencies up to the end of each class or category. It tells how many data points are **less than or equal to** the upper boundary of each class.

$$\text{Cumulative Frequency (cf)} = \sum_{i=1}^{k} f_i$$

3. Bar Charts

Bar charts are used for **categorical data** (e.g., types, names, categories). Each bar represents a category, and the height shows the frequency or percentage.

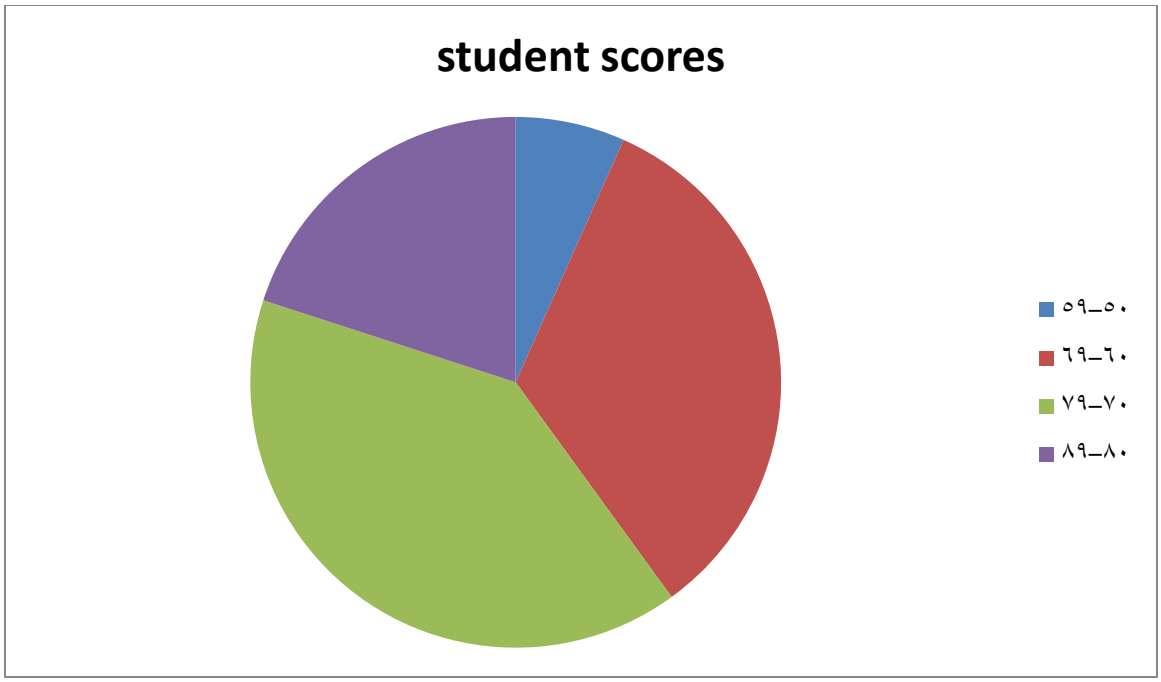


4. Pie Charts

Pie charts show **part-to-whole relationships** as slices of a circle.

Calculation:

$$\text{Angle} = \left(\frac{\text{Frequency}}{\text{Total}}\right) \times 360^\circ$$



Histograms

A **histogram** is a graphical representation of the **distribution of numerical data**. It displays data using **adjacent rectangles (bars)** where the height of each bar represents the **frequency** (or relative frequency) of data within a specific **interval (bin)**.

Construction Steps

Step 1: Collect and Sort the Data

Example dataset (test scores for 20 students):
`45, 52, 53, 55, 56, 60, 61, 62, 63, 65, 66, 68, 70, 72, 75, 77, 78, 80, 83, 85`

Step 2: Find Range

$$\text{Range} = \text{Max} - \text{Min} = 85 - 45 = 40$$

Step 3: Decide the Number of Bins

A common rule is **the following formula**:

$$k = 1 + 3.322 \log_{10}(n)$$

$$k = 1 + 3.322 \log_{10}(20) \approx 1 + 3.322 \times 1.301 \approx 5.32 \approx 5 \text{ bins}$$

Step 4: Determine Bin Width

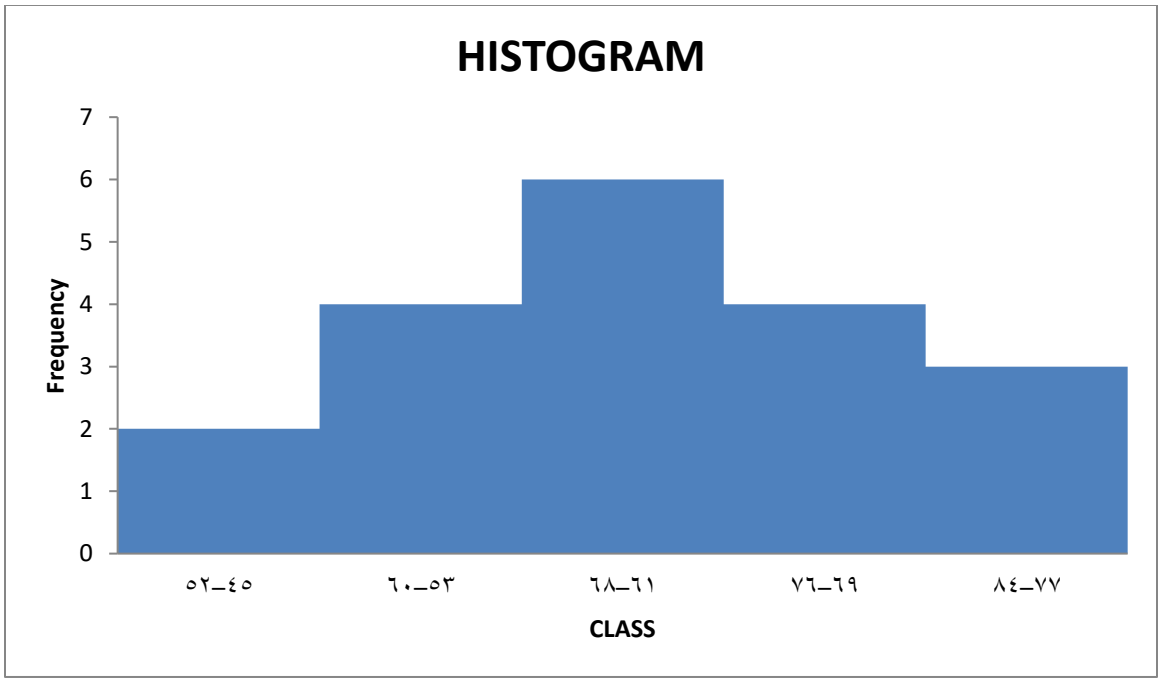
$$\text{Bin Width} = \frac{\text{Range}}{k} = \frac{40}{5} = 8$$

Step 5: Create Class Intervals

Class Interval	Frequency (f)
45–52	2
53–60	4
61–68	6
69–76	4
77–84	3
85–92	1

Step 6: Draw Histogram

- X-axis: 45 to 92 (class intervals)
- Y-axis: Frequency
- Bars:
    - Width = 8 (constant)
    - Height = frequency in each class



Types of Frequency in Histograms:

a. Simple Frequency Histogram

Each bar's height = count (f) in the bin.

b. Relative Frequency Histogram

Each bar's height = relative frequency:

c. Density Histogram

Height = frequency / (total data × bin width) — used when bin widths are **unequal**.

Interpretation

Histograms can reveal:

- **Shape of the distribution:**
  - **Symmetric:** Bell-shaped (normal distribution)
  - **Skewed Right (Positive):** Tail on the right
  - **Skewed Left (Negative):** Tail on the left
  - **Uniform:** All bars are similar in height
  - **Bimodal/Multimodal:** Two or more peaks
- **Central Tendency:** Where most data cluster
- **Spread:** How spread out data is
- **Outliers:** Unusual gaps or isolated bars