MATH 108: Elementary Probability and Statistics

Ramapo College of New Jersey

Instructor: Dr. Atul Anurag

Semester: Fall 2025 Date: September 11, 2025

Bar Plots, Histograms, and Types of Distributions

Frequency and Relative Frequency Distributions

Definition

A **frequency distribution** is a table that shows how often each value or category appears in a dataset.

Example 1

Suppose we surveyed 20 students about the number of books they read last month:

$$2, 3, 3, 4, 2, 5, 3, 4, 4, 5, 2, 1, 3, 4, 2, 3, 5, 4, 2, 3$$

Steps

1. **List all unique values**: 1, 2, 3, 4, 5
2. **Count the frequency of each value**

Number of Books	Frequency
1	1
2	5
3	6
4	5
5	3

Relative Frequency Distribution

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total Observations}}$$

Number of Books	Frequency	Relative Frequency	Percentage (%)
1	1	0.05	5%
2	5	0.25	25%
3	6	0.30	30%
4	5	0.25	25%
5	3	0.15	15%

Example 2: Favorite Fruit

Responses: Apple, Banana, Apple, Orange, Banana, Apple, Apple, Banana, Grape, Apple, Orange, Banana, Banana, Grape, Apple, Orange, Banana, Apple, Apple, Grape

Fruit	Frequency	Relative Frequency
Apple	8	0.40
Banana	6	0.30
Orange	3	0.15
Grape	3	0.15

Discrete vs. Continuous Variables

Discrete Variables

Quantitative and countable (e.g., number of students, cars, etc.)

Continuous Variables

Quantitative and uncountable (e.g., height, time, temperature)

Examples

Discrete Data Example

Score	Frequency	Relative Frequency
78	2	0.13
85	3	0.20
90	4	0.27
92	2	0.13
95	2	0.13
100	2	0.13

Continuous Data Example

Height Range	Frequency	Relative Frequency
150–159	2	0.10
160–169	5	0.25
170–179	8	0.40
180–189	4	0.20
190–199	1	0.05

Graphical Displays

Bar Graph (Categorical)

- X-axis: Categories
- Y-axis: Frequency or Relative Frequency
- Bars do not touch

Histogram

- X-axis: Intervals

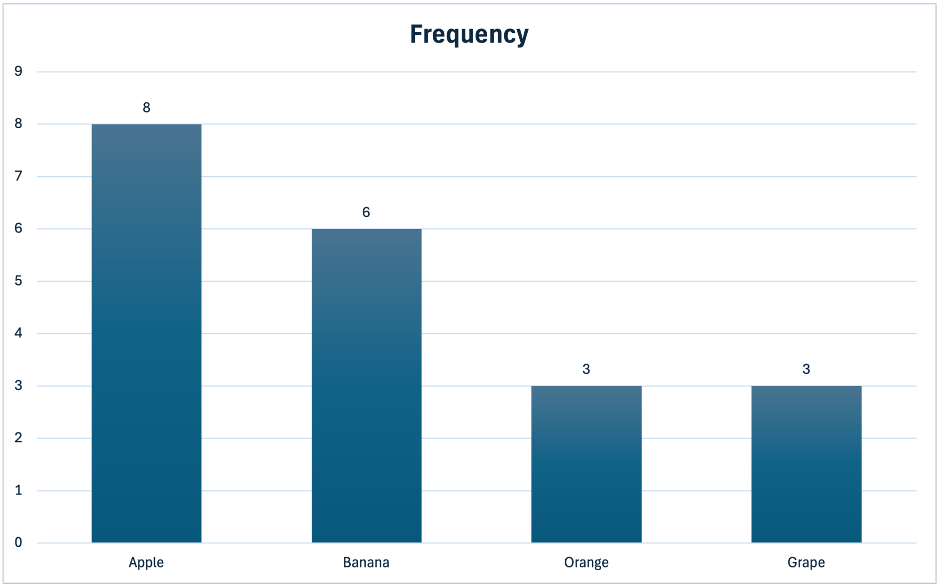

Figure 1: Frequency Bar Graph for Favorite Fruits

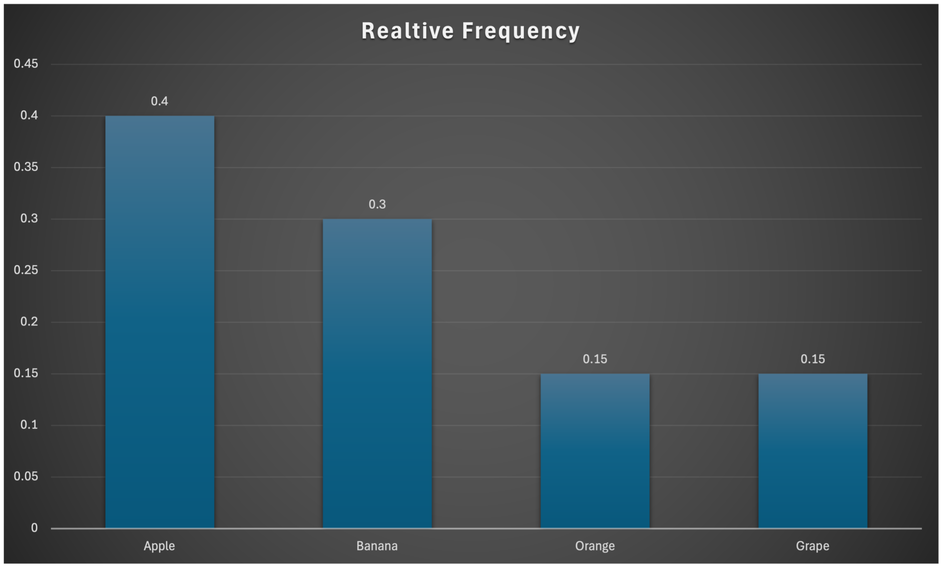

Figure 2: Relative Frequency Bar Graph for Favorite Fruits

- Y-axis: Frequency or Relative Frequency
- Bars touch (quantitative data)

Pie Chart

Dot Plot

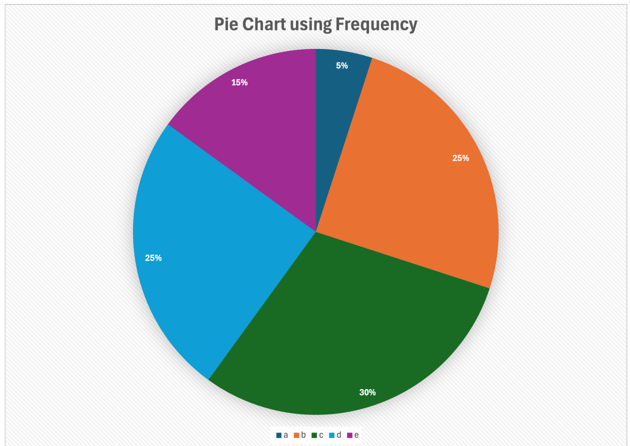

Figure 3: Pie Chart for Favorite Fruits

Definition

A **dot plot** is a simple graphical display of data using dots. Each dot represents one observation and is stacked above its corresponding value on the number line or category. Dot plots are especially useful for small datasets as they preserve the individual data points and show the distribution clearly.

Example: Test Scores

Suppose we have the following test scores from a small class:

$$78, 78, 85, 85, 85, 90, 90, 90, 90, 92, 92, 95, 95, 100, 100$$

This data can be represented as a dot plot:

78	**85**	**90**	**92**	**95**	**100**
● ●	● ● ●	● ● ● ●	● ●	● ●	● ●

- Each dot corresponds to one student's test score.
- The dots stacked vertically show how many students scored each value.
- This plot helps visualize clusters, gaps, and outliers in the data.

Objectives

- Determine the standard deviation of a variable from raw data
- Differentiate between population and sample standard deviation
- Understand degrees of freedom
- Apply the conceptual and computational formulas for standard deviation

1 Why Use Standard Deviation?

Measures of dispersion tell us how spread out data is. Standard deviation reflects how far each data point is from the mean, on average.

Problem: The average of deviations from the mean is always 0.

$$\sum(x_i - \mu) = 0 \quad \text{and} \quad \sum(x_i - \bar{x}) = 0$$

To avoid this, we either:

1. Use absolute values ⇒ Mean Absolute Deviation (MAD)
2. Square the deviations ⇒ Variance, then take the square root ⇒ Standard Deviation

2 Population Standard Deviation

Notation

- x_i: Data values
- μ: Population mean
- N: Number of observations
- σ: Population standard deviation

Conceptual Formula

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

2.1 Population variance

Population variance is a measure of how much the values in an entire population are spread out or dispersed around the population mean. It quantifies the average of the squared differences between each data point and the population mean.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

3 Sample Standard Deviation

Notation

- x_i: Sample values
- $\bar{x}$: Sample mean
- n: Sample size
- s: Sample standard deviation

Conceptual Formula

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

4 Why Divide by $n-1$?

Because the deviations from the mean must sum to zero, the last value is fixed once the first $n-1$ values and the mean are known.

$\Rightarrow$ **Degrees of Freedom** $= n-1$

Only $n-1$ values are free to vary.

Example: If $\bar{x}=4$, $n=3$, and $x_1=2$, $x_2=3$, then:

$$\frac{2+3+x_3}{3}=4 \Rightarrow x_3=7$$

4.1 Sample Variance

The sample variance s^2 is the average of the squared distances of the data points from the sample mean.

It is obtained by squaring the sample standard deviation s.

Conversely, the sample standard deviation s is the square root of the sample variance s^2.

In formula form:

$$s^2 = \text{sample variance} s = \sqrt{s^2}$$

5 Example: Sample Standard Deviation

Data: $4, 7, 6, 9$

Step 1: Compute the mean

$$\bar{x}=\frac{4+7+6+9}{4}=6.5$$

Step 2: Compute squared deviations

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
4	-2.5	6.25
7	0.5	0.25
6	-0.5	0.25
9	2.5	6.25

$$\sum (x_i-\bar{x})^2 = 13$$

Step 3: Compute standard deviation

$$s=\sqrt{\frac{13}{3}}\approx\sqrt{4.33}\approx 2.08$$

Answer: Sample standard deviation $s\approx 2.08$

Example 1

Find the sample standard deviation of the data set:

$$-2, 0, 2, 6, 8, 10$$

Step 1: Calculate the sample mean $\bar{x}$:

$$\bar{x} = \frac{-2+0+2+6+8+10}{6} = \frac{24}{6} = 4$$

Step 2: Calculate squared deviations $(x_i - \bar{x})^2$:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
-2	-6	36
0	-4	16
2	-2	4
6	2	4
8	4	16
10	6	36

Step 3: Sum of squared deviations:

$$36 + 16 + 4 + 4 + 16 + 36 = 112$$

Step 4: Sample variance s^2:

$$s^2 = \frac{112}{6-1} = \frac{112}{5} = 22.4$$

Step 5: Sample standard deviation s:

$$s = \sqrt{22.4} \approx 4.733$$

—

Example 2

Find the sample standard deviation of the data set:

$$1, -2, 5, 2, 1$$

Step 1: Calculate the sample mean $\bar{x}$:

$$\bar{x} = \frac{1+(-2)+5+2+1}{5} = \frac{7}{5} = 1.4$$

Step 2: Calculate squared deviations $(x_i - \bar{x})^2$:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	$1 - 1.4 = -0.4$	0.16
-2	$-2 - 1.4 = -3.4$	11.56
5	$5 - 1.4 = 3.6$	12.96
2	$2 - 1.4 = 0.6$	0.36
1	$1 - 1.4 = -0.4$	0.16

Step 3: Sum of squared deviations:

$$0.16 + 11.56 + 12.96 + 0.36 + 0.16 = 25.2$$

Step 4: Sample variance s^2:

$$s^2 = \frac{25.2}{5-1} = \frac{25.2}{4} = 6.3$$

Step 5: Sample standard deviation s:

$$s = \sqrt{6.3} \approx 2.510$$

Empirical Rule

If the distribution of data is roughly bell-shaped (normal distribution), then:

– Approximately 68% of the data will lie within 1 standard deviation of the mean.

$$\mu - \sigma \leq X \leq \mu + \sigma$$

– Approximately 95% of the data will lie within 2 standard deviations of the mean.

$$\mu - 2\sigma \leq X \leq \mu + 2\sigma$$

– Approximately 99.7% of the data will lie within 3 standard deviations of the mean.

$$\mu - 3\sigma \leq X \leq \mu + 3\sigma$$

Note: We can apply the Empirical Rule to sample data by using the sample mean $\bar{x}$ in place of the population mean μ, and the sample standard deviation s in place of the population standard deviation σ.

End of Lecture #4