Understanding Variability: Measures of Dispersion in Real Life

Objective:

The goal of this project is to apply measures of dispersion (range, variance, standard deviation, and interquartile range) to real-world data, interpret the results, and draw meaningful conclusions about how variability impacts decision-making in practical scenarios.

I. Data Collection:

The dataset will focus on the prices of a common product, specifically **Toothpaste**, sold in different stores in Dasmariñas, Cavite, Philippines. The data reflects prices from two different types of retail outlets:

- **Sari-sari stores**: Prices range from 10 PHP to 13 PHP per piece.
- **Grocery stores**: Prices range from 8 PHP to 10 PHP per piece, but purchasing one piece only is not possible in grocery stores.

The dataset below shows the 15 prices from two types of stores, Sari-sari stores and grocery stores.

Store Type	Store Name	Price (PHP)
Sari-sari	Tope Store	10
Sari-sari	Robena Store	11
Sari-sari	Dali's Mart	12
Sari-sari	Brito Store	13
Sari-sari	Tina's Store	11
Sari-sari	Nena's Store	12

Sari-sari	Roxas Store	10
Sari-sari	Juanita's Store	13
Sari-sari	Pablo's Sari-Sari	10
Grocery	Bautista Grocery	10
Grocery	Santos Grocery	8
Grocery	Carla's Store	8
Grocery	Del Mundo Grocery	9
Grocery	Ace Grocery	8
Grocery	Pwedo's Grocery	9

II. Compute Measures of Dispersion:

The Data below shows the price of Toothpaste in 15 different stores:

10	**11**	**12**	**13**	**11**
12	**10**	**13**	**10**	**10**
8	**8**	**9**	**8**	**9**

The Data below shows the price of Toothpaste in 15 different stores arranged in ascending order:

8	**8**	**8**	**9**	**9**
10	**10**	**10**	**10**	**11**
11	**12**	**12**	**13**	**13**

- **Range**: The range is the difference between the maximum and minimum values.

  The range is the difference between the maximum and minimum values in the dataset. It gives a quick measure of how spread out the values are.

- ❖ Maximum value = 13
- ❖ Minimum value = 8

Therefore, the range is calculated as:

Range = Maximum − Minimum = 13 − 8 = **5**

This tells us that the prices in the combined data set vary by 5 units from the lowest to the highest price.

- **Variance**: is the measure of how far each number in the dataset is from the mean, and therefore how spread out the numbers are. To calculate the variance, we need to subtract the mean from each data point, square the result, and then find the summation of those squared.

Find the mean (Ẍ) of the combined dataset: The combined dataset has 15 prices:

(8, 8, 8, 9, 9, 10, 10, 10, 10, 11, 11, 12, 12, 13, 13).

$\ddot{X} = \Sigma x/n = 154/15 = 10.2666$

Mean (Ẍ) is: 10.27

Price	Deviation (X - Ẍ)	(X - Ẍ)²
8	$(8 - 10.27)^2$	$(-2.27)^2 = 5.1529$
8	$(8 - 10.27)^2$	$(-2.27)^2 = 5.1529$
8	$(8 - 10.27)^2$	$(-2.27)^2 = 5.1529$

9	$(9 - 10.27)^2$	$(-1.27)^2 = 1.6129$
9	$(9 - 10.27)^2$	$(-1.27)^2 = 1.6129$
10	$(10 - 10.27)^2$	$(-0.27)^2 = 0.0729$
10	$(10 - 10.27)^2$	$(-0.27)^2 = 0.0729$
10	$(10 - 10.27)^2$	$(-0.27)^2 = 0.0729$
10	$(10 - 10.27)^2$	$(-0.27)^2 = 0.0729$
11	$(11 - 10.27)^2$	$(0.73)^2 = 0.5329$
11	$(11 - 10.27)^2$	$(0.73)^2 = 0.5329$
12	$(12 - 10.27)^2$	$(1.73)^2 = 2.9929$
12	$(12 - 10.27)^2$	$(1.73)^2 = 2.9929$
13	$(13 - 10.27)^2$	$(2.73)^2 = 7.4529$
13	$(13 - 10.27)^2$	$(2.73)^2 = 7.4529$
Σ 154		Σ 40.9335

The table shows the deviations of each price from the mean (10.27) and their squared values.

$$s^2 = \Sigma(X - \ddot{X})^2 / n - 1$$

$$s^2 = 40.9335 / 15 - 1$$

$$s^2 = 40.9335 / 14$$

$$s^2 = 2.923821429 = 2.92$$

The variance (s^2) is calculated by dividing the sum of squared deviations (40.9335) by n−1, where n is the number of data points (15). This gives the formula 40.9335 / 14 = **2.92**. The variance of **2.92** shows the average squared deviation from the mean, highlighting the spread of the data points around the mean.

- **Standard Deviation**: The square root of the variance.

$$s = \sqrt{\Sigma(X - \ddot{X})^2 / n - 1}$$

$$s = \sqrt{2.923821429}$$

$$s = 1.709918545$$

$$s = 1.71$$

The standard deviation (s) is the square root of the variance, which is 2.923821429. Taking the square root gives a standard deviation of **1.71**, indicating that, on average, each data point deviates from the mean by **1.71**.

- **Interquartile Range (IQR)**: The difference between the first (Q1) and third quartile (Q3), representing the middle 50% of the data.

$$Q_1 = (1/4)(n + 1) \qquad Q_3 = (3/4)(n + 1)$$

$$Q_1 = (0.25)(15 + 1) \qquad Q_3 = (0.75)(15 + 1)$$

$$Q_1 = (0.25)(16) = 4th\ data\ point \qquad Q_3 = (0.75)(16) = 12th\ data\ point$$

$$Q_1 = 9 \quad Q_2 = 10 \quad Q_3 = 12$$

$$IQR = Q_3 - Q_1 = 12 - 9 = 3$$

$$IQR = 3$$

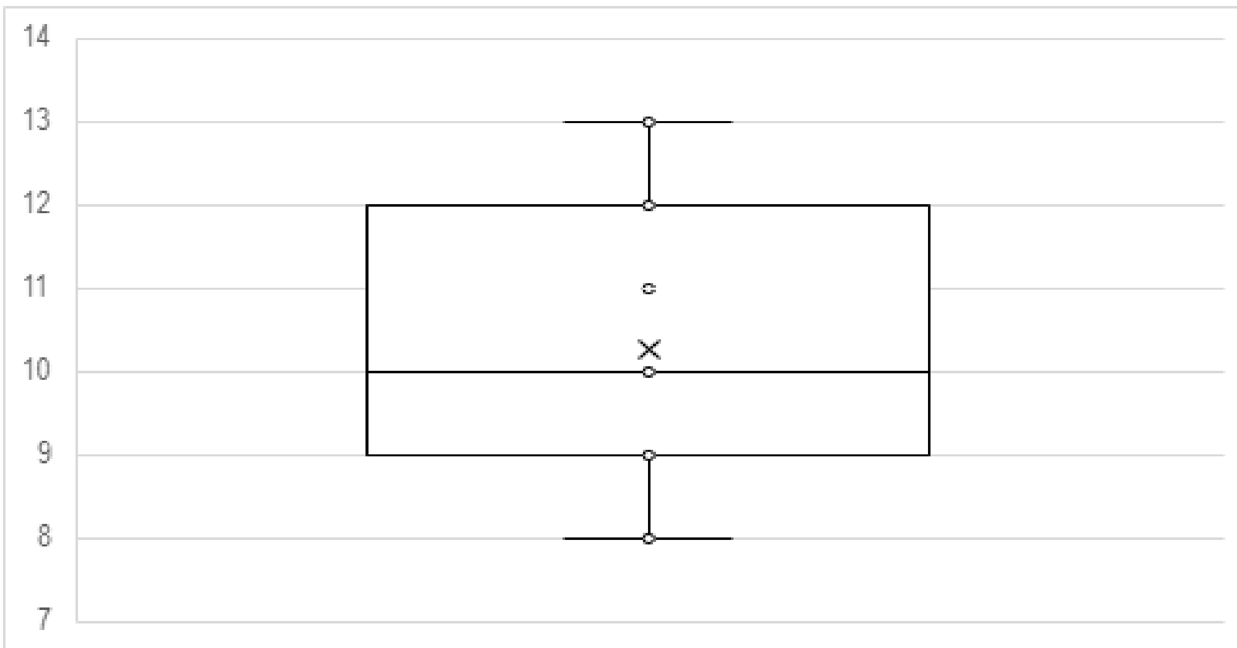
In order to calculate the interquartile range (IQR), we need to determine the positions of the first quartile (Q1) and third quartile (Q3). Using the formula Q1=(1/4)(n+1) and Q3=(3/4)(n+1), with n as 15, we find Q1 at the 4th data point and Q3 at the 12th data point in the list. The values of these positions are

Q1 = 9 and Q3 = 12, therefore the IQR is Q3 − Q1 = 12 − 9 = **3**. This means the middle 50% of the data falls within a range of **3** units

III. Visualization:

To visually present the data:

Box and Whisker Plot



The Box in the Box and Whisker plot starts at the first quartile (Q1), which is 9, and ends at the third quartile (Q3), which is 12, representing interquartile range (IQR). The whiskers, which shows the spread of the data, extend from the center of the box: the bottom whisker stretches from Q1 (9) to the minimum value of 8, while the top whisker extends from Q3 (12) to the maximum value of 13. The line inside the box on the other hand shows the median (Q2), which is 10, marking the center of the box.

V. Analysis and Interpretation:

The variation in the toothpaste prices shows that while the prices are spread out, they are generally close to the average price of 10.27. The **range** of 5 (from 8 to 13) shows us that the prices differ, but not too much. The **standard deviation** of 1.71 means that most of the prices are within 1.71

units of the average, showing a moderate amount of spread. The **variance** of 2.92 supports this too, by suggesting that while the prices aren't the same, they are not extremely different. The **interquartile range (IQR)** of 3 (between 9 and 12) shows that half of the prices are pretty close to each other, with less variation in the middle of the dataset.

There are not any **outliers** since all the prices fall within the range of 8 to 13. The prices are consistent, therefore the consumers can expect similar prices across stores, although some stores might have slightly higher or lower prices. Overall, the dataset shows that the prices of toothpaste are stable, and the small differences in price don’t affect the reliability of the data much.