Measures of Dispersion (Spread)

Measures of dispersion, also known as measures of spread or variability, are statistical tools used to quantify how spread out the data points are in a data set relative to a central tendency measure (like mean or median). In simpler terms, they tell you how much the data points vary around the "center" of the data set.

1. Range

The range, in statistics, is the simplest measure of dispersion (spread) used to describe how spread out the data points are in a data set. It tells you the distance between the extreme values in your data, indicating how much the data varies from the lowest to the highest value.

Calculating the Range:

The range is calculated by subtracting the minimum value (smallest number) from the maximum value (largest number) in your data set. Here's the formula:

$$Range = MaximumValue - MinimumValue$$

For example, consider the data set: {2, 7, 9, 14, 18}.

- Maximum value = 18
- Minimum value = 2

Therefore, the range:

$$Range = 18 - 2 = 16$$

2. Variance

Variance, in statistics, is a measure of dispersion (spread) that tells you how far, on average, each data point in your set is from the mean (average). It essentially quantifies how much the data points vary around their central tendency.

Calculating Variance:

1. **Find the mean:** Calculate the average value of your data set.
2. **Calculate squared deviations:** For each data point, subtract the mean and then square the difference. This removes negative signs and emphasizes the magnitude of the deviation.
3. **Average the squared deviations:** Sum the squared deviations from all data points and divide this sum by the total number of data points (n).

Here's the formula for variance:

$$Variance = \Sigma(x_i - \mu)^2/n$$

Where:

- Σ (sigma) represents the sum of all the squared deviations $(x_i - \mu)^2$.
- x_i represents each individual value in your data set.
- μ (mu) represents the mean of the data set.
- n represents the total number of values in your data set.

3. Standard deviation (SD)

Standard deviation (SD), denoted by the symbol σ (sigma), is a fundamental measure of dispersion (spread) in statistics. It tells you how much, on average, the data points in your set deviate from the mean (average). In simpler terms, it quantifies how much variation or spread exists around the central tendency of your data.

Calculating Standard Deviation:

1. **Find the mean:** Calculate the average value of your data set.
2. **Calculate squared deviations:** For each data point, subtract the mean and then square the difference. This removes negative signs and emphasizes the magnitude of the deviation.
3. **Average the squared deviations:** Sum the squared deviations from all data points and divide this sum by the total number of data points (n). This gives you the variance.
4. **Take the square root:** Finally, take the square root of the variance to get the standard deviation.

Here's the formula for standard deviation:

$$Standard Deviation (SD) = \sqrt{(Variance)} = \sqrt{[\Sigma (x_i - \mu)^2 / n]}$$

Where:

- Σ (sigma) represents the sum of all the squared deviations $(x_i - \mu)^2$.
- x_i represents each individual value in your data set.
- μ (mu) represents the mean of the data set.
- n represents the total number of values in your data set.

IQR

Interquartile Range, is a measure of dispersion (spread) used in statistics to describe the variability within the middle 50% of your data set. It tells you how spread out the majority of the data points are, focusing on the central part of the distribution and offering a more robust measure compared to the range, which can be easily influenced by outliers.

Concept:

Imagine your data set ordered from least to greatest. Divide it into four equal parts (quartiles):

- **First Quartile (Q1):** This represents the 25th percentile. It signifies the value below which 25% of the data points lie.
- **Second Quartile (Q2):** This is the median (50th percentile). It's the middle value when your data is arranged in order.
- **Third Quartile (Q3):** This represents the 75th percentile. It signifies the value below which 75% of the data points lie.

Calculation:

The IQR is simply the difference between the third quartile (Q3) and the first quartile (Q1). Here's the formula:

$$IQR = Q3 - Q1$$