# Data Science Notes by Sarowar Ahmed

📊 **Chapter: Descriptive Statistics**

📊 **Topic: Variance, Standard Deviation**

Hey, LinkedIn family! Ever wondered how we measure the "spread" or "variability" in a set of data? 🤔 Let's break down two fundamental concepts that do just that!

📐 Variance: Imagine you and your friends measure the length of fish you catch on a fishing trip. Some fish are big, some small, but most are around the same size. Variance tells us how much the fish lengths differ from the average length. It's like asking, "On average, how much do our fish lengths vary from the 'typical' fish length we catch?" To find it, we calculate the difference between each fish's length and the average length, square those differences (to get rid

of negative numbers), and then find the average of those squared differences.

🔍 Standard Deviation: If variance is how we measure the spread of our fish lengths squared, standard deviation is the square root of that variance. Why take the square root? It brings our measure of spread back to the same units as our original data, making it easier to understand. So, if we're measuring fish in centimeters, standard deviation gives us a spread in centimeters too, not squared centimeters!

🎨 Why They Matter: Both these statistics help us understand the diversity or uniformity in our data. If you're analyzing anything from heights in a classroom to temperatures during a month, knowing how spread out your data is can provide deep insights. High variance or standard deviation means your data points are spread out widely; low variance or standard deviation means they're more clustered closely.

## Variance and Standard Deviation

**Variance Formula:**

For a dataset with values $x_1, x_2, ..., x_n$ and mean (average) $\mu$, the variance ($\sigma^2$) is calculated as:

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

In simpler terms, variance measures the average of the squared differences from the Mean.

**Standard Deviation Formula:**

Standard deviation ($\sigma$) is the square root of variance:

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

Standard deviation provides a measure of the dispersion or variation in a set of values, in the same units as the data.

📈 Example to Differentiate Both:

Imagine you have recorded the heights (in cm) of five sunflowers in your garden: 150, 155, 160, 165, and 170.

Mean (Average) Height:

$\mu=(150+155+160+165+170)/5=160$ cm

Variance Calculation:

$\sigma^2=((150-160)^2+(155-160)^2+(160-160)^2+(165-160)^2+(170-160)^2)/5$

$\sigma^2=((-10)^2+(-5)^2+0^2+5^2+10^2)/5=(100+25+0+25+100)/5=50$ cm 2

Standard Deviation Calculation:

$\sigma=\sqrt{50}\approx7.07$ cm

🌟 Understanding the Difference:

Variance (50 cm²) tells us that, on average, the heights of the sunflowers vary 50 square centimeters from the mean height. It's a bit abstract because we're dealing in square units, making it hard to visualize the variability of the heights directly.

Standard Deviation (approximately 7.07 cm), on the other hand, is much more intuitive. It tells us that, on average, the heights of the sunflowers deviate about 7.07 cm from the mean height. This gives us a clearer, more relatable picture of the spread of the sunflower heights, directly in the units of the original data.

Remember, data tells stories, and understanding the spread of our data helps us get to the heart of these tales. Whether it's the consistency of your morning coffee's temperature or the fluctuation in your daily steps, variance and standard deviation are your keys to unlocking these narratives.

Got any questions about Variance and Standard Deviation! ? Feel free to ask me via Linkedin! Let's keep learning together.

<u>My LinkedIn</u>                                    Date: 03/04/2024