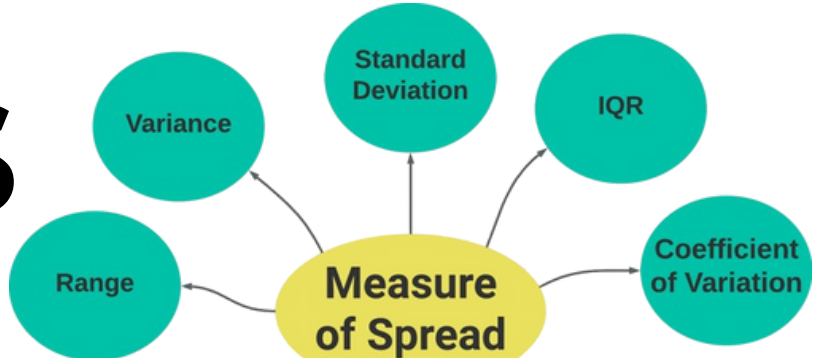
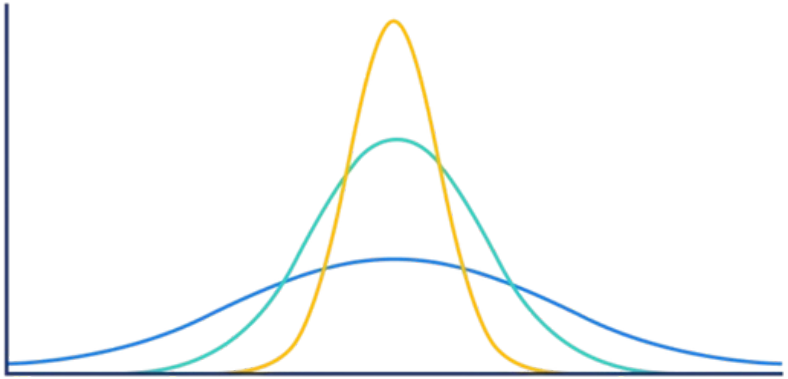
Data Analyst Perspective

STATISTICS



DESCRIPTIVE STATISTICS

Variability Part-1

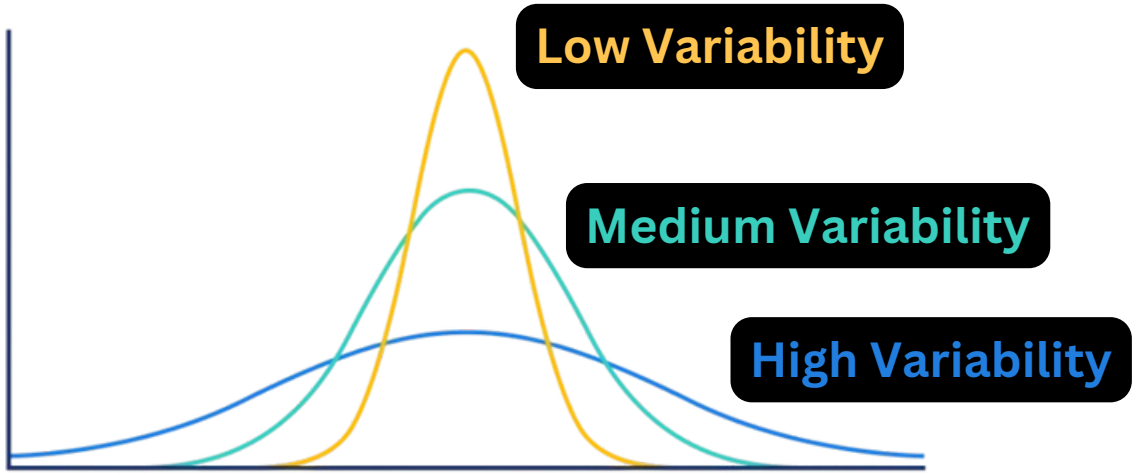


Agenda

- WHAT IS VARIABILITY
- KEY COMPONENTS OF VARIABILITY
- GET TO KNOW EACH COMPONENTS
- CHARACTERISTICS & APPLICATIONS OF EACH COMPONENTS

WHAT
IS VARIABILITY

- Also known as the **measure of dispersion,** describes how spread out or scattered the data points are in a dataset.
- By measuring variability, we can compare the spread of different datasets, even if their means or medians are similar.
- Variability is **crucial in statistical inference**, as it affects **confidence intervals, hypothesis tests, and other inferential statistics.**

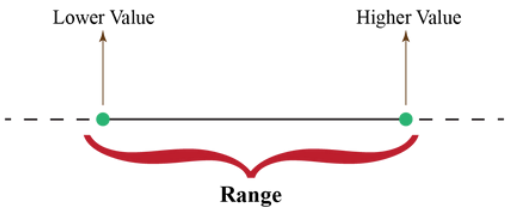


- Try to understand the distribution of 3 datasets/samples and find out which factor is distinguishing them from each other ??

KEY COMPONENTS
OF VARIABILITY

- Range
- Variance
- Standard Deviation
- Interquartile Range (IQR)
- Coefficient of Variation (CV)

RANGE



- Simplest measures of variability in descriptive statistics. It provides a basic idea of how spread out the values in a data set are. The range is calculated as:

  Range=Maximum Value−Minimum Value

- The range only takes into account the extreme values and **does not provide information** about the distribution of the values between the extremes.

VARIANCE

- Variance tells us how each value in the data set **differs from the mean (average)**. It provides insight into the overall dispersion of values within the dataset.
- Variance quantifies the **average squared deviation** of each data point from the mean of the dataset.

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

Population Variance

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

Sample Variance

μ : Population mean

N :Total number of Data points

$\bar{x}$: Sample mean

Variance Continue.....

- In the previous page you must have noticed a slight difference in formula of Sample Variance & Popular Variance.
- You'll encounter this **(n-1)** everywhere ,whenever we'll talk variability for sample.
- **Interview Question : Why do we use n-1 instead of n for a sample's variance ?**

STANDARD DEVIATION

STANDARD DEVIATION = $\sqrt{\text{VARIANCE}}$

- A lower standard deviation indicates that the data points tend to be close to the mean.
- A higher standard deviation indicates that the data points are spread out over a larger range of values.
- The standard deviation is commonly used because **it is in the same units as the data, making it more interpretable.**

- **Interview Question :** Ok so **Standard Deviation** is simply the under root of **Variance**, then why do we square the deviations when calculating variance ?

APPLICATION
OF VARIANCE & SD

- **Outlier Detection:** High standard deviation might **indicate potential outliers**. By identifying data points that lie far from the mean, you can investigate anomalies or errors.
- **Financial Analysis:** Standard deviation is often used to **assess the risk associated with an investment**. Higher standard deviation indicates higher volatility and risk.
- **Error Analysis:** In regression, the standard deviation of the residuals (errors) helps assess the **goodness of fit of the model**. Lower residual variance indicates a better fit.

TASK FOR YOU

- Today, we've encountered two intriguing interview questions on our learning journey, and one of them is particularly renowned and significant.
- I'd love to hear your thoughts and answers in the comments—**let's learn and grow together!**
- Also share atleast 1 application of these variability measures

THANK YOU

Share your thoughts and feedback !!