



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data
science and Analytics

II Year / IV Semester

Unit 2 – Descriptive Analytics

Describing variability





What is Variability?



Variability measures how data points differ from the central value (average). It helps in understanding the spread of data. High variability suggests unpredictable outcomes, while low variability indicates consistency. This is important for making data-driven decisions.



Why is Variability Important?



Variability allows for comparison between datasets, highlighting how spread out the data is. It indicates the level of consistency by showing how closely data points cluster around the mean. High variability suggests unpredictability, and low variability suggests stability.



Measure 1: Range?



The range is the difference between the maximum and minimum values in a dataset. It provides a simple measure of spread. For example, if a dataset is [10, 20, 30, 40, 50], the range is 40 (50 - 10). However, range may be sensitive to outliers.



Measure 2: Interquartile Range (IQR)



IQR is the range of the middle 50% of data, calculated as the difference between the third quartile (Q3) and first quartile (Q1). For the dataset [10, 20, 30, 40, 50], Q1 is 20 and Q3 is 60, so the IQR is 40. It gives a better understanding of spread by ignoring outliers.



Measure 3: Variance



Variance measures the average squared deviation from the mean. It shows the overall spread of data around the mean. For the dataset [10, 20, 30], with a mean of 20, the variance is 66.67. High variance indicates a wide spread in data points.



Measure 4: Standard Deviation (SD)



Standard Deviation is the square root of variance and gives variability in the same units as the data. If the variance is 66.67, then the standard deviation is 8.16. A higher standard deviation means more spread out data, while a lower SD indicates data points are closer to the mean.



Measure 5: Coefficient of Variation (CV)



The Coefficient of Variation (CV) expresses the standard deviation as a percentage of the mean. For a dataset with a mean of 20 and an SD of 8.16, the CV is approximately 40.8%. This allows comparisons of variability across datasets with different means.



Visualizing Variability



Variability can be visualized using boxplots, histograms, and line plots. A boxplot, for instance, shows the IQR and the range visually. It's useful for identifying outliers and understanding the spread and central tendency of data at a glance.



Applications of Variability



In business, variability helps assess sales consistency across regions. In education, it highlights performance gaps in student test scores. In healthcare, understanding variability in recovery times aids in optimizing treatment plans for patients. Variability is essential for data-driven decisions.



THANK YOU