CODE BASICS

CHAPTER TITLE	Measures Of Central Tendency and Dispersion

LECTURE TITLE	Descriptive vs Inferential Statistics

1 Inferential statistics involve making predictions or drawing conclusions about a population based on a sample.

2 Descriptive statistics are used to summarize and describe data, providing an overview of its main characteristics.

CODE BASICS

CHAPTER TITLE	Measures Of Central Tendency and Dispersion

LECTURE TITLE	Measures of Central Tendency: Mean, Median, Mode

1 **Mean** - Synonym for average.

2 **Median**-Middle value of a dataset when it is ordered in ascending order.

3 If the dataset has an **even number of values,** the median is the **average of the two middle values.**

4 **Mode** - Most frequently occurring data value.

CODE BASICS

| CHAPTER TITLE | Measures Of Central Tendency and Dispersion |

| LECTURE TITLE | Percentile |

1 A percentile is a statistical measure used to rank a value within a dataset, indicating what percentage of the data falls below or is equal to that value.

2 **25th Percentile:** The value below which 25% of the data falls; it represents the lower quartile.

3 **50th Percentile (Median):** The middle value of the dataset; it divides the data in half.

4 **75th Percentile:** The value below which 75% of the data falls; it corresponds to the upper quartile

1 "IQR" and "Range" also referred to as measures of **dispersion or variability**.

2 Range, calculated as **Maximum Value - Minimum Value,** reflects data spread.

3 Unlike Range, IQR(Inter Quartile Range) is less influenced by **outliers**, making it a robust measure.

4 Quartiles **Q1, Q2, and Q3** correspond to the **25th, 50th, and 75th** percentiles, respectively.

5 The **50th** percentile is commonly known as the median.

6 IQR is the difference between **Q3 and Q1**, showing the spread of the **middle 50%** of data.

CODE BASICS

CHAPTER TITLE	Measures Of Central Tendency and Dispersion

LECTURE TITLE	Box or Whisker Plot

1 Box/Whisker Plots provide a visual summary of the **central tendency, spread, and presence of outliers** in a dataset.

2 The "**box**" in a Box Plot shows the **middle 50%** of the data, with the line inside representing the **median**. It gives insights into the central data values and their spread.

3 The "**whiskers**" in a Box Plot show the **minimum and maximum** values within a specific range.

CODE BASICS

CHAPTER TITLE	Measures Of Central Tendency and Dispersion

LECTURE TITLE	Outlier treatment using IQR and box plot

1 The **Interquartile Range (IQR)** is determined by subtracting the first quartile (Q1) from the third quartile (Q3).

2 The Lower Limit is calculated as the first quartile (Q1) **minus 1.5 times** the IQR.

3 The Upper Limit is derived by **adding 1.5 times** the IQR to the third quartile (Q3).

CODE BASICS

KEY TAKEAWAYS

CHAPTER TITLE	Measures Of Central Tendency and Dispersion

LECTURE TITLE	Measures of Dispersion: Variance and Standard Deviation

1 Variance is a measure of how spread out a distribution is. It is calculated as the **average of the squared differences from the mean.**

2 The smaller the variance, the less spread out the data is. Conversely, the larger the variance, the **more spread out** the data is.

3 Standard deviation is a measure of the amount of variation or dispersion of a set of values. It is calculated as the **square root** of the variance.

4 The smaller the standard deviation, the closer the data points are to the mean. Conversely, the larger the standard deviation, the more spread out the data points are.

5 The stock market's volatility is the best use case for variance and standard deviation.

CHAPTER TITLE | Measures Of Central Tendency and Dispersion

LECTURE TITLE | Correlation

1 Correlation is a statistical measure that shows the degree to which two variables are related.

2 A correlation coefficient can range from -1 to 1
  ○ -1 (perfect negative correlation) < 0 (no correlation) < 1 (perfect positive correlation)

CODE BASICS

CHAPTER TITLE	Measures Of Central Tendency and Dispersion

LECTURE TITLE	Correlation vs Causation

1 **Correlation:** A statistical relationship between two variables, where changes in one variable are associated with changes in another, but it doesn't imply causation.

2 **Causation:** A cause-and-effect relationship between variables, where changes in one variable directly lead to changes in another.