Statistics

Descriptive Statistics:

- Descriptive statistics involve the presentation and organization of data to provide a clear summary.
- Common tools include measures of central tendency and measures of dispersion.

Measures of Central Tendency:

- These measures represent the center or average of a data set.
- Mean = average of the data set ⇒ sum of elements divided by the number of elements
- Mean of {2, 4, 8, 6, 10} = (2 + 4 + 8 + 6 + 10) / 5 = 6
- Median = middle number of sorted data set, if number of elements is even then median = average of these two numbers
- Median of {2, 4, 6, 8, 10} is 6
- Mode = The most frequently occurring value
- Mode of {2, 2, 3, 4, 5} is 2

Measures of Dispersion or Variability:

- Range = Max - Min
- Range of {12, 5, 8, 15, 7} = 15 - 5 = 10
- Variance = sum of ((xi - mean of the sample) power 2) / n - 1
- xi = every element in the data set, n = number of elements
- Variance of {4, 7, 1, 9, 3} = 8
- Standard deviation = square root of variance
- Standard Deviation of {4, 7, 1, 9, 3} ≈ 2.83

Quartiles:

- Q1 (First Quartile): The median of the lower half of the dataset.
- Q2 (Second Quartile/Median): Same as the median.
- Q3 (Third Quartile): The median of the upper half of the dataset.
- Example: For the dataset {2, 7, 1, 8, 4, 10, 12}, Q1 = 3, Q2 = 7, Q3 = 10. The IQR = Q3 - Q1 = 7.
- Outliers = [Q1 - 1.5 * IQR - Q3 + 1.5 * IQR]