

Statistical Summaries for AI

Madushanka Wadumulla
PGC in IA
Beng, IIESL, AMEI



Outline

- Types of Matrices
- Transpose
- Determinant
- Inverse

Mean of Data

- Mean is also known as average of all the numbers in the data set which is calculated by below equation.

$$\text{Mean} = \frac{\text{Sum of all data values}}{\text{Number of data values}}$$

$$\bar{X} = \frac{\sum X_i}{n}$$

Variance of Data

- Variance is the numerical values that describe the variability of the observations from its arithmetic mean. Standard Deviation is the square root of the variance.

$$V(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Median of Data

- Median is mid value in this ordered data set.

First, arrange the observations in an ascending order.

**If the number of observations (n) is odd:
the median is the value at position**

$$\left(\frac{n+1}{2} \right)$$

If the number of observations (n) is even:

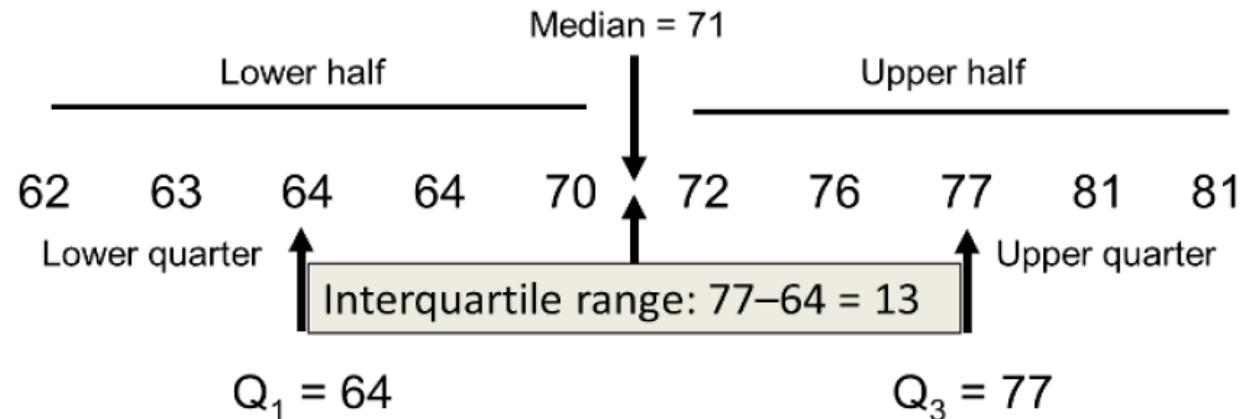
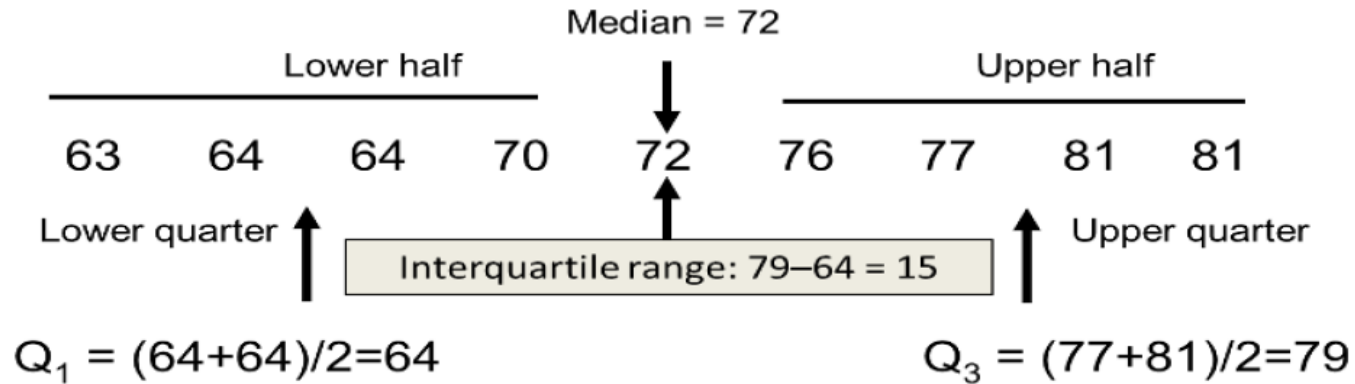
1. Find the value at position $\left(\frac{n}{2} \right)$

2. Find the value at position $\left(\frac{n+1}{2} \right)$

3. Find the average of the two values to get the median.

Five Number Summary of Data

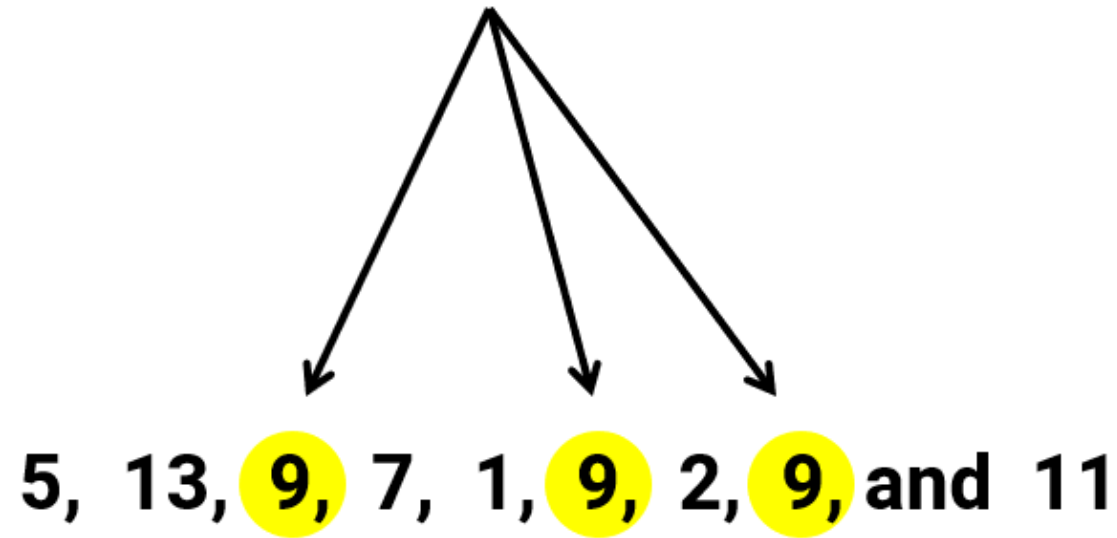
- Minimum, 1st Quartile, 2nd Quartile (Median), 3rd Quartile, Maximum



Mode of Data

- The mode is the value that has highest number of occurrences in a set of data. Unlike mean and median, mode can have both numeric and character data.

Shows up the most!



Mode = 9

Covariance of Two Numerical Variables

- Covariance will measure joint variation of two numerical variables. This is a measure of the relationship between two variables.
- Strength of the association is very high with the increase of covariance value.

$$COV(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) (Y_i - \bar{Y})$$

Correlation of Two Numerical Variables

- Correlation is also a measurement of the relationship between two numerical variables. It is lying between -1 and +1.

$$CORR(X, Y) = \frac{COV(X, Y)}{\sqrt{V(X)V(Y)}}$$

