Estimates of Location

PROF. SARALA MARY

Topics To Be Covered

- Mean

- Median

- Robust

- Outlier

- Mode

Measures of Central Tendancy

- Mean

- Median

- Mode

Estimation of Allocation

It is defined as a single value that is used to describe the "center" of the data

Mean for Ungrouped Data

It is the most commonly used measure of the center of data

It is also referred as the "arithmetic average"

- Computation of Sample Mean

$$\bar{X} = \frac{\Sigma x}{N} = \frac{x_1 + x_2 + x_3 + \ldots x_n}{N}$$

Example:
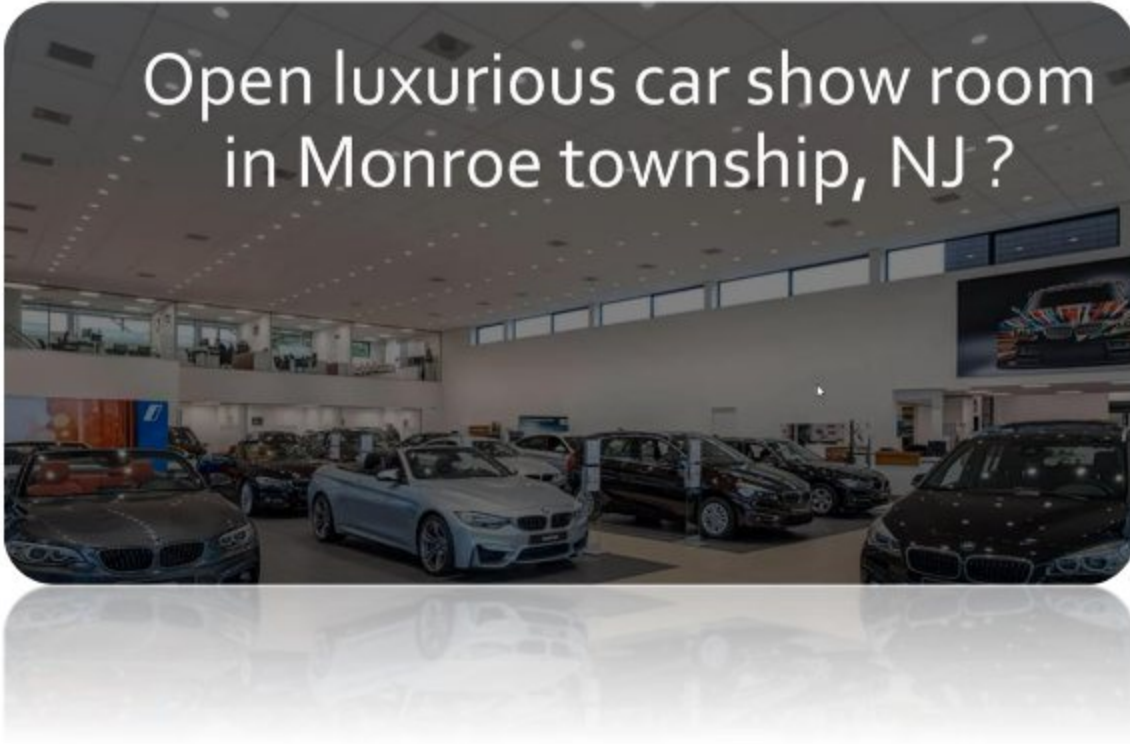Consider the set of numbers {3,5,1,2}

Mean = (3+5+1+2)/4
Mean = 2.75

Example Problem

Find the mean of the data set $\{5,23,8,12\}$

Find the mean of the data set $\{2,7,5,4,6,3\}$

Example: Mean in Data Science

Open luxurious car show room in Monroe township, NJ ?

Find the Average Income

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000
Average	6250

Conclusion: Cannot open showroom

If Suppose…

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000
Elon Musk	10 million
Average	1.43 million

Conclusion: Open the Showroom.
It is a wrong decision.

How to solve this problem?

MEDIAN

- Median is what divides the scores in the distribution into two equal parts.

- Fifty percent (50%) lies below the median value and 50% lies above the median value.

Step 1: Sort the Data

Nina	Rob	Rafiq	Tao	Sofia	Mohan	Elon Musk
4000	5000	6000	7000	7500	8000	10 million

Median

Step 2: Find the middle value (odd values)

Nina	Rob	Rafiq	Tao	Sofia	Mohan	Elon Musk
4000	5000	6000	7000	7500	8000	10 million

Median = 7000

Even values – Find the average of the middle 2 values

Nina	Rob	Rafiq	Tao	Prem	Sofia	Mohan	Elon Musk
4000	5000	6000	7000	8000	7500	8000	10 million

Median = 7500

Median – Example 2

Use case # 2: Handling missing values

Name	Monthly Income ($)	Credit Score	Approve Loan?
Rob	5000	650	No
Rafiq	6000	400	No
Nina	4000	780	Yes
Sofia	??	810	Yes
Mohan	8000	410	No
Tao	7000	850	Yes
Elon Musk	10 million	880	Yes

How to add the missing values?

What if use mean for missing value?

Name	Monthly Income ($)	Credit Score	Approve Loan?
Rob	5000	650	No
Rafiq	6000	400	No
Nina	4000	780	Yes
Sofia	1.6 million	810	Yes
Mohan	8000	410	No
Tao	7000	850	Yes
Elon Musk	10 million	880	Yes

Wrong Conclusion

What if use the Median for the missing value?

4000   5000   6000   7000   8000   10 million

Median = 6500

Name	Monthly Income ($)	Credit Score	Approve Loan?
Rob	5000	650	No
Rafiq	6000	400	No
Nina	4000	780	Yes
Sofia	6500	810	Yes
Mohan	8000	410	No
Tao	7000	850	Yes
Elon Musk	10 million	880	Yes

Median

- The problem is caused due to the outlier value.

- Median solves the problem caused and gives a better result and accuracy

Example - Problem

Find the median of this data:

1, 4, 2, 5, 0

Find the median of this data:

10, 40, 20, 50

How Mean and Median is used in Data Science?

Opening car show room

Descriptive Analysis

Loan approval model

Data Cleaning (Filling NA Values)

What is Mode?

Mode means most frequently occurring value in a dataset

Example: From the given dataset, decide the dish for dinner

Mode here is Mexican

Name	Restaurant Vote
Rob	Mexican
Rafiq	Mexican
Nina	Italian
Sofia	Thai
Mohan	Italian
Tao	Mexican
Bantu	Indian

Example - Problem

Find the mode of the data:

0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 5

Find the mode of the data:

0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 4