



Unit 2 Statistics of One Variable

Day 5 Measures of Central Tendency



Measures of Central Tendency

- In this section we attempt to describe a set of numeric data using a single value.
- The value you calculate will describe the centre of the set of data.
- There are various ways to describe this notion of centre:
 - The **mode** is the value that occurs most frequently
 - The **median** is the middle value of the data when it is organized from lowest to highest
 - The **mean** is the average of the data values

Ungrouped Data

Ungrouped data has not been organized into intervals or categories. You are dealing with the raw data.

Example 1

Two classes that wrote the same physics exam had the following results.

Class A	45	51	55	64	66	70	71	71	71	73	82	83	84	90	95
---------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Mode: 71 – since 71 occurs most frequently

Median: 71 – since 71 is the middle value

$$\begin{aligned}\text{middle} &= \frac{n + 1}{2} = \frac{15 + 1}{2} \\ &= 8\end{aligned}$$

Class A	45	51	55	64	66	70	71	71	71	73	82	83	84	90	95
---------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

$$\begin{aligned}\text{Mean: } \mu &= \frac{\sum x}{n} \\ &= \frac{45 + 51 + 55 + \dots + 90 + 95}{15} \\ &= \frac{1071}{15} \\ \therefore \mu &= 71.4\end{aligned}$$

- The Greek letter Σ (sigma) indicates all the individual values of x are added together.
i.e. $\Sigma x = x_1 + x_2 + \dots + x_n$
- The sum is divided by the number of values in the set, n .

Class B 12 15 44 50 54 61 61 63 69 73 74 75 80 80 81 84 91 92

Mode:

- There are two modes – 61 and 80
- This is called a bimodal set of data

$$\text{Median: } \frac{69 + 73}{2} = 71 \quad \text{middle} = \frac{n + 1}{2} = 9.5$$

Since there are 18 data values the middle is between the 9th and 10th values

$$\begin{aligned} \text{Mean: } \mu &= \frac{\sum x}{n} = \frac{12 + 15 + 44 + \dots + 91 + 92}{18} \\ &= \frac{1159}{18} \\ \therefore \mu &\doteq 64.4 \end{aligned}$$

	Class A	Class B
Mode	71	61, 80
Median	71	71
Mean	71.4	64.4

Observations:

- Mean for Class A is significantly higher however the medians for both classes are the same
- The measures of central tendency for Class A are all similar but this is not true for Class B
- If you look closely at the two sets of data, they are very similar except for two low scores of 12 and 15 in Class B. Data points like these that are much higher or much lower than the rest of the data are called **outliers**.
- If the outliers are removed, the mean for Class B would be 70.1, which demonstrates the effect that outliers can have on central tendency.

Grouped Data

Example 2

A sample of car owners was asked how old they were when they got their first car. Determine the measures of central tendency.

AGE	16-20	21-25	26-30	31-35	36-40
Frequency	10	18	12	8	2

Mode – most car sampled were in 21-25 years old

- The mode is 21-25

Median – There are 50 data points. The middle of the data will be between the 25th and 26th values.

- The median age is 21-25.

$$\text{middle} = \frac{50 + 1}{2} = 25.5$$

$$\text{Mean } \mu = \frac{\sum f \times m}{\sum f}$$

Sum of the frequencies

Sum of the frequency of each interval times the midpoint of the interval

Age	f frequency	m midpoint	$f \times m$
16-20	10		
21-25	18		
26-30	12		
31-35	8		
36-40	2		

$$\sum f =$$

$$\sum f \times m =$$

$$\mu = \frac{\sum f \times m}{\sum f}$$



Weighted Mean

Sometimes certain data within a set are more significant than others.

For example, in MDM4U your term work is weighted at knowledge 35%, application 35%, communication 15% and thinking 15%.

The formula for a weighted mean is

$$\bar{x}_w = \frac{\sum xw}{\sum w}$$

where x represents a data value and w represents its weight or frequency for that data value.



Example 3

In MDM4U the 70% term categories are weighed as follows: knowledge 24.5%; application 24.5%; communication 4.5%; Thinking 4.5%. A student's marks in these categories are 85, 80, 82 and 60, respectively. Calculate the student's mark.

Test Scores	Weight	Test Scores \times Weight
Total =		