Lesson One :Descriptive Statistics: Measures of Central Tendency

Learning Outcomes

By the end of this lesson, students will be able to:

1. Define **descriptive statistics** and explain its purpose in summarizing numerical data.
2. Identify and calculate the three main **measures of central tendency**: mean, median, and mode.
3. Explain the significance of **measures of dispersion** and calculate range and standard deviation.
4. Interpret data distributions (normal and skewed) and understand their impact on central tendency measures.
5. Use graphs (bar charts, histograms, line graphs, pie charts, scattergrams) to present and interpret quantitative data.
6. Differentiate between correlation types (positive, negative, zero) and understand the difference between correlation and causation.

Introduction

Descriptive statistics are essential tools in research for summarizing and interpreting numerical data. They allow researchers to identify patterns, trends, and typical values in a dataset. These statistics are divided into two main categories:

1. **Measures of Central Tendency** – provide a single value that best represents a 'typical' or average observation in a dataset. This includes the **mean**, **median**, and **mode**.
2. **Measures of Dispersion** – describe how spread out the data values are. Common measures include the **range** and **standard deviation**.

Descriptive statistics are often complemented by **visual representations** such as bar charts, histograms, line graphs, and pie charts, which help communicate findings clearly. Understanding the distribution of data—whether **normal** or **skewed**—is crucial for accurate interpretation.

Key terms

Descriptive statistics – The use of graphs, tables and summary statistics to identify trends and analyse sets of data.

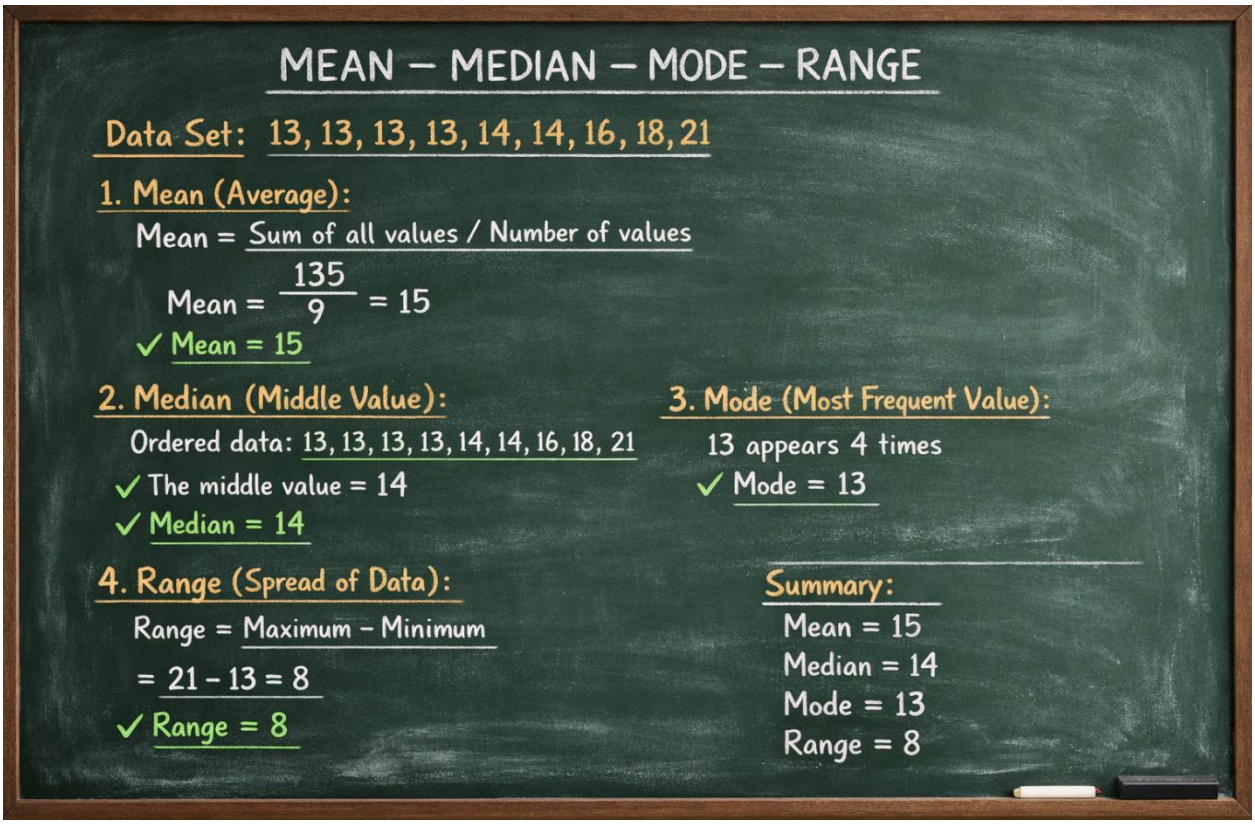
Measures of central tendency –A descriptive statistic that provides information about a 'typical' value for a data set.

Mean – The arithmetic average calculated by adding up all the values in a set of data and dividing by the number of values there are.

Median – The middle value of a data set when the items are placed in rank order..

Mode – The most frequently occurring value in a set of data..

Quantitative Data measured in numbers.



Measures of central tendency

Measures of central tendency are 'averages' which give us information about the most *typical* values in a set of data. There are three of these to consider: the **mean**, the **median** and the **mode**.

Mean

The mean is what most of us will recognise as the average. It is The arithmetic average. Calculated as: by adding up all the scores or values in a data set and dividing this gure by the total number of scores there are.

Example: For scores 5, 7, 7, 9, 10, 11, 12, 14, 15, 17 → Mean = 107 ÷ 10 = 10.7

The mean is the most sensitive of the measures of central tendency as it includes all of the scores/values in the data set within the calculation. This means it is more representative of the data as a whole.

However, the mean is easily distorted by extreme values. If we replace 17 in the data above with the number 98, the mean becomes 18.8 which does not really seem to represent the data overall!

Median : The middle value when data are ordered. For an even number of values, the median is the average of the two middle scores.

Example: Median = (10 + 11) ÷ 2 = 10.5

The strength of the median, unlike the mean, is that extreme scores do not affect it, so whether 98 replaces 17 in the data above or not, the median remains the same. It is also easy to calculate (once you have arranged the numbers in order). However, it is less sensitive than the mean as not all scores are included in the nal calculation.

Mode
The mode is the value that appears most frequently in a data set. Some data sets may have **two modes (bi-modal)** or no mode at all if all values are unique.

While calculating the mode is straightforward, it is considered a relatively simple measure and may not fully represent the overall data. For example, in the data set above, the mode is 7, which differs noticeably from both the mean and the median, and thus may not accurately reflect the central tendency.

For categorical data, however, the mode is often the only measure of central tendency that can be applied. For instance, if you asked your class to name their favourite dessert, identifying the most common choice (the mode) would be the best way to determine the "typical" preference.

Strengths & Limitations:

- Mean is sensitive to extreme values (outliers).
- Median is robust to outliers.
- Mode is simple but less representative of the data overall.

Apply it

Methods: And the results are in….
The table below shows the results of our energy drink experiment using a repeated measures. The score for each participant is the number of words said in the ve minutes after consuming each drink (participants were lmed and the number of words spoken after each drink were counted).

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
SpeedUpp	110	59	206	89	76	141	152	98	198	57
Water	122	45	135	90	42	87	131	113	129	62

Questions
1. Calculate the **mean**, **median** and **mode** for the *SpeedUpp* condition and the water condition above. Give all answers rounded up to the nearest whole number.

2. What can you conclude from these calculations?