NexGen Tutoring

21 Catherine St, Bentley WA 6102

Topic:

- **Data Representation and Interpretation – Measures of Centre and Surveying/Sampling**

Learning Intentions:

- Understand that mean (also called average) and median are different measures of centre for numerical data.
- Understand that the mean of a set of data can be affected significantly by an outlier, whereas the median is not affected.
- Know how to calculate the mean, median, and mode for a set of numerical data.
- Know the meaning of the terms population, sample, survey, census, symmetric, skewed, and bi-modal.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

Overview: Measures of Centre and Surveying/Sampling

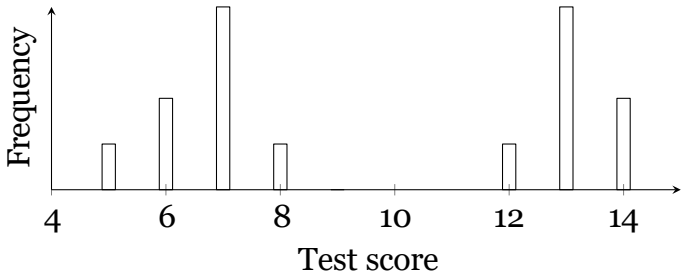
When we collect data (like test marks, survey answers, or daily steps), we often want a single number that represents the "middle" or "typical" value. These are called **measures of centre**.

Measures of Centre

- **Mean (average):** Add up all the numbers, then divide by how many there are. Example: If your scores are 8, 10, 12, the mean is $(8 + 10 + 12)/3 = 10$. The mean is very common, but it can be pulled up or down a lot by extreme values (called *outliers*).
- **Median:** Line up the numbers from smallest to largest. The median is the middle one. Example: For 5, 7, 9, the median is 7. If there are two middle numbers (e.g., 5, 7, 9, 11), take their average. The median is less affected by extreme values, so it's often more "fair" when there are outliers.
- **Mode:** The value that shows up the most. Example: If five students scored 12 and only one scored 15, the mode is 12. Data can be *unimodal* (one mode), *bi-modal* (two modes), or even have several modes.

  Bi-modal Example:

  Sometimes data has more than one "peak." For example, test scores may cluster into two groups: some students score low, others high, with fewer in the middle.



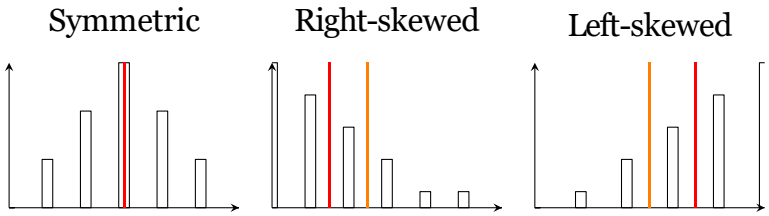
  Two clear peaks (around 7 and 13) show that this is a **bi-modal distribution**.

- **Skew:** How the data is "tilted."
  - *Symmetric:* Data is balanced; mean and median are about the same.
  - *Right-skewed:* There's a long "tail" on the right (big outliers); mean is larger than median.
  - *Left-skewed:* There's a long "tail" on the left (small outliers); mean is smaller than median.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

Visualising Skewness

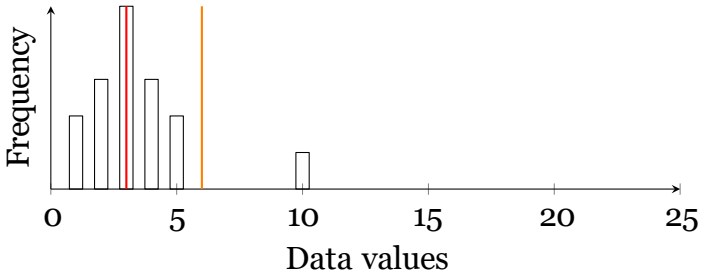
Data can "lean" in different directions. Look at the shapes below:



Notice: - In **symmetric** data, the mean and median line up. - In **right-skewed** data, the mean is "pulled" to the right. - In **left-skewed** data, the mean is "pulled" to the left.

Outliers and Their Effect on Mean and Median

Outliers are values that are much higher or lower than the rest of the data. These extreme values can significantly affect the mean but have little effect on the median.



In the graph above, there is an outlier at 10. Notice how the mean (orange line) is pulled right, while the median (red line) remains stable. This demonstrates how outliers influence the mean more significantly than the median.

Population, Sample, Survey, Census

When we study data, it is important to understand the context of where the data is coming from. Below are the key terms used in this process:

- **Population:** The entire group we care about (e.g., all Year 9 students in your school).
  - Example: The population of all students at a school if you want to know their average study time.
- **Sample:** A smaller group taken from the population to represent it (e.g., 30 randomly chosen students).
  - Example: A survey of 30 randomly selected students to estimate the average study time of all students.
- **Census:** A survey of the whole population (everyone is asked).

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

  - Example: A census where every student in the school is asked about their study habits.

- **Survey:** Collecting data from only a sample, then making conclusions about the population.
  - Example: Asking a sample of students about their favorite subjects and generalizing it to the entire school population.

In practice, conducting a census can be difficult and time-consuming, so often a representative sample is used to make reasonable estimates about the population.

Representativeness of Samples

For a sample to provide useful information, it must be **representative** of the entire population. A sample is considered representative if it accurately reflects the characteristics, behaviors, or qualities of the larger population. This can be achieved through random sampling and ensuring diversity in the sample.

Neutral vs. Leading Questions

The way a survey question is asked can influence the responses. Questions can be classified as either **neutral** or **leading**, and this distinction is crucial for ensuring unbiased results.

- **Neutral Question:** A neutral question does not influence the respondent's answer. It is unbiased and allows the respondent to provide their true opinion.

  Example:

  "What is your favorite type of music?"

- **Leading Question:** A leading question suggests a particular answer or influences the respondent in a certain direction, making it biased.

  Example:

  "Don't you agree that rock music is the best?"

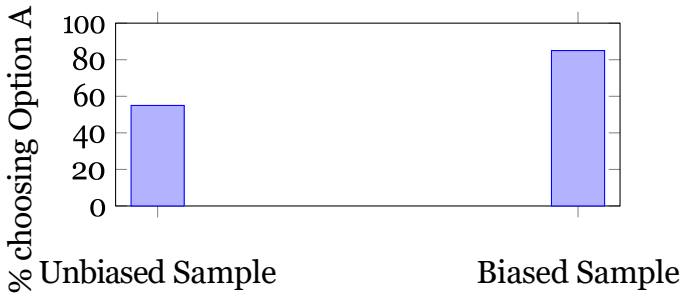
  A leading question like this one assumes that the respondent agrees with the statement, thus skewing the results.

  It is important to avoid leading questions in surveys to obtain honest and accurate data.

Sampling Bias Illustration

Bias happens when the sample doesn't really represent the population. For example, if a survey only includes students from a sports team, the results may be skewed.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102



Example: If you only ask your basketball teammates about "favourite sport," you'll likely get a biased result. A better way is to randomly sample students from all groups.

Worked Example: Interpreting Survey Results

A class of 30 students was asked where they like to study:

Location	Count
Home desk	12
Library	9
Kitchen table	6
Outdoors	3

How to read this:

- Home desk is most popular: 12/30 = 40%.
- Library is next (9 students = 30%).
- Together, they make up 70% of all answers.
- Outdoors is least common (10%).

Conclusion from this data: Most students prefer quiet, indoor places to study. When you interpret survey data, always report both the *counts* and the *percentages* so it's easier to compare.

When we collect data (like test marks, survey answers, or daily steps), we often want a single number that represents the "middle" or "typical" value. These are called **measures of centre**.

Creating Datasets with Specific Properties

In data analysis, sometimes you need to create datasets that satisfy specific conditions like a given mean, median, or modality. Below are examples of how to create datasets that meet certain requirements:

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

1. Dataset with a Specific Mean and Median

To create a dataset where the mean and median are predefined:

- First, calculate the total sum of the dataset using the formula for the mean:

$$\text{Sum} = \text{Mean} \times \text{Number of Values}.$$

- Ensure the **median** is positioned correctly in the dataset by sorting the values. For an odd number of values, the median is the middle value; for an even number, it is the average of the two middle values.
- Choose values that add up to the sum calculated and place values around the median value to achieve the desired result.

Example: Create a dataset with 8 values where the mean = 10 and the median = 9. - The sum of the dataset must be $10 \times 8 = 80$. - The values must be chosen such that the 4th and 5th values are both 9 to make the median 9.
Dataset:

$$\{6, 7, 8, 9, 9, 10, 12, 13\}$$

- **Mean:** $\frac{6+7+8+9+9+10+12+13}{8} = 10$ - **Median:** The middle two values are both 9, so the median is $\frac{9+9}{2} = 9$.

2. Bi-modal Dataset

A bi-modal dataset contains two values that appear most frequently. To create a bi-modal dataset:

- Select two values to be the modes. These should appear more frequently than other values in the dataset.
- Ensure the **median** is placed in the correct position based on the ordered dataset.

Example: Create a dataset with 9 values that is bi-modal. The modes are 7 and 8, and the median is 7.
Dataset:

$$\{4, 5, 6, 7, 7, 8, 8, 10, 11\}$$

- **Modes:** 7 and 8 (both appear twice). - **Median:** The middle value is 7.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

Section A: Multiple Choice

1. Which measure is most *affected* by an outlier?
   a. Median b. Mode
   c. Mean d. Range

2. For symmetric data, which statement is usually true?
   a. Mean < Median b. Mean ≈ Median
   c. Mean > Median d. Mean = Mode always

3. A dataset has two equally common values. It is:
   a. Unimodal b. Bi-modal
   c. Left-skewed d. Uniform

4. Which method is most likely to produce a *representative* sample?
   a. Ask the first 20 students you meet b. Ask volunteers from your class
   c. Randomly select 30 students from all Year 8 rolls d. Ask only your friends

5. A survey of all students in the school is a:
   a. Sample b. Census
   c. Convenience sample d. Biased sample

6. In right-skewed data, typically:
   a. Mean < Median b. Mean ≈ Median
   c. Mean > Median d. Mean = Mode

7. Which question is most *neutral*?
   a. "Don't you agree that homework takes too long?"
   b. "How many minutes of homework do you do on a school night?"
   c. "Homework is bad, right?"
   d. "You never do more than 20 minutes, do you?"

8. Which is *not* a source of bias?
   a. Leading questions b. Low response rate
   c. Random sampling d. Convenience sampling

9. The mode is most useful when:
   a. Data are skewed b. You need a typical *category*
   c. Outliers dominate d. You need a resistant measure

10. The median of 8, 9, 11, 12 is:
    a. 9 b. 10
    c. 10.5 d. 11

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

Section B: Short Answer

11. Compute the **mean**, **median**, and **mode** for:
6, 7, 7, 8, 10.

12. Compute the **mean** and **median** for:
12, 12, 13, 13, 14, 90.
Which measure is more appropriate to describe the centre? Justify briefly.

13. The set 3, 5, 6, 9, x has a median of 6. What is the possible range of values for x if the data is ordered?

14. The mean of five numbers is 8. Four of the numbers are 5, 7, 9, 12. Find the value of the fifth number.

15. The mean of n values is m. If an extra value equal to the current mean is added, what is the new mean? Explain briefly.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

16. A dataset is left-skewed. Which is more likely?
    a. mean < median b. mean = median c. mean > median
    Explain your choice in one sentence.

17. A sample should be representative. Suggest one practical way to ensure the sample is representative when surveying Year 8 students.

18. Provide one example of a **leading question** and one example of a **neutral question** on the topic "screen time."

19. A shop owner surveys customers only on Saturday afternoon to estimate average spending. Identify one potential bias and briefly explain how to reduce it.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

Section C: Applications & Interpretation

20. The heights (cm) of 12 plants:
22, 25, 24, 21, 26, 25, 120, 24, 23, 25, 26, 24.

1. Calculate the mean and median.

2. Which measure better represents a typical height? Explain referencing the data.

21. Two classes took a 20-mark quiz. Summary:

	Median	Mean
Class A	14	14
Class B	12	15

1. Which class likely has more high outliers? Explain.

2. Which class is more symmetric? Why?

22. Data: 4, 4, 5, 5, 5, 6, 7, 9.

1. Find mean, median, mode.

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

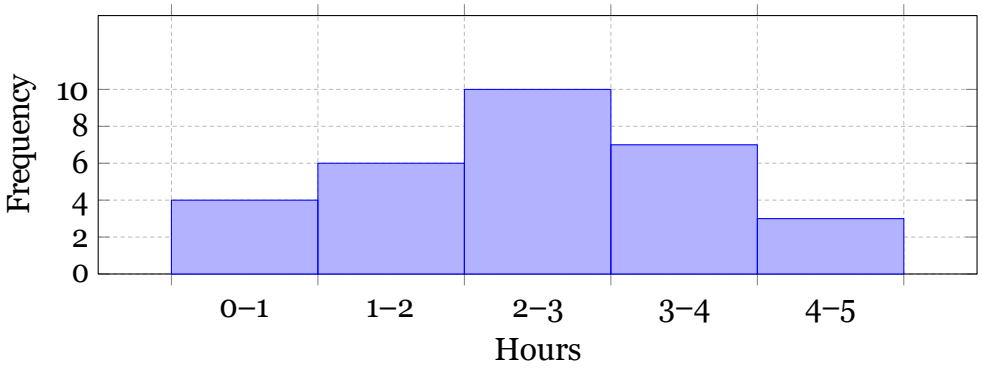
2. If 12 is added, state the new mean and comment on the median.

23. A survey asks: "How much time do you *waste* on homework each night?"

1. Identify the bias in the wording.

2. Rewrite it to be neutral.

24. **Chart reading (Histogram).** The histogram below summarises daily screen times (hours) for a class.



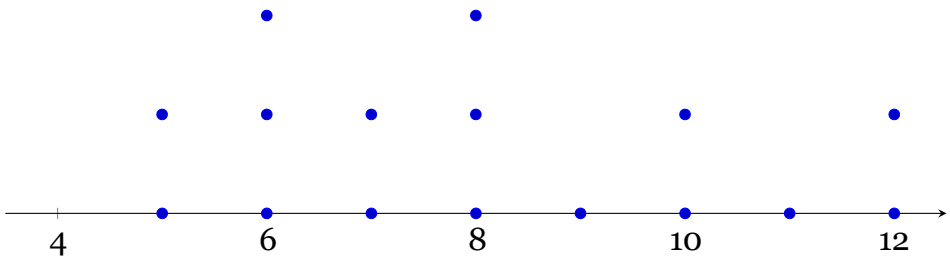
1. What is the modal class?

2. Is the distribution roughly symmetric, left-skewed, or right-skewed?

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102

3. Which measure of centre (mean/median) is more appropriate here? Why?

25. **Chart reading (Dot plot).** Dot plot of daily steps (thousands) for 16 students:



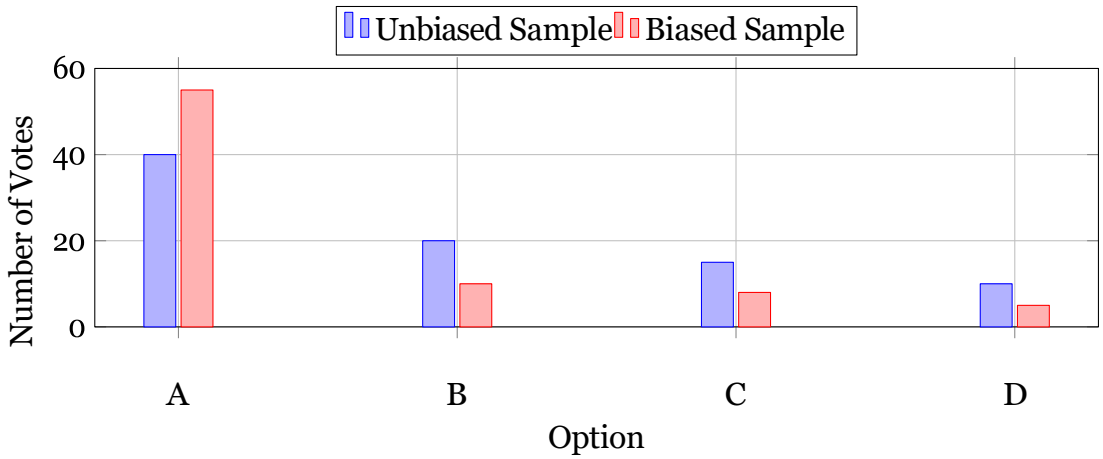
1. Estimate the median steps (in thousands).

2. State the mode(s).

3. Comment on skewness, if any.

26. **Chart reading (Bar chart and bias).** Preferences for lunch option from two surveys:

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102



1. Which survey likely used convenience sampling? Explain.

2. If the school must choose fairly, which survey should be trusted more and why?

info.nexgentutoring@gmail.com
+61 468 536 435
Tutoring.NexGen
21 Catherine St,
Bentley WA 6102