WORKSHEET # 015 | Australian Curriculum Year 8 Mathematics | WM9T4U7

ACACIA TUTORING

Data Representation and Interpretation 8

Free and always will be!

Focus: A set of questions and solutions for Year 8 students focused on 'Data Analysis' under the "Statistics and Probability" strand, tailored to the Australian Curriculum:

1. Understanding Data Analysis:

a) Define what data analysis is and why it is important.

Solution:
Data analysis involves collecting, organising, summarising, presenting, and interpreting data to draw conclusions or make informed decisions. It's important because it helps in understanding trends, making predictions, and solving problems in various contexts.

b) What are the different types of data (qualitative vs. quantitative)?

Solution:
Qualitative Data: Descriptive data that describes qualities or characteristics (e.g., colour, type).
Quantitative Data: Numerical data that can be measured or counted (e.g., height, number of students).

2. Data Collection Methods:

Describe two methods of collecting data.

Solution:
Surveys/Questionnaires: Collecting information directly from individuals through questions.
Observation: Gathering data by watching and recording behaviours or events as they occur.

3. Organising Data:

a) What are the common ways to organise data for analysis?

Solution:

Tables: For raw data or organised sets.
Frequency Tables: To show how often each value occurs.
Tally Charts: For counting occurrences in real-time.

b) Organise the following data into a frequency table: $\{ 2, 3, 3, 4, 2, 5, 3, 2, 4 \}$.

Solution:

Write the 'Number' column first, filling in the different numbers in the set. Then, as you are counting the amount of times each number occurs, cross off each number as it is counted and add the total to the 'Frequency' column. (Crossing off reduces the chance of error / accidentally counting the same number twice, and it makes it easier to see the numbers left over (still to count)).

E.g. after you have counted the twos:
$\{ \not{2}, 3, 3, 4, \not{2}, 5, 3, \not{2}, 4 \}$.

Number	Frequency
2	3
3	3
4	2
5	1

4. Data Representation:

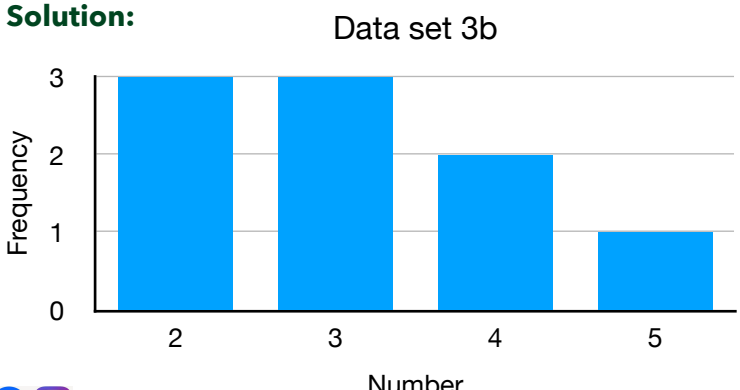
a) What are three common types of graphs used in data analysis?

Solution:

Bar Graphs: Good for comparing quantities.
Histograms: Used for continuous data to show distribution.
Pie Charts: For showing proportions or percentages of a whole.

b) Create a bar graph for the frequency table from question 3b.

Solution:



Description: The x-axis would have numbers $2, 3, 4, 5$, and the y-axis would count frequency. Bars would be drawn to represent the frequency of each number.

5. Measures of Central Tendency:

a) Define mean, median, and mode.

Solution:

Mean: The arithmetic average of a set of numbers, calculated by summing all values and dividing by the count.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \text{Mean} = \frac{\text{Sum of scores}}{\text{Number of Scores}}$$

Median: The middle value in an ordered list of data. If there's two middle numbers, add them together and divide by two. Order the dataset from lowest to highest, and the numbers (or pair of numbers), that are in the middle, are the median.

Mode: The value that appears most frequently in the data set.

b) Calculate the mean, median, and mode for the dataset: $\{ 5, 7, 7, 8, 9, 10 \}$ **.**

Solution:

Mean:

$$\text{Mean} = \frac{\sum \text{Scores}}{\text{Number of Scores}}$$

$\sum$ (Sum), means, add up all the numbers.

$$= \frac{5 + 7 + 7 + 8 + 9 + 10}{6}$$

$$= \frac{46}{6}$$

$$\approx 7.67 .$$

Median:

First, put the numbers in ascending order: $5, 7, 7, 8, 9, 10$,

Then, cross off numbers in pairs (on either side of the set),

until you reach the middle number (or middle numbers):

$$\cancel{5}, \cancel{7}, \boxed{7, 8}, \cancel{9}, \cancel{10} .$$

$$\text{Median} = \frac{7 + 8}{2}$$

$$= 7.5 .$$

Mode:

7 (appears twice, more than any other number).

c) Calculate the mean, median, and mode of the following dataset: $\{ 4, 8, 6, 5, 3, 4 \}$ **.**

Solution:

Mean:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{Mean} = \frac{\text{Sum of scores}}{\text{Number of Scores}}$$

$$= \frac{4 + 8 + 6 + 5 + 3 + 4}{6}$$

$$= \frac{30}{6}$$

$$= 5 .$$

Median:

Ordered: $3, 4, 4, 5, 6, 8 ,$

$$\cancel{3}, \cancel{4}, \boxed{4, 5,} \cancel{6}, \cancel{8} .$$

$$\text{Median} = \frac{4 + 5}{2}$$

$$= 4.5 .$$

Mode:

4 (appears twice, more than any other number).

6. Measures of Spread:

a) What is the range of a dataset?

Solution:

The range is the difference between the highest and lowest values in the data set.

b) Find the range for the dataset in question 5b.

Solution:

$$\text{Range} = 10 \text{ (highest)} - 5 \text{ (lowest)}$$

$$= 5 .$$

a) Define 'spread' in data analysis. Why is it important alongside measures of central tendency?

Solution:

Spread describes how data is dispersed or distributed around the central tendency. It's important because central tendency alone doesn't show how variable the data is; spread gives context to the average or typical value.

c) What are the common measures of spread?

Solution:

Range:

The difference between the highest and lowest values.

Range = Higest − Lowest

Interquartile Range (IQR):

The spread of the middle $50\,\%$ of data, calculated as (see Question 6e. for example):

$IQR = Q3 - Q1$.

Variance and Standard Deviation:

Measures of how spread out the numbers in a dataset are.

d) Find the range for the dataset from question 2a.

Solution:

Ordered $\{ 3, 4, 4, 5, 6, 8 \}$:

Range $= 8$ (highest) $- 3$ (lowest),

$= 5$.

e) Calculate the interquartile range for this dataset: $\{ 2, 5, 7, 9, 12, 15 \}$.

Solution:

To find quartiles, firstly, break the dataset into two even parts.

$\{ 2, 5, 7, \} \{ 9, 12, 15 \}$, then:

Q1 (First Quartile), is given by the median of the lower half:

$\{ 2, 5, 7 \} = 5$,

Q3 (Third Quartile), is given by the median of the upper half:

$\{ 9, 12, 15 \} = 12$,

$$IQR = Q3 - Q1$$
$$= 12 - 5$$
$$= 7.$$

f) Calculate the interquartile range for this dataset: $\{ 3, 6, 7, 9, 11, 13, 15 \}$.

Solution:
To find quartiles, firstly, break the dataset into two even parts.

$\{ 3, 6, 7, \} \ 9, \{ 11, 13, 15 \}$, then:

Q1 (First Quartile), is given by the median of the lower half:
$\{ 3, 6, 7 \} = 6$,

Q3 (Third Quartile), is given by the median of the upper half:
$\{ 11, 13, 15 \} = 13$,

$$\begin{aligned} IQR &= Q3 - Q1 \\ &= 13 - 6 \\ &= 7. \end{aligned}$$

Note: (Quartile 2), in this case, is 9 (the median of the original dataset)

7. Interpretation of Data:

If the mean height of students in a class is $160\,cm$ but the median is $155\,cm$, what might this suggest about the distribution of heights?

Solution:
It suggests there might be a few very tall students pulling the mean up, indicating a skew towards higher values.

8. Practical Application:

A teacher wants to analyse the scores of a test where scores range from 0 to 100. After plotting a histogram, she notices most scores are between 60 and 80. What can she infer from this?

Solution:
She can infer that the majority of students performed moderately well, with a concentration of scores in the $60 - 80$ range, possibly indicating a moderate level of difficulty for the test or good understanding of the material by most students.

9. Misleading Data:

How can data representation be misleading? Give an example.

Solution:
Data can be misleading through:
Scale manipulation: Changing the scale on a graph to exaggerate or minimise differences.

Example: A bar graph with a truncated y-axis might make small differences in data look much larger than they are.

10. Understanding Central Tendency:

a) What does 'central tendency' mean in statistics?

Solution:
Central tendency refers to the central or typical value for a set of data, providing a single value that represents the whole set. There are three measures of central tendency:

Mean (typically referred to as average),
Median (the middle number), and
Mode (the most common number).

b) If a dataset has a mean of 50 and a large standard deviation, what does this tell you about the data?

Solution:
It suggests that while the average value is 50, the data points are widely spread out from this mean, indicating high variability among the values.

c) What implications would a small range and a median close to the mean have on data interpretation?

Solution:
This would imply that the data is tightly clustered around the central value, indicating low variability and a more consistent dataset.

11. Comparing Datasets:

Two classes took the same test. Class A has a mean score of 75 with a standard deviation of 10, while Class B has a mean of 70 with a standard deviation of 5. Comment on the performance and variability of each class.

Solution:
Class A: Higher average score but with more variability, suggesting some students scored much higher or lower than 75.

Class B: Lower average but with less variability, indicating scores are more consistent around 70, with less extreme highs or lows.

12. Practical Application:

A teacher recorded the heights of students in $centimetres$: $\{ 150, 155, 160, 160, 165, 170, 175 \}$. Calculate and interpret the mean height and the range.

Solution:

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{\text{Sum of scores}}{\text{Number of Scores}}$$

$$= \frac{150 + 155 + 160 + 160 + 165 + 170 + 175}{7}$$

$$= \frac{1135}{7}$$

$$\approx 162.14\, cm\ .$$

This indicates the average height of the students.

$$\text{Range} = 175\ (\text{highest}) - 150\ (\text{lowest})$$

$$= 25\, cm\ .$$

This shows the spread from the shortest to the tallest student is quite significant.

13. Outliers and Their Effects:

How might an outlier affect the mean and median of a dataset?

Solution:

An outlier can significantly increase or decrease the mean, pulling it towards its value. The median, being less sensitive to outliers, might remain unchanged or change only slightly if the dataset is large enough.

Additional Notes for Teachers:

Learning Outcomes: Students should be adept at collecting, organising, representing, and analysing data, understanding both basic statistics and the potential for misrepresentation.

Teaching Strategies: Use real or collected class data for hands-on activities in data analysis. Encourage critical thinking about how data is presented and could be manipulated. Incorporate technology for dynamic data visualisation and analysis.

Assessment: Assess through tasks where students must analyse datasets, interpret graphs, or create their own data representations to answer specific questions.

Resources: Spreadsheet software for data entry and analysis, graph paper for manual plotting, or online tools for interactive data exploration.

This question set aligns with the Australian Curriculum for Year 8, focusing on the proficiencies of understanding, fluency, problem-solving, and reasoning in statistics and probability, specifically in the context of data analysis.

IMPORTANT: At Acacia Tutoring we believe all educational resources should be free, as education, is a fundamental human right and a cornerstone of an equitable society. By removing financial barriers, we ensure that all students, regardless of their socioeconomic background, have equal access to high-quality learning materials. This inclusivity promotes fairness, helps bridge achievement gaps, and fosters a society where every individual can reach their full potential.

Furthermore, free resources empower teachers and parents, providing them with tools to support diverse learners and improve outcomes across communities. Education benefits everyone, and making resources universally accessible ensures we build a more informed, skilled, and prosperous future for all.

All documents are formatted as a **.pdf** file, and are completely **FREE** to use, print and distribute - as long as they are not sold or reproduced to make a profit.

ACCAIA
PER REVERENTIA INTEGRITAS SCIENTIA ET MISERICORDIA
Through Respect Integrity Knowledge and Compassion

N.B. Although we try our best to produce high-quality, accurate and precise materials, we at Acacia Tutoring are still human, these documents may contain errors or omissions, if you find any and wish to help, please contact Jason at info@acaciatutoring.com.au .