ACACIA TUTORING

Data Representation and Interpretation 9

Free and always will be!

Focus: A set of questions and solutions for Year 9 students on Data Representation and Interpretation, tailored to the Australian Curriculum under the strand 'Statistics and Probability':

1. Understanding Time Series:

a) Define what time series data is and give an example.

b) List three characteristics of time series data.

2. Analysing Time Series:

a) Given the following monthly temperatures for Brisbane in 2025 (in $^{\circ}C$) :

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temp	37	35	32	28	24	21	20	22	25	30	35	38

Identify any trend, seasonality, or anomalies in this data. (Graph first if needed, Q2b).

b) How would you graph this time series data to visualise trends and seasonal patterns?

3. Forecasting with Time Series:

a) If the data shows a consistent increase of $2^{\circ}C$ per month from July to December, predict the temperature for January.

b) Discuss how moving averages can be used to forecast future data points in a time series.

4. Practical Application:

a) A company has recorded its sales (in thousands) for the last 6 months as follows:
$\{ \ 12, \ 15, \ 18, \ 20, \ 22, \ 23 \ \}$ **. Predict the sales for the next month assuming linear growth.**

b) Explain how businesses might use time series data for decision-making.

5. Data Smoothing:

a) Describe how a moving average can be used to smooth time series data.

b) What might be the drawbacks of using moving averages?

6. Constructing a Box Plot:

a) Here are the test scores of 10 **students :**$\{\ 65,\ 78,\ 85,\ 90,\ 88,\ 72,\ 95,\ 82,\ 79,\ 84\ \}$ **. Order the data, construct a box plot for this data, and calculate the interquartile range.**

b) Explain what each part of a box plot represents.

7. Interpreting Box Plots:

a) Given two box plots for the test scores of two different classes, how can you compare the spread and central tendency of the scores?

b) What does it mean if one box plot has a longer whisker on one side compared to the other?

8. Box Plots with Outliers:

a) If the test scores from question 6a now include an outlier of 46, how would you modify the solution?

b) How does the presence of outliers affect the interpretation of data using a box plot?

Solutions

1a.

Time series data is a sequence of data points collected or recorded at successive, equally spaced points in time. An example could be the monthly sales figures for a retail store over a year.

b.

Trend: The long-term movement in the data, whether it's upward, downward, or stable.
Seasonality: Regular patterns that repeat over known, fixed periods (e.g., daily, weekly, annually).
Cyclical Variations: Fluctuations occurring at irregular intervals, often influenced by economic or business cycles.

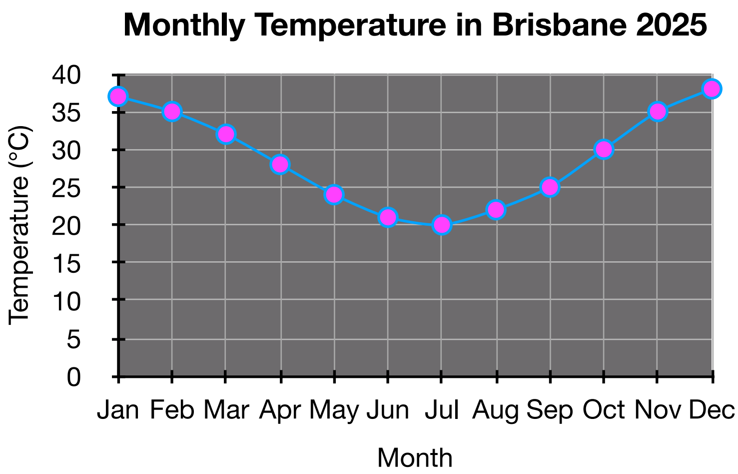
2a.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temp	37	35	32	28	24	21	20	22	25	30	35	38

Trend: There's a decreasing trend from January to July, then an increasing trend from July to December, which indicates an annual cycle.
Seasonality: There's clear seasonality with temperatures peaking in December – January and dipping in winter months (June – July), suggesting a seasonal pattern related to the year's climate cycle.
Anomalies: No clear anomalies are visible from this data, but any unexpected temperature drop or spike outside the seasonal norm would be considered an anomaly.

b.

Use a **line graph** with: Month on the x-axis. Temperature on the y-axis.
Plot each point for each month, connecting them with lines to show the progression of temperature over time. Consider adding a moving average line to smooth out short-term fluctuations and highlight the trend more clearly.



3a.

If the trend continues, January's temperature would be:
December's temperature plus two degrees,
$\rightarrow (38^\circ C) + 2^\circ C$
$= 40^\circ C.$

b.

A moving average smooths out short-term fluctuations and highlights longer-term trends or cycles. For forecasting: Calculate a moving average (e.g., a 3-month or 6-month moving average) to identify the trend. Project the trend forward by assuming the rate of change observed in the moving average will continue. For example, if the 3-month moving average shows an increase of $1^{\circ}C$ each period, you might forecast a similar increase for the next period.

4a.

As the growth is stated to be linear, with an increase averaging 2 thousand each month, given by:

$$\{\ 12,\ 15,\ 18,\ 20,\ 22,\ 23\ \}$$

$\rightarrow$ Difference between each number in the set : $3, 3, 2, 2, 1$

Find the average i.e. add them, then divide by 5 :

$$\rightarrow \frac{3+3+2+2+1}{5}$$
$$= \frac{11}{5}$$
$$= 2.2$$
$$\approx 2\ thousand\ sales\ per\ month\,.$$

Predicted sales for next month = this month plus average difference:

$= 23 + 2$

$= 25$ thousand.

b.

Inventory Management: Predicting demand patterns to stock appropriate levels of goods.
Marketing Strategies: Timing promotions or product launches based on seasonal trends or dips in sales.
Financial Planning: Forecasting revenue for budgeting, setting sales targets, or planning expansions.
Resource Allocation: Scheduling staff or production based on expected busy or slow periods.

5a.

A moving average is calculated by taking the average of a set number of data points (called the window) and moving this window over the data series: For a 3-month moving average, you'd average the first three months, then move one month forward and average the next three, and so on.This process reduces noise and highlights the underlying pattern or trend in the data.

b.

Lag: Moving averages are inherently lagging indicators because they use past data, potentially missing rapid changes in trends.
Boundary Effects: At the start and end of the time series, fewer data points are available for averaging, which can distort the trend.
Loss of Detail: Smoothing can mask short-term fluctuations that might be significant for certain analyses.

6a.

Order the data: $\{65, 72, 78, 79, 82, 84, 85, 88, 90, 95\}$

Identify the minimum value : Min $= 65$

Find Q_1 (First Quartile) : Median of the lower half $(65, 72, 78, 79, 82)$ $Q_1 = 78$

Find the median (Q_2) : Between 82 and 84 $= \dfrac{82+84}{2}$ Median $= Q_2 = 83$

Find Q_3 (Third Quartile) : Median of the upper half $(84, 85, 88, 90, 95)$ $Q_3 = 88$

Identify the maximum value : Max $= 95$

Calculate the Interquartile Range (IQR) : $IQR = Q3 - Q1 = 88 - 78$ IQR $= 10$

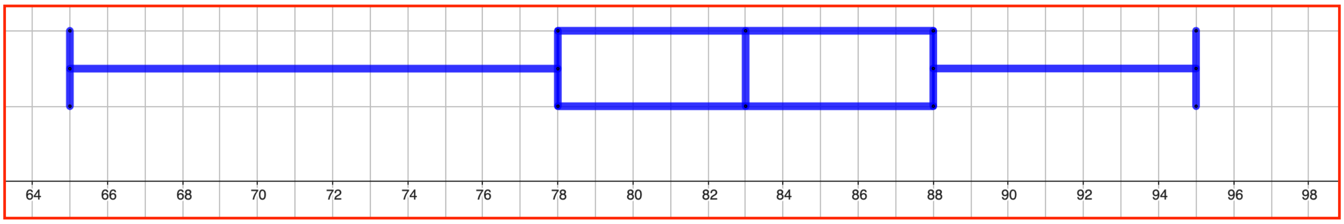
Drawing the Box Plot:

Draw a number line from 65 to 95 .

Plot points for: $Min,\ Q_1,\ Median,\ Q_3,$ and Max .

Draw a box from Q_1 to Q_3 with a line at the median.

Draw whiskers from the box to the minimum and maximum values.



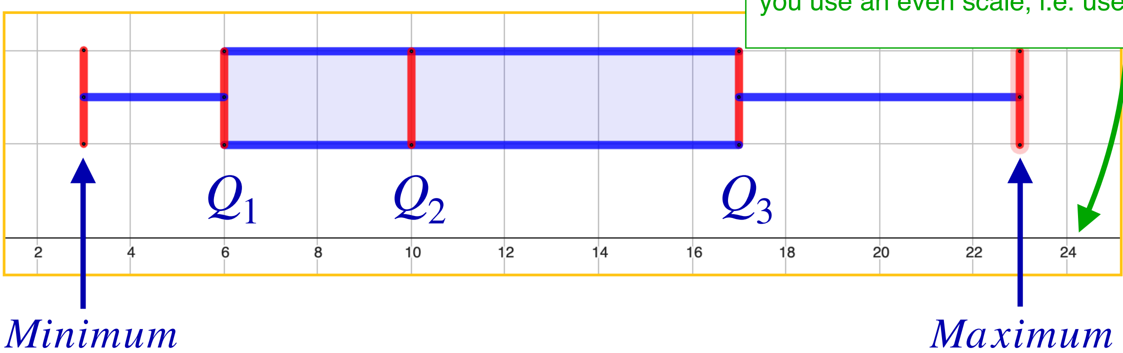
b.

A box plot is a visual representation of a five number summary.

Minimum Value:	The smallest data point excluding outliers.
First Quartile (Q_1) :	The median of the lower half of the data.
Median (Q_2) :	The middle value when data is ordered (median of whole data set).
Third Quartile (Q_3) :	The median of the upper half of the data.
Maximum Value:	The largest data point excluding outliers.

Box-and-Whisker Plot

Important: When drawing box-and-whisker plots, ensure you use an even scale, i.e. use



7a.

Central Tendency: Compare the medians. The class with the higher median has a higher central score.

Spread: **IQR:** The class with the larger IQR has a wider spread in the middle 50% of its scores, indicating more variability in the typical scores.
Whisker Length: Longer whiskers suggest a wider total range of scores.
Symmetry and Skewness: If the median is not centred in the box, it indicates skewness. If the median is closer to Q_1 , it's positively skewed; if closer to Q_3 , negatively skewed.

b.

A longer whisker on one side indicates that there's more variability in the data on that side of the median. For instance, a longer upper whisker means there are higher scores that are more spread out, potentially indicating some students performed much better than the rest.

8a.

Order the new data: $\{46, 65, 72, 78, 79, 82, 84, 85, 88, 90, 95\}$

Identify the minimum value : Min $= 46$
Find Q_1 (First Quartile) : Median of the lower half $(46, 65, 72, 78, 79)$ $Q_1 = 72$
Find the median (Q_2) : Middle number of the set $= Q_2 = 82$
Find Q_3 (Third Quartile) : Median of the upper half $(84, 85, 88, 90, 95)$ $Q_3 = 88\ (unchanged)$
Identify the maximum value : Max $= 95$

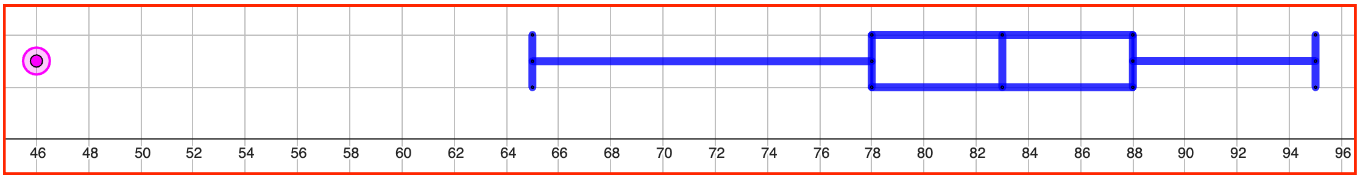
Calculate the Interquartile Range (IQR) : $IQR = Q3 - Q1 = 88 - 72$ IQR $= 16$

Determine Outliers:

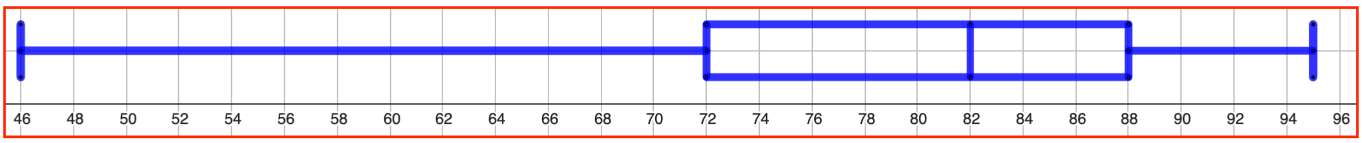
Anything,
below $Q_1 - 1.5 \times IQR$ ($72 - 1.5 \times 16 = 72 - 24 = 48$) or,
above $Q_3 + 1.5 \times IQR$ ($88 + 1.5 \times 16 = 88 + 24 = 112$) ,
is considered an outlier.

Here, 46 is an outlier.

Modified Box Plot:
Draw the box plot as before however :
Mark 46 as an outlier point below the 65 whisker.



Without modification of data (i.e. without removing outlier) :



It can be seen, the outlier skews the data to the left significantly (negative skew), i.e. the left hand side of the box plot drops down and the middle is slightly affected.

b.

Skewness: Outliers can make the distribution look skewed, affecting how we perceive the central tendency or spread.
Variability: An outlier increases the perceived spread of the data, making the dataset appear more variable than it might be without the outlier.
Central Tendency: Outliers do not affect the median but can significantly impact the mean, thus changing how we might interpret the 'average' score if considering both.

Additional Notes for Teachers:

Learning Outcomes:

Students should understand how to analyse, interpret, and forecast from time series data, recognising patterns like trends and seasonality. Students should be able to construct, interpret, and compare box plots, understanding how they represent data distribution, including the impact of outliers.

Teaching Strategies:

Use real or simulated data sets for practical exercises. Encourage students to graph data by hand or using software to visualise patterns. Use real or simulated data for students to practice creating box plots. Discuss the implications of outliers in various contexts like sports scores or test results.

Assessment:

Assess through tasks where students must analyse time series data, predict future values, and explain their reasoning based on observed patterns. Provide datasets for students to create box plots and ask questions about comparison, spread, and the influence of outliers.

Resources:

Excel or other spreadsheet software for plotting and analysing time series data, online datasets for real-world examples, or educational time series analysis tools. Graph paper for manual drawing, statistical software for digital creation of box plots, or online tools for interactive learning about data representation.

This set of questions aligns with the Australian Curriculum for Year 9, focusing on understanding and applying concepts related to data representation and interpretation in statistics and probability.