General Mathematics Units 3 & 4

Chapter 5 — Investigating and modelling time series

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Attribution: Classbasics Pty Ltd, vwx.au

Aboriginal and Torres Strait Islander content has been excluded as accuracy cannot be guaranteed, and it is better to be respectful than cover the entire curriculum inaccurately.

Significant effort has been made to ensure this content is legal. Please send any areas of concern to admin@vwx.au with appropriate document/legal references so errors can be verified and changes be made.

Contents

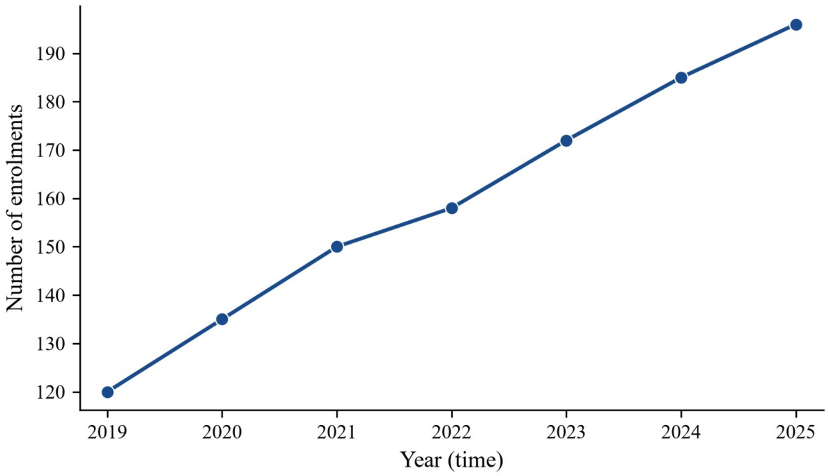
Theory

A **time series** is a numerical variable that is recorded at **successive times**, usually at equally spaced intervals — every day, week, month, quarter or year. The monthly rainfall at a weather station, a company's quarterly sales, and a country's annual population are all time series. The defining feature is *order in time*: each value belongs to a particular moment, and the sequence in which the values occur is part of the information.

When we study a time series, **time is the explanatory variable** and the recorded quantity is the response variable. Because time drives the recording, it is always placed on the horizontal axis.

The time series plot.

A **time series plot** displays each recorded value against its time. The points are plotted as (time, value) and then **joined by straight line segments**, in time order. Joining the points is what separates a time series plot from an ordinary scatterplot: the line emphasises the movement of the variable *from one time to the next* and makes patterns over time easy to see. Reading a time series plot means moving left to right along the time axis and reading the height of the graph at each time.

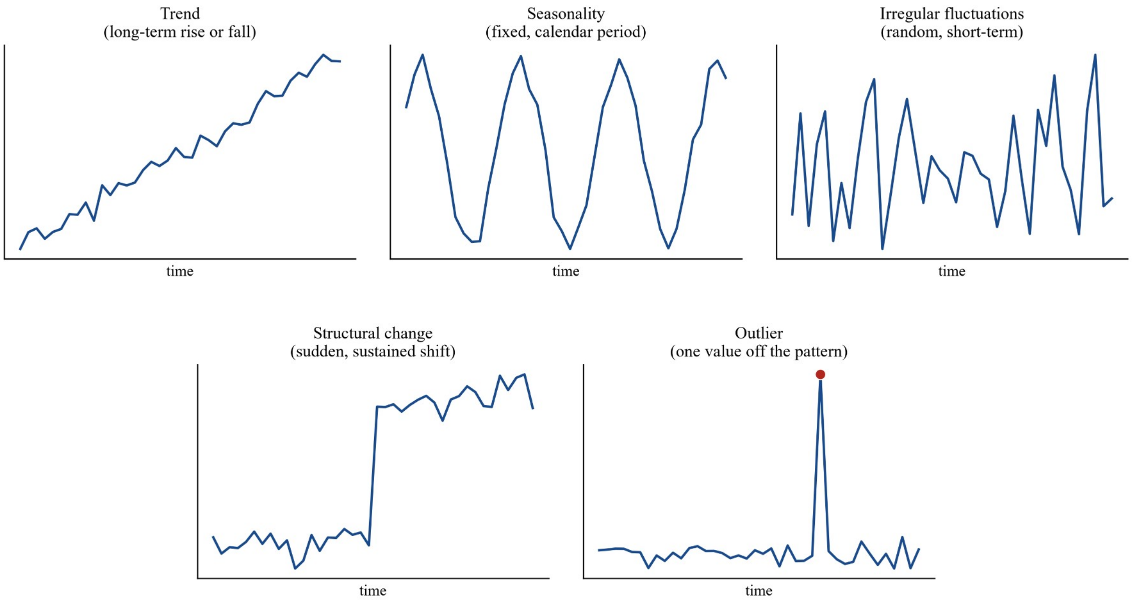


Describing a time series.

Once a time series is plotted, we describe its **qualitative features**. A real series may show several of the following at once, so a good description names each feature that is present and always interprets it **in the context of the data**.

- **Trend** is the long-term upward or downward movement of the series across the whole period of observation. A series may show an **increasing trend**, a **decreasing trend**, or **no trend**. Trend is about the general direction over the long run, not the small ups and downs from one time to the next.
- **Seasonality** is a **systematic, calendar-related** pattern that repeats over a **fixed, known period** of one year or less. Ice-cream sales that peak every summer and dip every winter, or electricity use that rises each weekday and falls at the weekend, are seasonal. The **period** of the seasonality is the number of observations in one complete cycle: for **monthly** data the period is **12**, for **quarterly** data it is **4**, and for **daily** data with a weekly pattern it is **7**. A rise and fall that has no fixed length — the multi-year booms and downturns of an economy, say — is *not* seasonal, because it is not tied to a fixed calendar period.
- **Irregular (random) fluctuations** are the unsystematic, short-term ups and downs that remain once trend and seasonality are accounted for. They are unpredictable and have no repeating pattern.
- **Structural change** is a **sudden, sustained change** in the behaviour of the series — an abrupt shift in its level or a change of direction that persists. It shows up as a discontinuity in the plot, often caused by a one-off event that permanently alters the situation, such as a new factory opening or a change in the law.
- **Outliers** are individual values that lie **well away from the surrounding pattern**. An outlier is usually caused by a **one-off event** — a strike, a natural disaster, a public holiday sale — and, unlike a structural change, it affects only a single time before the series returns to its earlier behaviour.

The five panels below show each feature on its own.



Reading features in context.

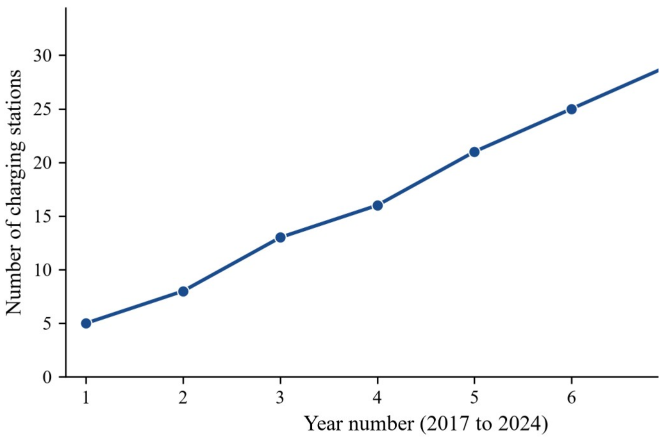
Describing a time series is not just naming features — it is saying *what they mean for the variable being measured*. Rather than “there is an increasing trend”, write “the number of subscribers shows an increasing trend over the ten years”. Rather than “there is seasonality”, write “sales are seasonal, peaking in the summer quarter each year”. Always refer to the actual variable and the actual times.

Using CAS. General Mathematics is technology-active. In practice you enter the time and the recorded values into your CAS calculator’s statistics editor and have it draw the time series plot for you. The by-hand skill you must master is *reading and describing* the resulting plot — identifying the trend, any seasonality, irregular fluctuations, structural change and outliers, and interpreting them in context.

Worked examples

Example 1 — scaffolded

The number of public electric-vehicle charging stations in a town, recorded at the end of each year from 2017 to 2024, was 5, 8, 13, 16, 21, 25, 29, 33. Construct a time series plot and describe the trend.

Working	Comment
Plot the points $(2017,5),(2018,8),\ldots,(2024,33)$ with year on the horizontal axis and join them in order.	Time (year) is the explanatory variable, so it goes on the horizontal axis; the points are joined by line segments.
	Reading left to right, the graph climbs steadily throughout.
The series rises from 5 stations in 2017 to 33 in 2024, an increase of $33-5=28$ over the 7 years.	An overall rise ⇒ an **increasing trend**.

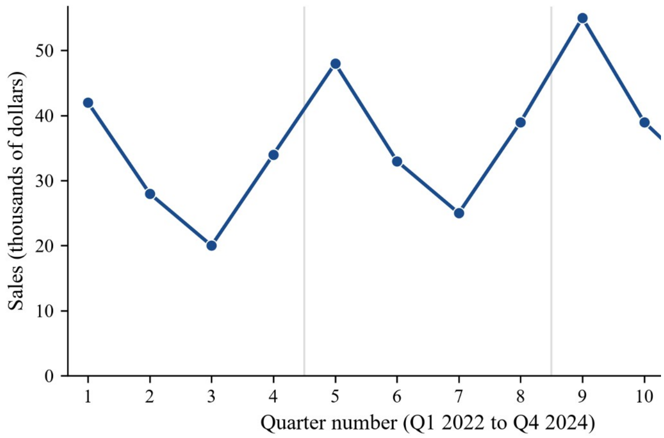
Working	Comment
Average increase $=\dfrac{28}{7}=4$ stations per year.	Total change divided by the number of yearly intervals.

The plot shows a clear **increasing trend**: the number of charging stations grew by about 4 per year, from 5 in 2017 to 33 in 2024.

Example 2 — scaffolded

The table shows a swimwear retailer's quarterly sales (in thousands of dollars) over three years. Plot the series and describe its features.

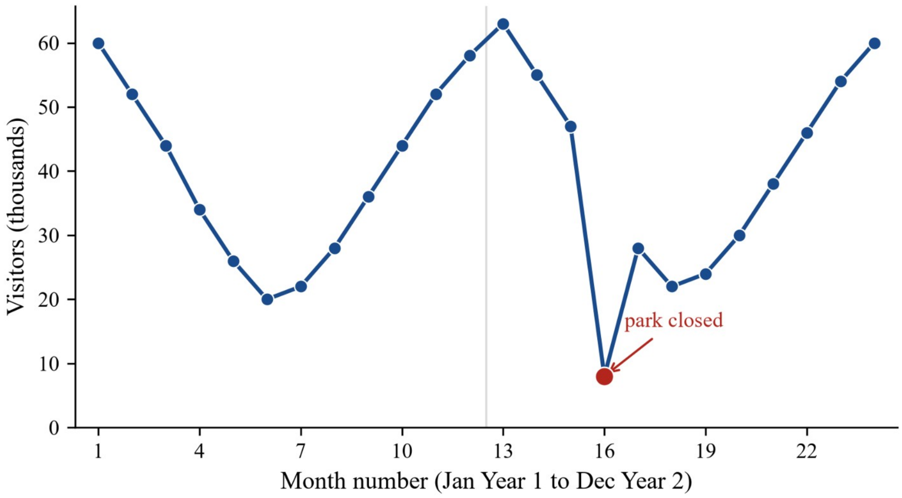
Year	Q1	Q2	Q3	Q4
2022	42	28	20	34
2023	48	33	25	39
2024	55	39	30	45

Working	Comment
Plot the 12 quarterly figures in time order and join them.	
The same shape repeats each year: a Q1 peak, falling to a Q3 low, then rising again.	The pattern repeats over a **fixed period of 4 quarters** ⇒ **seasonality**.
Sales are highest in Q1 (summer) and lowest in Q3 (winter) every year.	Describe *when* the peaks and troughs occur, in context.
Yearly means: $\dfrac{42+28+20+34}{4}=31$, $\dfrac{48+33+25+39}{4}=36.25$, $\dfrac{55+39+30+45}{4}=42.25$.	Averaging over each year removes the seasonal swing so the trend is visible.
The yearly means rise: $31 \rightarrow 36.25 \rightarrow 42.25$.	A rising average ⇒ an **increasing trend**.

The series is **seasonal** with a period of 4 quarters — sales peak in summer (Q1) and bottom out in winter (Q3) each year — and it also shows an **increasing trend**, the yearly mean rising from 31 to 42.25 thousand dollars.

Example 3 — unscaffolded

The time series plot shows the monthly number of visitors (in thousands) to a national park over two years. Describe the features of the series in context.



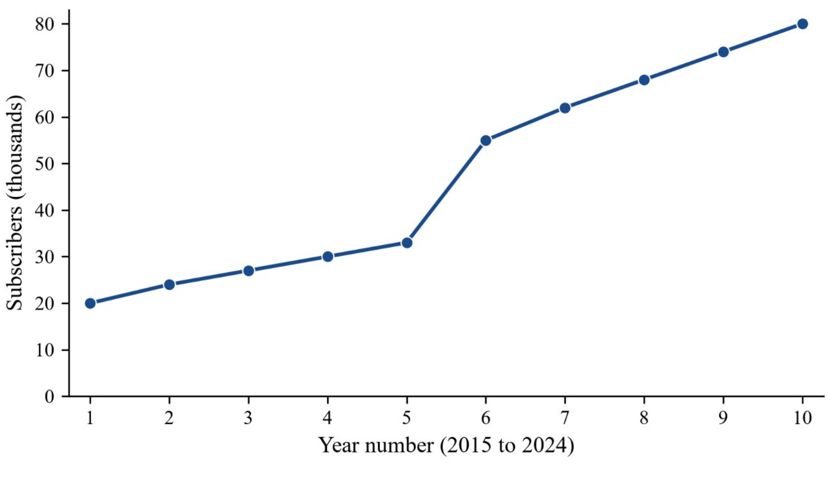
The series is strongly **seasonal**, with a period of 12 months: visitor numbers peak in the summer months (around 60 thousand each January) and fall to a trough in the winter months (around 20 thousand each June–July). Comparing the two years, the peaks and troughs are a little higher in the second year, suggesting a slight **increasing trend** in park visits.

There is also one clear **outlier**. In April of the second year (month 16) only 8 thousand visitors were recorded, far below the roughly 34 thousand seen in April of the first year — a drop of about 26 thousand. This one-off dip, most likely caused by a temporary closure of the park, does not fit the seasonal pattern and the series returns to that pattern the following month.

∴ the series shows seasonality (period 12) with a slight upward trend, plus a single outlier in the second April caused by a one-off closure.

Example 4 — unscaffolded

A streaming service records the number of subscribers (in thousands) at the end of each year from 2015 to 2024 as 20, 24, 27, 30, 33, 55, 62, 68, 74, 80. Describe the features of this time series in context.



Overall the number of subscribers shows an **increasing trend** across the ten years, rising from 20 thousand in 2015 to 80 thousand in 2024. The rise is not smooth, however. For the first five years the growth is gentle, adding about 3 thousand subscribers a year ($24-20=4$, then $27-24=3$, $30-27=3$, $33-30=3$, an average of about 3.25 per year).

Between 2019 and 2020 the series jumps abruptly from 33 to 55 thousand — a rise of $55-33=22$ thousand in a single year, far larger than any earlier yearly change — and then continues to grow steadily from this new, higher level. This sudden and lasting shift is a **structural change**, most plausibly caused by the service launching in a new country during 2020.

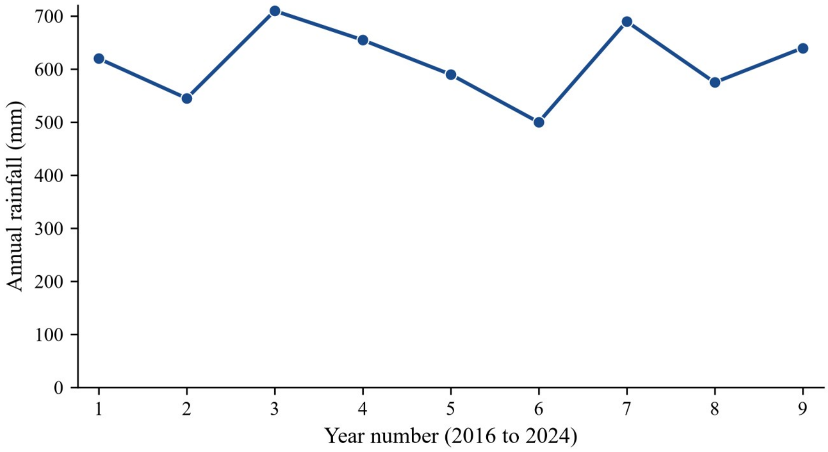
∴ the series has an overall increasing trend, with a structural change between 2019 and 2020 that lifts subscribers to a permanently higher level.

Practice questions

Scaffolded

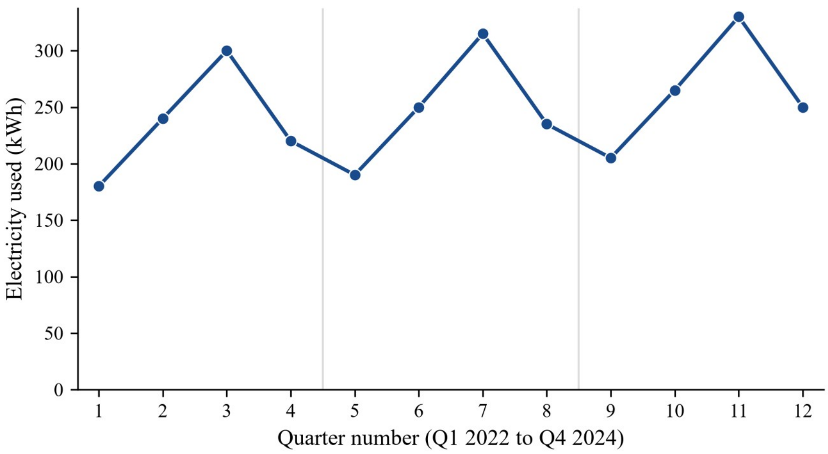
Q1. The population of a town (in thousands), recorded at the end of each year from 2019 to 2025, was 12, 13, 15, 16, 18, 19, 21. (a) Which variable is placed on the horizontal axis of the time series plot? (b) Construct the time series plot. (c) Describe the trend. (d) State the total change in population over the seven years.

Q2. The time series plot shows the annual rainfall (mm) at a weather station from 2016 to 2024.



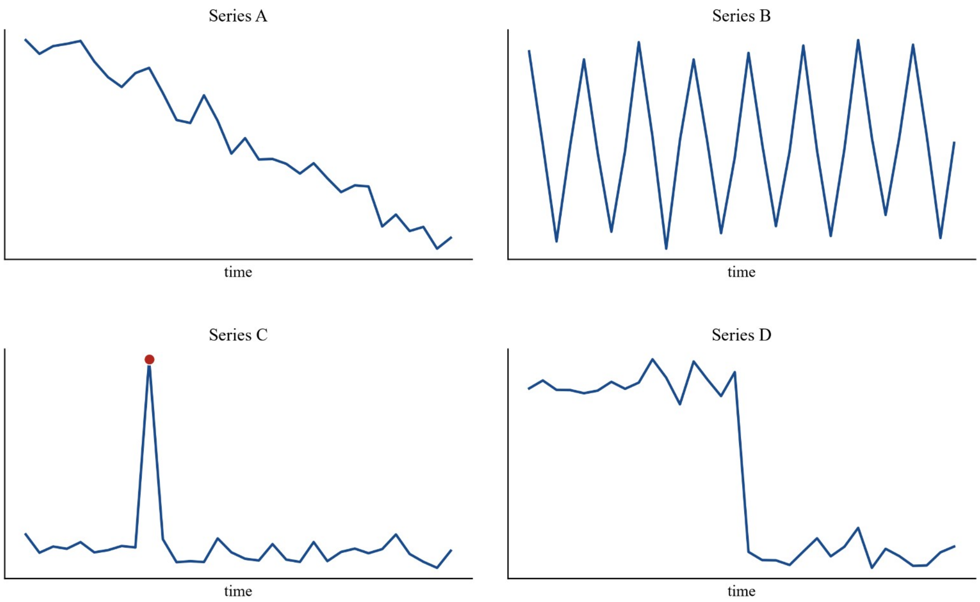
(a) State the rainfall recorded in 2018.
(b) In which year was the rainfall greatest, and what was it?
(c) In which year was the rainfall least, and what was it?
(d) Does the series show a clear long-term trend? What feature best describes it instead?

Q3. The time series plot shows a household's quarterly electricity use (kWh) over three years.



(a) How many quarters are there in one year?
(b) State the period of the seasonality.
(c) In which quarter is electricity use highest each year? Suggest why.
(d) Using the yearly means 235, 247.5 and 262.5 kWh, describe the trend.

Q4. Each panel below shows a different time series. For each, name the single feature (trend, seasonality, outlier or structural change) that best describes it.

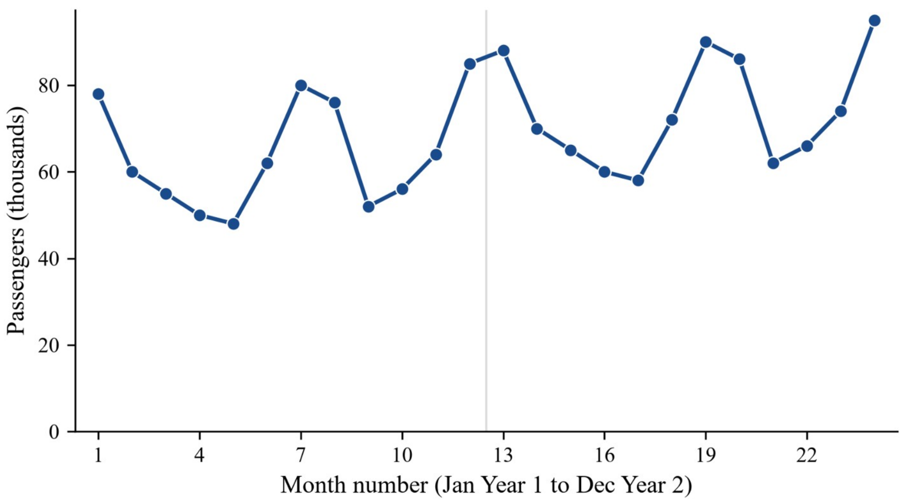


Q5. A shop records the number of customers on each of 14 consecutive days: 120, 115, 130, 125, 118, 122, 128, 260, 124, 119, 131, 126, 120, 123. (a) Construct the time series plot. (b) Identify the outlier: on which day did it occur and what was its value? (c) The typical daily count is about 123. Suggest a possible cause for the outlier.

Q6. For each of the following, state whether the data form a time series (a numerical variable recorded at successive times) or not, and give a brief reason. (a) The heights of 30 students, all measured on the same day. (b) A company's total revenue for each month over two years. (c) The daily maximum temperature at a city, recorded every day for a year. (d) The marks of the students in one class on a single test.

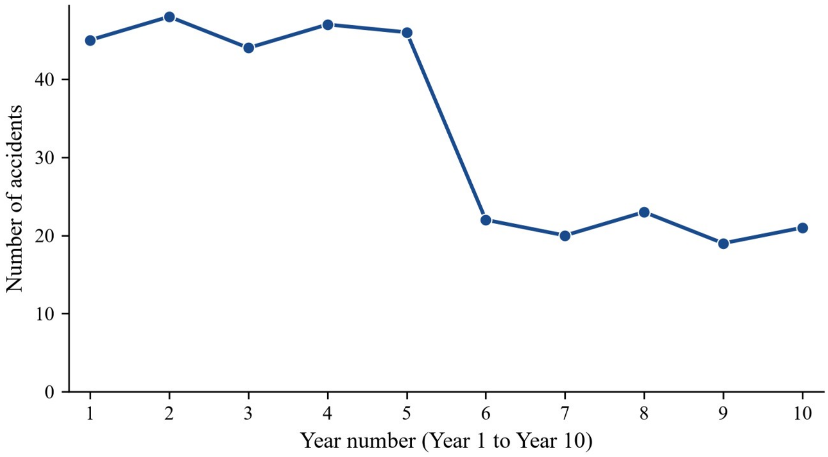
Unscaffolded

Q7. The time series plot shows the monthly number of airline passengers (in thousands) over two years. Describe all the features of the series in context.



Q8. Explain the difference between **seasonality** and **irregular fluctuations** in a time series, and give one example of each.

Q9. The time series plot shows the annual number of accidents at an intersection over ten years. A roundabout was installed at the end of Year 5. Describe what the plot shows, naming the feature involved and interpreting it in context.

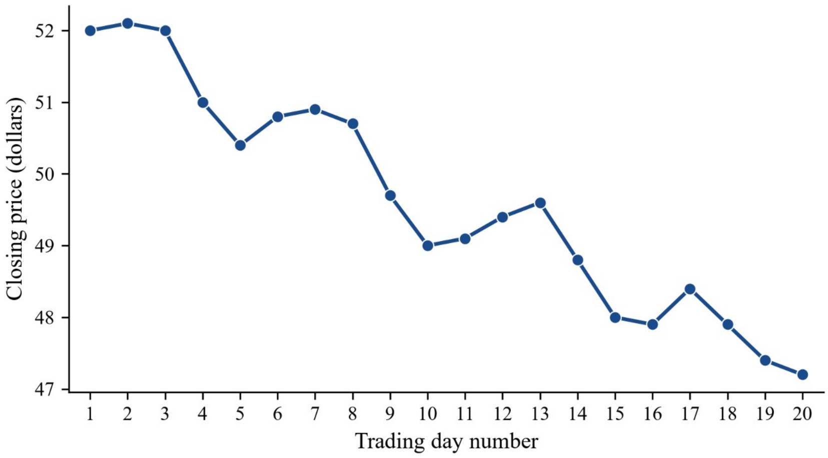


Q10. A ski resort's quarterly revenue (in thousands of dollars) over two years was

Year	Q1	Q2	Q3	Q4
2023	20	60	120	40
2024	25	68	132	46

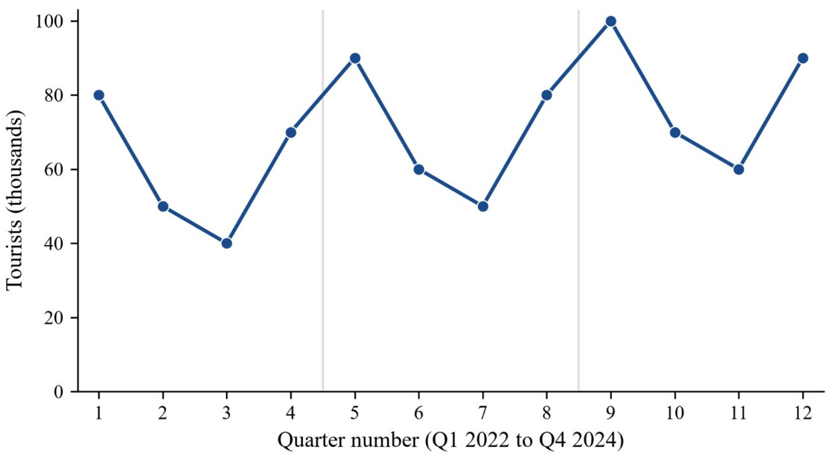
(a) Construct the time series plot. (b) Describe its features in context, stating the period of the seasonality and in which quarter revenue peaks. (c) Use the yearly means to comment on the trend.

Q11. The time series plot shows the daily closing price (dollars) of a share over 20 trading days. Describe the features of the series, and explain why it would not be described as seasonal.



Q12. In a time series plot the successive points are joined by straight line segments. (a) Explain what the line segments help the reader to see. (b) Explain why time is treated as the explanatory variable and placed on the horizontal axis.

Q13. The time series plot shows the quarterly number of tourists (in thousands) visiting a town over three years.



(a) State the period of the seasonality and the quarter in which tourism peaks.
(b) The yearly means are 60, 70 and 80 thousand. Describe the trend and state the average increase per year.

Q14. For each scenario, name the time series feature (trend, seasonality, irregular fluctuation, outlier or structural change) that it describes. (a) A retailer's sales rise sharply every December. (b) A factory's daily output jumps permanently after new machinery is installed. (c) Production falls for a single day because of a one-off power failure. (d) A city's electricity demand is higher on every weekday and lower every weekend. (e) A city's water use has fallen steadily over the last twenty years. (f) The small, unpredictable day-to-day changes in a share price.

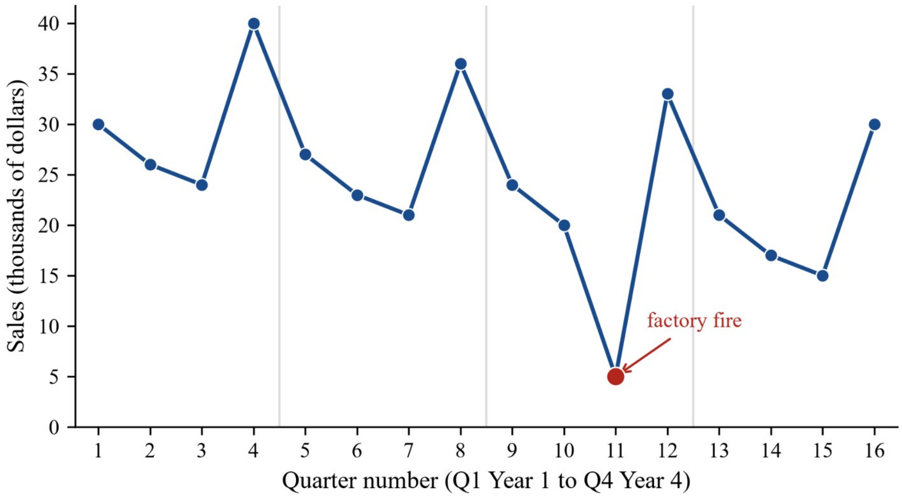
Q15. A sporting club's membership at the end of each year from 2018 to 2024 was 340, 372, 401, 438, 470, 505, 542. (a) Describe the trend. (b) Find the total change in membership over the period. (c) Find the average change per year, correct to two decimal places.

Q16. A CAS-produced time series plot of a store's monthly sales over four years shows a pattern that repeats each year, peaking every December, together with a gradual rise in the December peaks from one year to the next. (a) Name the feature responsible for the December peaks, and state its period. (b) Name the feature responsible for the year-on-year rise in the peaks.

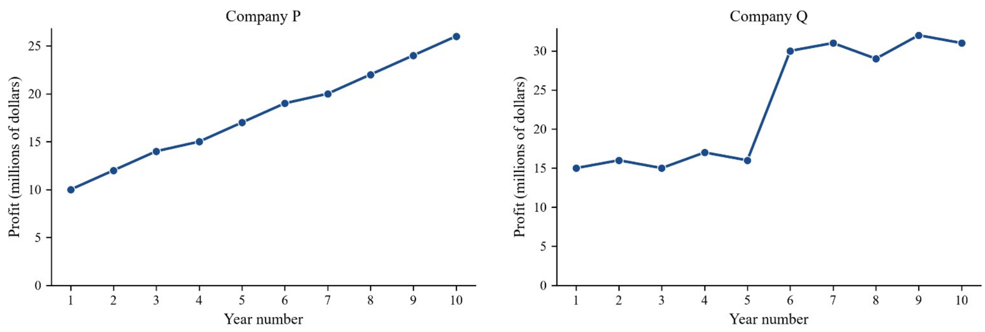
Q17. (a) Explain what is meant by an **outlier** in a time series. (b) Give two different real-world events that could produce an outlier in a series of monthly retail sales.

Q18. State the period of the seasonality you would expect in a time series made up of (a) monthly observations, (b) quarterly observations, (c) daily observations that follow a weekly pattern.

Q19. The time series plot shows a product's quarterly sales (in thousands of dollars) over four years. Describe the features of the series in context, including the period of any seasonality, the direction of the trend, and any outlier.



Q20. The two time series plots show the annual profit (in millions of dollars) of two companies over ten years. Describe the main feature of each series and compare them.

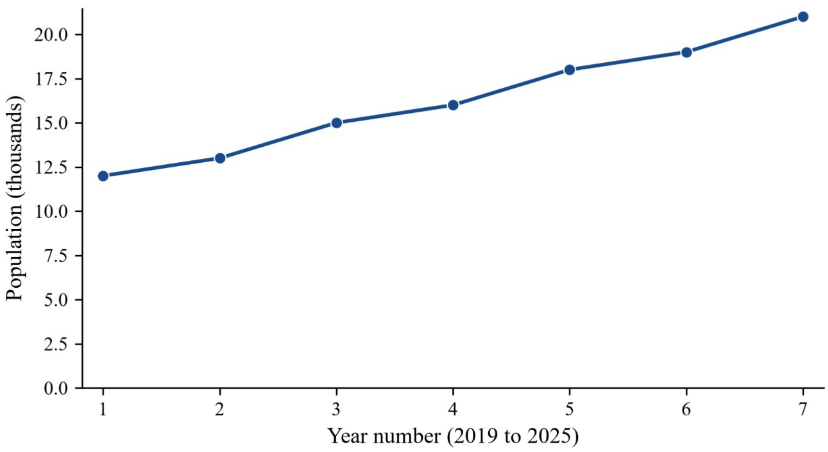


Answers

Q1. (a) Year (time). (b) Points $(2019,12),\ldots,(2025,21)$ joined in order. (c) Increasing trend. (d) $21-12=9$ thousand. **Q2.** (a) 710 mm. (b) 2018, 710 mm. (c) 2021, 500 mm. (d) No clear trend; the series shows irregular fluctuations. **Q3.** (a) 4. (b) 4 quarters. (c) Q3 (winter), because of heating. (d) The yearly means rise, so there is an increasing trend. **Q4.** A: (decreasing) trend. B: seasonality. C: outlier. D: structural change. **Q5.** (a) Points joined in day order. (b) Day 8, 260 customers. (c) A one-off event such as a holiday sale. **Q6.** (a) Not a time series (one time only). (b) Time series (monthly, in time order). (c) Time series (daily, in time order). (d) Not a time series (one test, not successive times). **Q7.** Seasonality (period 12), with peaks over the December–January holidays and again in July, and the lowest numbers in the autumn months, plus an increasing trend (every month exactly 10 thousand higher in Year 2). There are no irregular fluctuations, no outliers and no structural change. **Q8.** Seasonality is systematic — a calendar-related pattern repeating over a fixed period of one year or less (e.g. sales peaking each summer); irregular fluctuations are unsystematic, short-term ups and downs with no repeating pattern (e.g. the day-to-day wobbles in a share price). **Q9.** A structural change: accidents ran at about 46 per year before the roundabout and dropped to about 21 per year afterwards — a sudden, sustained fall caused by the roundabout. **Q10.** (a) 8 quarterly points joined in order. (b) Seasonal, period 4, peaking in Q3 (winter ski season) and lowest in Q1. (c) Yearly means $60 \to 67.75$, so an increasing trend. **Q11.** Irregular short-term fluctuations about a slight downward trend (from about 52 to about 47 dollars). Not seasonal: there is no fixed, repeating calendar pattern over the 20 days. **Q12.** (a) They show how the variable moves from one time to the next, making trends and repeating patterns easy to see. (b) Time drives the measurements (each value belongs to a time), so time is explanatory and is placed on the horizontal axis. **Q13.** (a) Period 4, peaking in Q1 (summer). (b) Increasing trend; average increase $\frac{80-60}{2}=10$ thousand per year. **Q14.** (a) Seasonality. (b) Structural change. (c) Outlier. (d) Seasonality (period 7). (e) Trend. (f) Irregular fluctuation. **Q15.** (a) Increasing trend. (b) $542-340=202$ members. (c) $\frac{202}{6}\approx 33.67$ members per year. **Q16.** (a) Seasonality, period 12. (b) Trend (an increasing trend). **Q17.** (a) A single value lying well away from the surrounding pattern, usually from a one-off event. (b) Any two, e.g. a Boxing Day sale spike; a store closure from a flood or fire. **Q18.** (a) 12. (b) 4. (c) 7. **Q19.** Seasonal, period 4, peaking in Q4 each year; an overall decreasing trend (Q4 peaks fall 40, 36, 33, 30); and an outlier in Q3 of Year 3 (value 5, far below the pattern), likely a one-off disruption. **Q20.** Company P shows a steady increasing trend (profit rising from about 10 to 26 million). Company Q shows a structural change: profit is roughly steady near 16 million, then jumps to about 30 million after Year 5 and stays there. P grows gradually; Q changes suddenly and then holds a higher level.

Fully-worked solutions

Q1. Time is the explanatory variable, so **year** is on the horizontal axis. Plotting $(2019,12),(2020,13),(2021,15),(2022,16),(2023,18),(2024,19),(2025,21)$ and joining the points in order gives a graph that climbs throughout, so the population shows an **increasing trend**. The total change is $21-12=9$ thousand people.

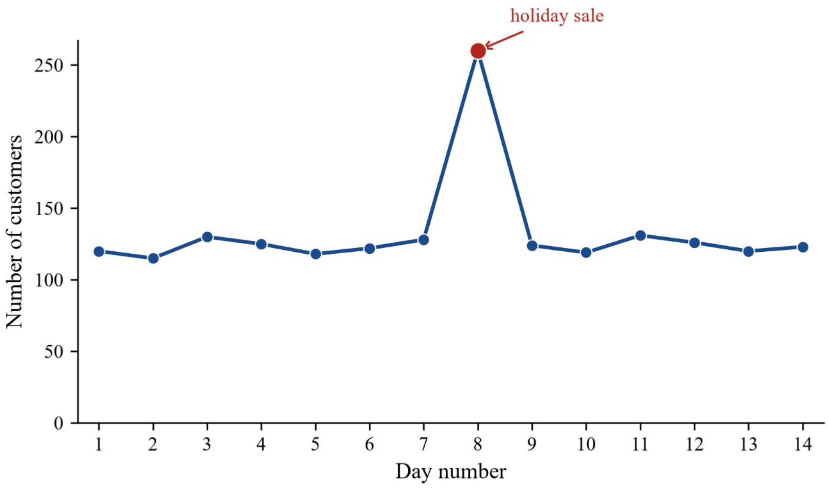


Q2. Reading heights from the plot: in 2018 the rainfall was **710 mm**, which is also the **greatest** value (2018, 710 mm). The **least** value is **500 mm**, in **2021**. The values move up and down from year to year with no sustained rise or fall, so there is **no clear long-term trend** — the series is dominated by **irregular fluctuations**.

Q3. (a) There are **4** quarters in a year. (b) The pattern repeats every year, so the period of the seasonality is **4 quarters**. (c) Use is highest in **Q3** (the winter quarter), most likely because of heating. (d) The yearly means are $235, 247.5, 262.5$; since they increase from year to year, the series has an **increasing trend** underneath the seasonal swing.

Q4. Series **A** falls steadily overall — a (decreasing) **trend**. Series **B** repeats the same shape at a fixed period — **seasonality**. Series **C** is flat apart from one value far above the rest — an **outlier**. Series **D** sits at one level, then jumps to a new level and stays there — a **structural change**.

Q5. Plot the 14 daily counts in order and join them. Thirteen of the days sit close together near 120–130 customers; **day 8** stands far above them at **260** customers. Since it lies well away from the surrounding values, it is an **outlier**, most plausibly caused by a one-off event such as a **holiday sale**. After day 8 the counts return to their usual level, confirming it is an outlier rather than a structural change.



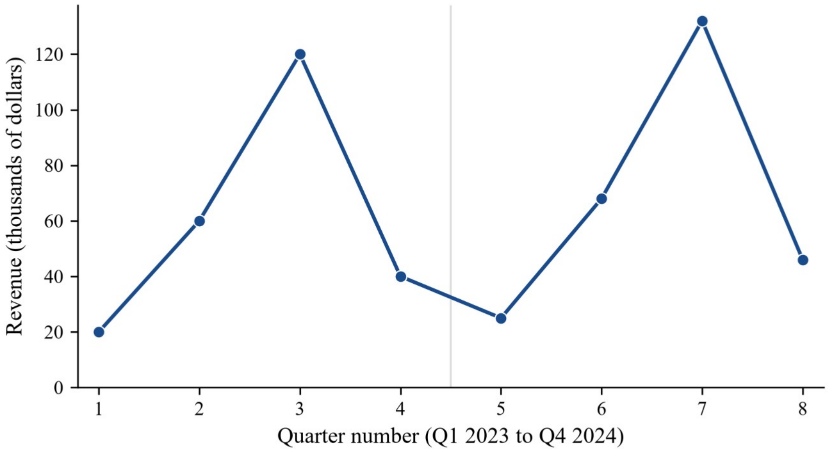
Q6. (a) **Not** a time series: all 30 heights are measured at the same single time, not at successive times. (b) **A time series**: revenue is recorded month after month, in time order. (c) **A time series**: the maximum temperature is recorded day after day, in time order. (d) **Not** a time series: the marks are all from one test at one time.

Q7. The series repeats the same yearly shape — peaking over the December–January summer holidays and again in July, and falling to its lowest values in the autumn months (April and May) — so it is **seasonal** with a **period of 12** months. Comparing the two years, every month sits exactly 10 thousand passengers higher in the second year, so there is also an **increasing trend**. Because the second year reproduces the first year's shape exactly, every rise and fall belongs to that repeating shape: the series shows **no irregular fluctuations**, no outliers and no structural change. In context: passenger numbers are seasonal, with holiday peaks in summer and again in July, and are rising year on year.

Q8. Seasonality is a **systematic** pattern: it is tied to the calendar and repeats over a **fixed, known period** of one year or less — for example, ice-cream sales peaking every summer and falling every winter (period 12 for monthly data). Because the period is fixed, the pattern is predictable, and you can say in advance when the next peak is due. **Irregular fluctuations** are **unsystematic**, short-term ups and downs that are left once trend and seasonality are accounted for — for example, the day-to-day wobbles in a share price. They follow no repeating pattern, so they cannot be predicted. The essential difference is that seasonality recurs on a fixed calendar schedule, whereas irregular fluctuations recur on no schedule at all.

Q9. For the first five years the number of accidents fluctuates around **46 per year** $\left(45+48+44+47+46/5=46\right)$. After the roundabout is installed at the end of Year 5, the count drops to around **21 per year** $\left(22+20+23+19+21/5=21\right)$ and stays at that lower level. This sudden and sustained fall — a drop of about $46-21=25$ accidents per year — is a **structural change** caused by the roundabout, not an outlier, because the lower level persists.

Q10. (a) Plot the eight quarterly figures in order and join them. (b) The same shape repeats each year — a low Q1, a high **Q3**, then a fall — so the series is **seasonal** with a **period of 4** quarters, peaking in **Q3** during the winter ski season and lowest in Q1. (c) The yearly means are $20+60+120+40/4=60$ and $25+68+132+46/4=67.75$ thousand dollars; because the mean rises, there is an **increasing trend**.



Q11. Over the 20 trading days the price jitters up and down from one day to the next — these are **irregular (random) fluctuations** — while drifting gently downward from about **52** to about **47** dollars, a slight **decreasing trend** $(52.0-47.2=4.8)$. The series is **not seasonal** because there is no fixed, repeating calendar pattern in the data: the rises and falls come at no regular interval, and a share price has no built-in calendar period of the kind that drives ice-cream sales or electricity use.

Q12. (a) The line segments connect each value to the next in time order, showing at a glance how the variable **moves from one time to the next** and making trends, seasonal swings and sudden changes easy to see. (b) Each value is recorded *at* a particular time, so time is what the measurements depend on; it is therefore the **explanatory variable** and is placed on the **horizontal axis**.

Q13. (a) The pattern repeats every four quarters, so the period is **4**, and tourism peaks each year in **Q1** (summer). (b) The yearly means $60, 70, 80$ increase steadily, so the series has an **increasing trend**; the average increase is $\frac{80-60}{2}=10$ thousand tourists per year.

Q14. (a) A sharp rise at the same time each year is **seasonality**. (b) A permanent jump to a new level is a **structural change**. (c) A single one-off drop is an **outlier**. (d) A pattern that repeats with every week is **seasonality**; for daily data that is a period of **7**. (e) A steady long-term fall is a **trend**. (f) Small unpredictable day-to-day changes are **irregular fluctuations**.

Q15. (a) The membership figures $340, 372, 401, 438, 470, 505, 542$ increase every year, so there is an **increasing trend**. (b) The total change is $542-340=202$ members. (c) There are 6 yearly intervals, so the average change is $\frac{202}{6}\approx 33.67$ members per year.

Q16. (a) A pattern that repeats each year, peaking every December, is **seasonality**; with monthly data its period is **12**. (b) The steady rise in the peaks from one year to the next is the **trend** (an increasing trend) underlying the seasonal pattern.

Q17. (a) An **outlier** is a single data value that lies well away from the pattern of the surrounding values, typically produced by a **one-off event**, after which the series returns to its usual behaviour. (b) Two possible causes for an outlier in monthly retail sales are a **one-off sales event** (such as an unusually large Boxing Day promotion, giving a spike) and a **temporary store closure** (from a flood or fire, giving a dip).

Q18. The period equals the number of observations in one complete calendar cycle: (a) monthly observations → **12**; (b) quarterly observations → **4**; (c) daily observations with a weekly pattern → **7**.

Q19. The same shape repeats every four quarters, so the series is **seasonal** with a **period of 4**, peaking in **Q4** each year. Following the Q4 peaks — 40, 36, 33, 30 — they fall steadily from year to year, so the underlying **trend is**

decreasing. One point does not fit: in **Q3 of Year 3** sales are only **5** thousand dollars, far below the Q3 values in the other years, so it is an **outlier**, most likely caused by a one-off disruption (for example, the factory fire). In context, this product's sales are seasonal (a Q4 peak), declining over the four years, with a single outlier caused by a one-off event.

Q20. Company P rises fairly evenly from about 10 to 26 million dollars over the ten years — a steady **increasing trend**. **Company Q** stays near 16 million for the first five years, then jumps to about 30 million after Year 5 and remains at that higher level — a **structural change**. Comparing them: both end higher than they began, but P's profit grows *gradually and continuously*, whereas Q's is *steady, then rises suddenly* and holds the new level. A gradual trend and a one-off structural change are different ways a series can increase over time.

Chapter 5 Investigating and modelling time series — 5B Smoothing a time series using moving means

Theory

A **time series** is a set of measurements recorded in time order — daily temperatures, quarterly sales, monthly rainfall. When the points of a time-series plot are joined, the **trend** (the long-term upward or downward direction) is often hidden by **irregular fluctuations** and, in many data sets, by a repeating **seasonal** pattern. **Smoothing** damps down these short-term movements so that the underlying trend becomes visible.

The most common numerical method is the **moving mean** (or **moving average**). We replace each observation by the mean of itself and an equal number of neighbours on each side; as the "window" of values slides one step along the series, a new smoothed value is produced at each step. Because each smoothed value is an average of several observations, the ups and downs partly cancel and the smoothed series is flatter than the original.

Odd-order moving means. When the number of terms is **odd**, the window has a single middle value, so the moving mean lines up naturally with that middle time point. The **3-point moving mean** at time t is

$$\frac{y_{t-1}+y_t+y_{t+1}}{3},$$

and the **5-point moving mean** at time t is

$$\frac{y_{t-2}+y_{t-1}+y_t+y_{t+1}+y_{t+2}}{5}.$$

Each smoothed value is written against the **centre** point of its window. A 3-point mean cannot be found for the very first or very last observation (there is no neighbour on one side), so **one point is lost at each end**; a 5-point mean loses **two points at each end**. In general a k-point moving mean (with k odd) loses $\frac{k-1}{2}$ points at each end.

Choosing the number of terms. More terms give a smoother series, but lose finer detail and more end points. A key use of smoothing is to remove **seasonality**: if the pattern repeats every p time periods (for example $p=4$ for quarterly data, $p=12$ for monthly data, or $p=7$ for daily data with a weekly cycle), choose a moving mean with p **terms**, so that each smoothed value averages exactly one complete cycle and the seasonal swing is averaged out.

Even-order moving means and centring. A cycle length is often **even** (4 quarters, 12 months), but the mean of an even number of terms does **not** sit above a data point — it falls **halfway between** the two middle time points. To bring the smoothed value back onto a data point we apply a **centring** step: take the mean of two neighbouring even moving means. This is a 2-term moving mean *of* the moving means, sometimes written "2 × 4", and the result is a **centred moving mean**.

For a **4-point centred moving mean** at time t, the two 4-point windows either side of t average to

$$\frac{y_{t-2}+2y_{t-1}+2y_t+2y_{t+1}+y_{t+2}}{8}.$$

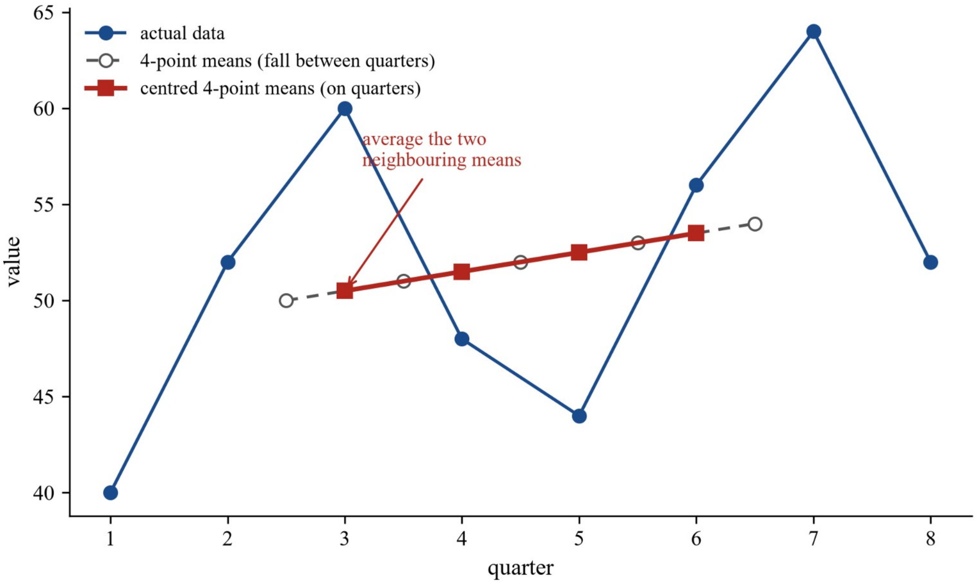
The three inner values carry a weight of 2 and the two outer values a weight of 1, and the weights add to $1+2+2+2+1=8$. A 4-point centred mean **loses two points at each end**.

For a **2-point centred moving mean** at time t the same idea gives

$$\frac{y_{t-1}+2y_t+y_{t+1}}{4},$$

which loses one point at each end.

The plot below shows the centring idea for quarterly data: the plain 4-point means (grey, open) fall between the quarters, and averaging each neighbouring pair brings the centred moving means (red) back onto a quarter.

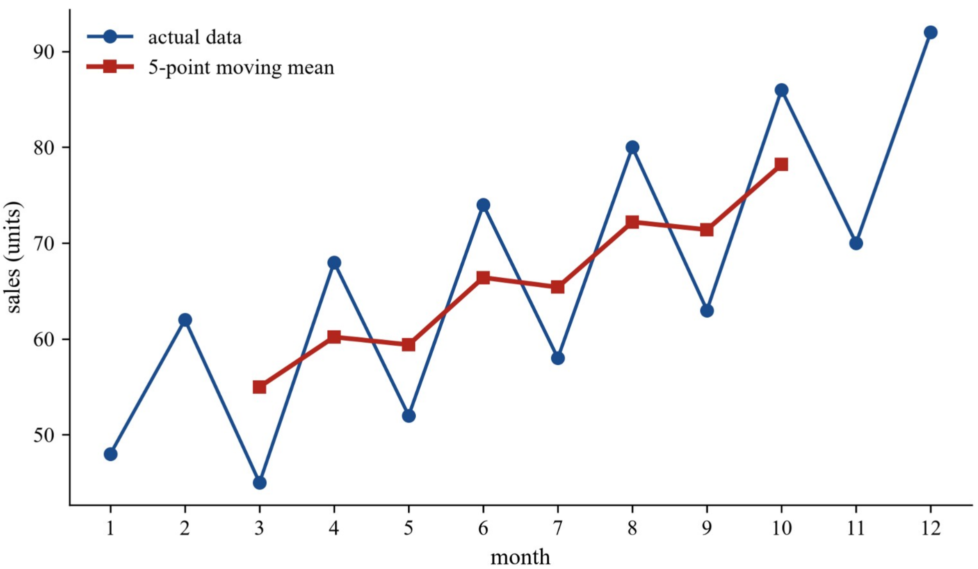


The table summarises the four most common cases.

Terms	Odd/even	Centring needed?	Points lost at each end
3-point	odd	no	1
5-point	odd	no	2
2-point centred	even	yes	1
4-point centred	even	yes	2

Present the smoothed values in a table beside the original series, then **plot the smoothed series on the same axes** as the raw plot: the smoothed line reveals the trend that the raw fluctuations obscured. Where a moving mean is not a whole number we give it as an exact terminating decimal, or correct to one decimal place.

The plot below shows monthly sales (blue) with strong month-to-month fluctuation; the 5-point moving mean (red) exposes a steady rising trend.



Using CAS. General Mathematics is technology-active: a CAS calculator or spreadsheet will compute moving means once the series is entered. Learn the by-hand method so that you understand what a smoothed value represents, can carry out centring correctly, and can check that the technology's output is reasonable. Smoothing with moving **medians**, which resist outliers, is treated separately.

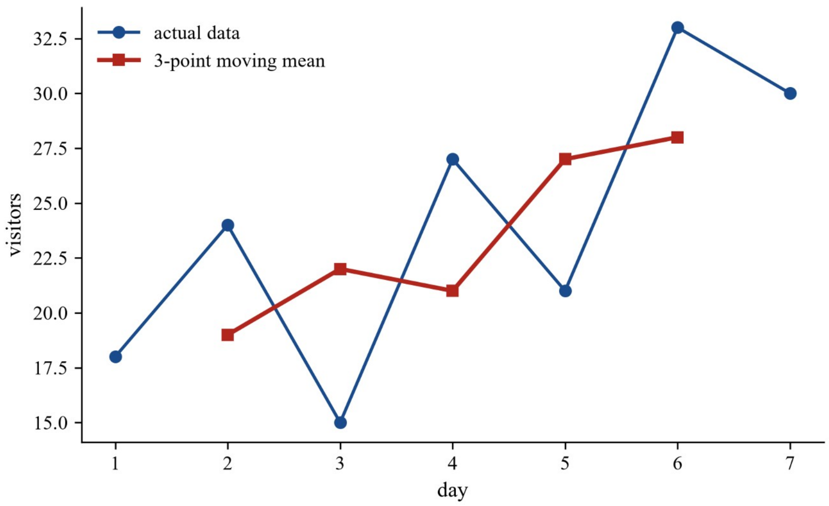
Worked examples

Example 1 — scaffolded

The numbers of visitors to a museum on each of seven days were 18, 24, 15, 27, 21, 33, 30. Calculate the 3-point moving means and use them to describe the trend.

Working	Comment
The first 3-point mean is centred on day 2: $\frac{18+24+15}{3}=\frac{57}{3}=19.$	Average days 1, 2, 3 and write it against the middle day. Day 1 has no mean.
Day 3: $\frac{24+15+27}{3}=\frac{66}{3}=22.$	Slide the window one day along: average days 2, 3, 4.
Day 4: $\frac{15+27+21}{3}=\frac{63}{3}=21.$	Average days 3, 4, 5.
Day 5: $\frac{27+21+33}{3}=\frac{81}{3}=27.$	Average days 4, 5, 6.
Day 6: $\frac{21+33+30}{3}=\frac{84}{3}=28.$	Average days 5, 6, 7. Day 7 has no mean.
Smoothed series (days 2–6): 19, 22, 21, 27, 28.	The first and last day have no 3-point mean.

The raw values zig-zag, but the smoothed series rises from 19 to 28, so the underlying trend is increasing.



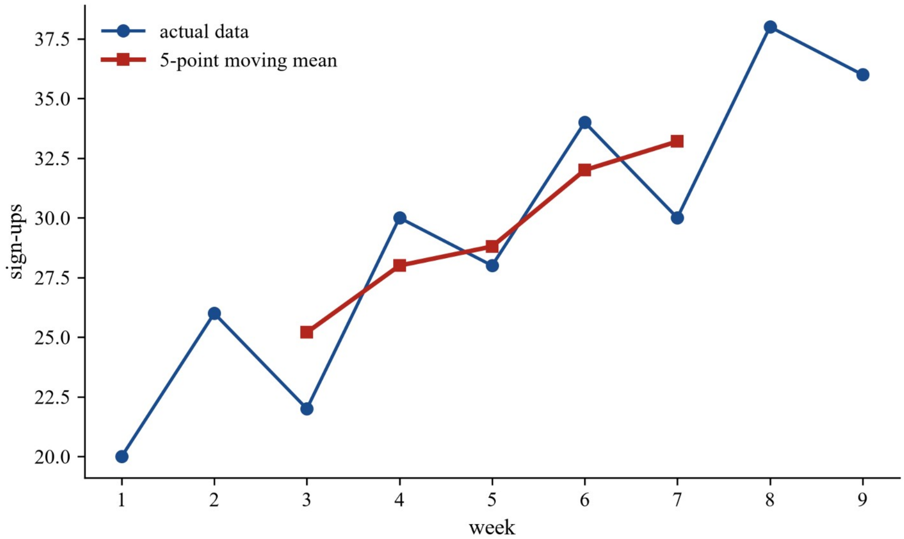
Example 2 — scaffolded

The weekly numbers of new sign-ups to a website over nine weeks were 20, 26, 22, 30, 28, 34, 30, 38, 36. Calculate the 5-point moving means, correct to one decimal place where needed.

Working	Comment
The first 5-point mean is centred on week 3: $\frac{20+26+22+30+28}{5}=\frac{126}{5}=25.2.$	A 5-point mean needs two values on each side, so weeks 1 and 2 have no mean.
Week 4: $\frac{26+22+30+28+34}{5}=\frac{140}{5}=28.$	Slide the window one week along: average weeks 2–6.
Week 5: $\frac{22+30+28+34+30}{5}=\frac{144}{5}=28.8.$	Average weeks 3–7.
Week 6: $\frac{30+28+34+30+38}{5}=\frac{160}{5}=32.$	Average weeks 4–8.

Working	Comment
Week 7: $\dfrac{28+34+30+38+36}{5}=\dfrac{166}{5}=33.2.$	Average weeks 5–9. Weeks 8 and 9 have no mean.
Smoothed series (weeks 3–7): 25.2, 28, 28.8, 32, 33.2.	Four points (the first two and last two) are lost.

The 5-point smoothing removes almost all of the week-to-week fluctuation, leaving a clear upward trend.



Example 3 — unscaffolded

The quarterly electricity use (kWh) of a household over two years was as follows.

Quarter	1	2	3	4	5	6	7	8
Use (kWh)	40	52	60	48	44	56	64	52

The data show a strong seasonal pattern, so we smooth with a 4-point moving mean, centring it so that each smoothed value lands on a quarter.

First find the 4-point moving means. Each averages a full year, so it falls **between** two quarters:

$$\frac{40+52+60+48}{4}=50,\ \frac{52+60+48+44}{4}=51,\ \frac{60+48+44+56}{4}=52,$$

$$\frac{48+44+56+64}{4}=53,\ \frac{44+56+64+52}{4}=54.$$

Now centre each neighbouring pair of 4-point means so that the value sits on a quarter:

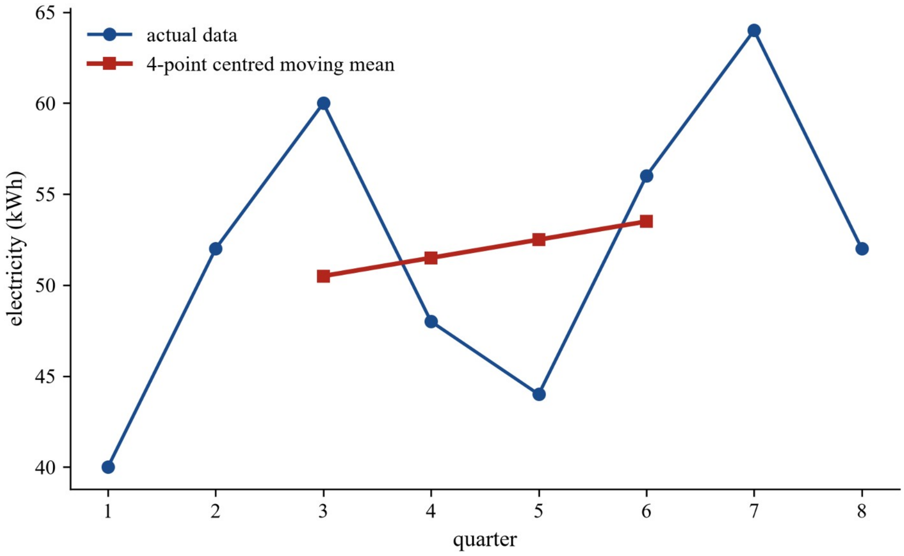
$$\text{Quarter 3: }\frac{50+51}{2}=50.5,\ \text{Quarter 4: }\frac{51+52}{2}=51.5,$$

$$\text{Quarter 5: }\frac{52+53}{2}=52.5,\ \text{Quarter 6: }\frac{53+54}{2}=53.5.$$

The same values come from the weighted formula; for quarter 3, $\dfrac{40+2(52)+2(60)+2(48)+44}{8}=\dfrac{404}{8}=50.5.$

Two quarters are lost at each end.

∴ the centred smoothed series for quarters 3–6 is 50.5, 51.5, 52.5, 53.5 kWh. The seasonal swings have been removed, revealing a steady upward trend.



Example 4 — unscaffolded

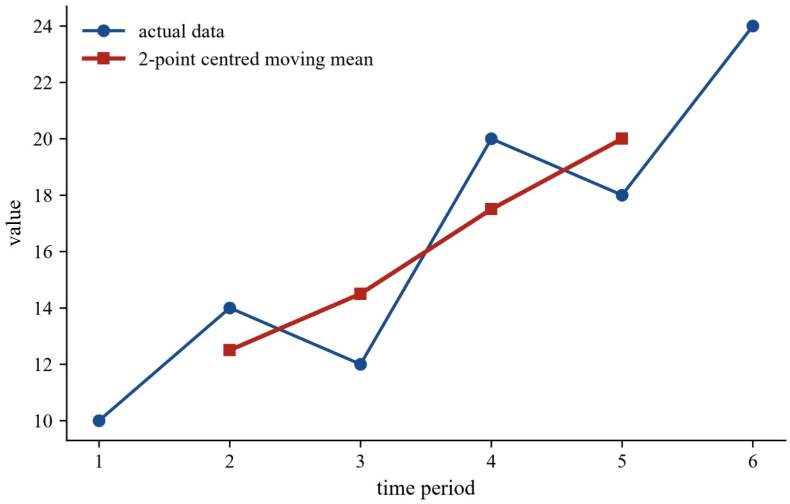
The values recorded at six equally spaced times were 10, 14, 12, 20, 18, 24. Smooth the series using a 2-point centred moving mean.

A 2-point moving mean averages two neighbouring values, so it falls between them; centring then averages two neighbouring 2-point means, bringing the result back onto a data point. Using the weighted form $\frac{y_{t-1}+2y_t+y_{t+1}}{4}$:

$$t=2:\frac{10+2(14)+12}{4}=\frac{50}{4}=12.5,\ t=3:\frac{14+2(12)+20}{4}=\frac{58}{4}=14.5,$$

$$t=4:\frac{12+2(20)+18}{4}=\frac{70}{4}=17.5,\ t=5:\frac{20+2(18)+24}{4}=\frac{80}{4}=20.$$

∴ the 2-point centred smoothed series (times 2–5) is 12.5, 14.5, 17.5, 20, which increases steadily.



Practice questions

Scaffolded

Q1. The daily numbers of customers at a café over seven days were 24, 15, 21, 30, 27, 36, 33. (a) Explain why no 3-point moving mean can be found for day 1. (b) Find the 3-point moving mean centred on day 2. (c) Find the 3-point moving means centred on days 3, 4, 5 and 6. (d) Write the smoothed series and state the direction of the trend.

Q2. The weekly numbers of visitors to a park over nine weeks were 40, 55, 45, 60, 50, 65, 55, 75, 65. (a) State which weeks will have no 5-point moving mean. (b) Find the 5-point moving mean centred on week 3. (c) Find the remaining 5-point moving means (weeks 4 to 7). (d) Write the smoothed series.

Q3. The values recorded at six equally spaced times were 16, 22, 20, 28, 26, 32. (a) Find the five 2-point moving means (they fall between the time points). (b) Centre them by averaging each neighbouring pair. (c) Write the 2-point centred smoothed values (times 2 to 5).

Q4. The quarterly sales (in thousands of dollars) of a firm over two years were 60, 80, 72, 52, 66, 86, 78, 58. (a) Find the five 4-point moving means (they fall between quarters). (b) Centre neighbouring pairs to obtain the smoothed values on quarters 3 to 6. (c) Confirm the quarter-3 value using the weighted formula $\frac{y_1+2y_2+2y_3+2y_4+y_5}{8}$.

Q5. The monthly numbers of memberships sold by a gym over six months were 12, 20, 16, 24, 20, 28. (a) Find the 3-point moving means (months 2 to 5). (b) Describe what the smoothing shows about the trend.

Q6. The maximum daily temperature (°C) at a town over eight days was 18, 21, 12, 24, 15, 27, 18, 30. (a) Find the 3-point moving means (days 2 to 7). (b) On the same axes, plot the raw temperatures and the smoothed series.

Unscaffolded

Q7. Find the 3-point moving means of the series 30, 24, 36, 30, 42, 36, 48.

Q8. Find the 5-point moving means of the series 12, 18, 15, 21, 18, 24, 21, 27, 24.

Q9. The quarterly rainfall (mm) at a station over two years was 120, 90, 150, 140, 130, 100, 160, 150. Find the 4-point centred moving means for quarters 3 to 6.

Q10. Find the 2-point centred moving means of the series 8, 12, 10, 16, 14, 20.

Q11. A series over nine time periods is 51, 63, 54, 66, 57, 69, 60, 72, 63. (a) Find the 3-point moving means. (b) Find the 5-point moving means. (c) State which smoothing gives the flatter (smoother) series, and explain why.

Q12. Monthly unemployment figures collected over several years show a pattern that repeats each year, peaking every January. (a) Name this qualitative feature of the time series. (b) How many terms should a moving mean use to smooth it out? (c) Will centring be needed? Explain.

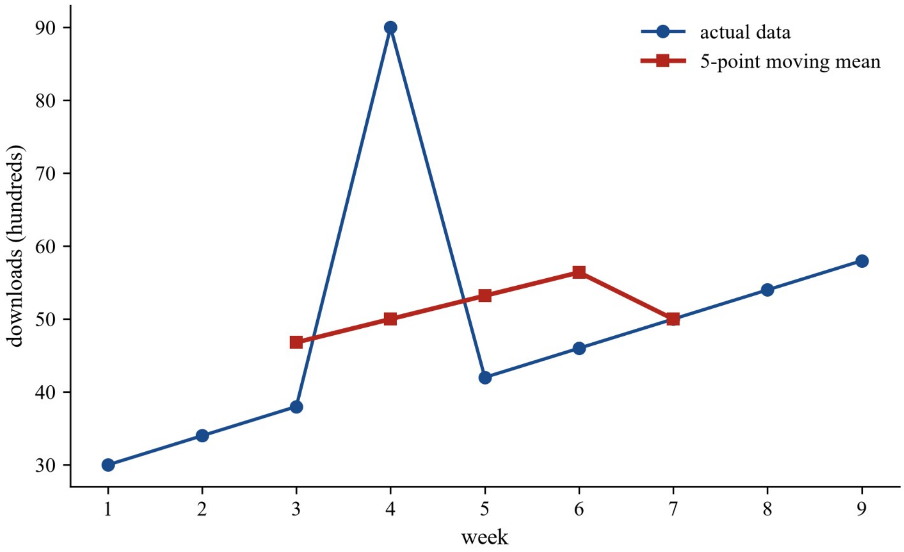
Q13. (a) A 3-point moving mean centred on week 4 equals 25. The values in weeks 3 and 5 were 22 and 26. Find the value recorded in week 4. (b) A 3-point moving mean centred on day 6 equals 40. The value on day 6 was 38 and the value on day 5 was 45. Find the value recorded on day 7.

Q14. Find the 3-point moving means of the series 30, 45, 33, 48, 36, 51, 39, and describe the trend.

Q15. Explain why a 4-point moving mean must be centred but a 3-point moving mean need not be. State what the centring step does.

Q16. The monthly sales of a store over eleven months were 22, 30, 26, 34, 28, 36, 30, 40, 34, 44, 38. Find the 7-point moving means (correct to one decimal place), and state how many points are lost at each end.

Q17. The weekly numbers of downloads (in hundreds) of an app over nine weeks are plotted below, together with the 5-point moving mean. A one-off media mention caused the spike in week 4.



(a) Identify the long-term trend, and the type of movement occurring in week 4.
(b) Explain how the 5-point moving mean helps reveal the trend.
(c) Comment on how well the moving mean deals with the week-4 spike.

Q18. The quarterly profit (in thousands of dollars) of a company over two years was 200, 260, 180, 300, 240, 320, 220, 360. Using the weighted formula $\frac{y_{t-2}+2y_{t-1}+2y_t+2y_{t+1}+y_{t+2}}{8}$, find the 4-point centred moving means for quarters 3 to 6.

Q19. The daily numbers of online orders taken by a business over nine days were 64, 58, 70, 60, 74, 64, 78, 68, 82. (a) Find the 5-point moving means. (b) On the same axes, plot the raw series and the smoothed series.

Q20. The quarterly numbers of visitors (in thousands) to a tourist attraction over two years were 18, 42, 30, 12, 24, 48, 36, 18. (a) Explain why a 4-point centred moving mean is an appropriate smoothing method here. (b) Find the smoothed values for quarters 3 to 6. (c) Describe the trend revealed, and explain how it differs from the raw pattern.

Answers

Q1. (a) Day 1 has no earlier neighbour, so a window of three centred on it cannot be formed. (b) 20. (c) 22, 26, 31, 32. (d) 20, 22, 26, 31, 32; increasing trend. **Q2.** (a) Weeks 1, 2, 8, 9. (b) 50. (c) 55, 55, 61, 62. (d) 50, 55, 55, 61, 62. **Q3.** (a) 19, 21, 24, 27, 29. (b), (c) 20, 22.5, 25.5, 28. **Q4.** (a) 66, 67.5, 69, 70.5, 72. (b) 66.75, 68.25, 69.75, 71.25. (c) $\frac{60+2(80)+2(72)+2(52)+66}{8}=\frac{534}{8}=66.75$. **Q5.** (a) 16, 20, 20, 24. (b) The smoothed values rise steadily, so the trend is increasing. **Q6.** (a) 17, 19, 17, 22, 20, 25. (b) See the plot in the solutions. **Q7.** 30, 30, 36, 36, 42. **Q8.** 16.8, 19.2, 19.8, 22.2, 22.8. **Q9.** 126.25, 128.75, 131.25, 133.75. **Q10.** 10.5, 12, 14, 16. **Q11.** (a) 56, 61, 59, 64, 62, 67, 65. (b) 58.2, 61.8, 61.2, 64.8, 64.2. (c) The 5-point series is smoother, because each value averages more observations, so more of the fluctuation cancels. **Q12.** (a) Seasonality. (b) 12 terms. (c) Yes; 12 is even, so the moving mean must be centred to align each value with a month. **Q13.** (a) 27. (b) 37. **Q14.** 36, 42, 39, 45, 42; the smoothed values rise overall, so the trend is increasing. **Q15.** A 3-point window has a middle value, so its mean already aligns with a data point. A 4-point window has no single middle, so its mean falls between two points; centring averages two neighbouring 4-point means to bring the value back onto a data point. **Q16.** 29.4, 32.0, 32.6, 35.1, 35.7; three points are lost at each end. **Q17.** (a) An increasing trend; week 4 is a one-off irregular fluctuation (an outlier). (b) Averaging each value with its neighbours cancels the week-to-week noise, so the rising trend stands out. (c) The moving mean dampens the spike but does not remove it; it spreads the extra downloads across weeks 3 to 6 (pulling those smoothed values up) and leaves a dip at week 7. **Q18.** 240, 252.5, 265, 277.5. **Q19.** (a) 65.2, 65.2, 69.2, 68.8, 73.2. (b) See the plot in the solutions. **Q20.** (a) The data are quarterly (a cycle of 4, which is even), so a 4-point centred moving mean averages one full year and removes the seasonal pattern. (b) 26.25, 27.75, 29.25, 30.75. (c) A steady upward trend, in contrast to the large seasonal swings of the raw data (a peak in the second quarter of each year and a trough in the fourth).

Fully-worked solutions

Q1. (a) A 3-point moving mean centred on day 1 would need a value on the day *before* day 1, which does not exist, so it cannot be formed. (b) Day 2: $\frac{24+15+21}{3}=\frac{60}{3}=20$. (c) Day 3: $\frac{15+21+30}{3}=\frac{66}{3}=22$; day 4: $\frac{21+30+27}{3}=\frac{78}{3}=26$; day 5: $\frac{30+27+36}{3}=\frac{93}{3}=31$; day 6: $\frac{27+36+33}{3}=\frac{96}{3}=32$. (d) Smoothed series 20, 22, 26, 31, 32; the values increase throughout, so the trend is increasing.

Q2. (a) A 5-point mean needs two values on each side, so the first two weeks (1, 2) and the last two weeks (8, 9) have no mean. (b) Week 3: $\frac{40+55+45+60+50}{5}=\frac{250}{5}=50$. (c) Week 4: $\frac{55+45+60+50+65}{5}=\frac{275}{5}=55$; week 5: $\frac{45+60+50+65+55}{5}=\frac{275}{5}=55$; week 6: $\frac{60+50+65+55+75}{5}=\frac{305}{5}=61$; week 7: $\frac{50+65+55+75+65}{5}=\frac{310}{5}=62$. (d) Smoothed series 50, 55, 55, 61, 62.

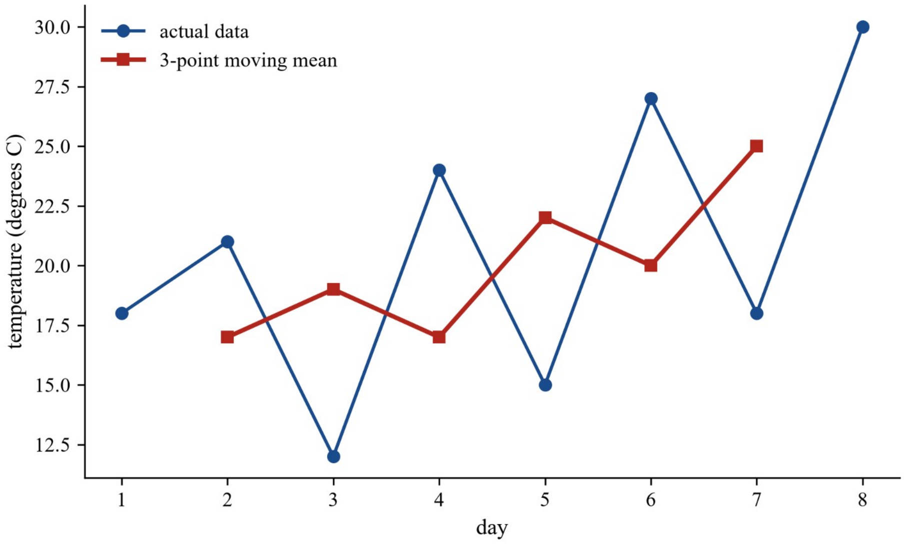
Q3. (a) The 2-point moving means are $\frac{16+22}{2}=19$, $\frac{22+20}{2}=21$, $\frac{20+28}{2}=24$, $\frac{28+26}{2}=27$, $\frac{26+32}{2}=29$. (b), (c) Centring averages neighbouring pairs: $\frac{19+21}{2}=20$, $\frac{21+24}{2}=22.5$, $\frac{24+27}{2}=25.5$, $\frac{27+29}{2}=28$, giving the centred values 20, 22.5, 25.5, 28 at times 2 to 5. (Equivalently, time 2 is $\frac{16+2(22)+20}{4}=\frac{80}{4}=20$.)

Q4. (a) The 4-point moving means are $\frac{60+80+72+52}{4}=66$, $\frac{80+72+52+66}{4}=67.5$, $\frac{72+52+66+86}{4}=69$, $\frac{52+66+86+78}{4}=70.5$, $\frac{66+86+78+58}{4}=72$. (b) Centring neighbouring pairs: quarter 3 $\frac{66+67.5}{2}=66.75$, quarter 4 $\frac{67.5+69}{2}=68.25$, quarter 5 $\frac{69+70.5}{2}=69.75$, quarter 6 $\frac{70.5+72}{2}=71.25$. (c)

$\dfrac{60+2(80)+2(72)+2(52)+66}{8}=\dfrac{60+160+144+104+66}{8}=\dfrac{534}{8}=66.75$, which agrees with the centred value for quarter 3.

Q5. (a) Month 2: $\dfrac{12+20+16}{3}=\dfrac{48}{3}=16$; month 3: $\dfrac{20+16+24}{3}=\dfrac{60}{3}=20$; month 4: $\dfrac{16+24+20}{3}=\dfrac{60}{3}=20$; month 5: $\dfrac{24+20+28}{3}=\dfrac{72}{3}=24$. (b) The smoothed values 16, 20, 20, 24 rise overall, so the underlying trend is increasing, even though the raw data go up and down from month to month.

Q6. (a) Day 2: $\dfrac{18+21+12}{3}=\dfrac{51}{3}=17$; day 3: $\dfrac{21+12+24}{3}=\dfrac{57}{3}=19$; day 4: $\dfrac{12+24+15}{3}=\dfrac{51}{3}=17$; day 5: $\dfrac{24+15+27}{3}=\dfrac{66}{3}=22$; day 6: $\dfrac{15+27+18}{3}=\dfrac{60}{3}=20$; day 7: $\dfrac{27+18+30}{3}=\dfrac{75}{3}=25$. (b) Plotting the smoothed series 17, 19, 17, 22, 20, 25 on the same axes as the raw data shows a gentle upward trend beneath the large day-to-day swings.



Q7. $\dfrac{30+24+36}{3}=30$, $\dfrac{24+36+30}{3}=30$, $\dfrac{36+30+42}{3}=36$, $\dfrac{30+42+36}{3}=36$, $\dfrac{42+36+48}{3}=42$. Smoothed series 30, 30, 36, 36, 42.

Q8. $\dfrac{12+18+15+21+18}{5}=\dfrac{84}{5}=16.8$, $\dfrac{18+15+21+18+24}{5}=\dfrac{96}{5}=19.2$, $\dfrac{15+21+18+24+21}{5}=\dfrac{99}{5}=19.8$, $\dfrac{21+18+24+21+27}{5}=\dfrac{111}{5}=22.2$, $\dfrac{18+24+21+27+24}{5}=\dfrac{114}{5}=22.8$. Smoothed series 16.8, 19.2, 19.8, 22.2, 22.8.

Q9. The 4-point moving means are $\dfrac{120+90+150+140}{4}=125$, $\dfrac{90+150+140+130}{4}=127.5$, $\dfrac{150+140+130+100}{4}=130$, $\dfrac{140+130+100+160}{4}=132.5$, $\dfrac{130+100+160+150}{4}=135$. Centring neighbouring pairs gives quarter 3 $\dfrac{125+127.5}{2}=126.25$, quarter 4 $\dfrac{127.5+130}{2}=128.75$, quarter 5 $\dfrac{130+132.5}{2}=131.25$, quarter 6 $\dfrac{132.5+135}{2}=133.75$.

Q10. The 2-point moving means are $\dfrac{8+12}{2}=10$, $\dfrac{12+10}{2}=11$, $\dfrac{10+16}{2}=13$, $\dfrac{16+14}{2}=15$, $\dfrac{14+20}{2}=17$. Centring neighbouring pairs: $\dfrac{10+11}{2}=10.5$, $\dfrac{11+13}{2}=12$, $\dfrac{13+15}{2}=14$, $\dfrac{15+17}{2}=16$. Centred series 10.5, 12, 14, 16.

Q11. (a) 3-point means: $\frac{51+63+54}{3}=56$, $\frac{63+54+66}{3}=61$, $\frac{54+66+57}{3}=59$, $\frac{66+57+69}{3}=64$, $\frac{57+69+60}{3}=62$, $\frac{69+60+72}{3}=67$, $\frac{60+72+63}{3}=65$, giving 56, 61, 59, 64, 62, 67, 65. (b) 5-point means: $\frac{51+63+54+66+57}{5}=58.2$, $\frac{63+54+66+57+69}{5}=61.8$, $\frac{54+66+57+69+60}{5}=61.2$, $\frac{66+57+69+60+72}{5}=64.8$, $\frac{57+69+60+72+63}{5}=64.2$, giving 58.2, 61.8, 61.2, 64.8, 64.2. (c) The 5-point series is smoother: the 3-point values still swing by up to about 5 (for example 59 up to 64), whereas the 5-point values change by at most about 3.6, because averaging five values cancels more of the fluctuation than averaging three.

Q12. (a) The repeating within-year pattern is **seasonality**. (b) There are 12 months in the repeating cycle, so a **12-point** moving mean averages exactly one full cycle and removes it. (c) Yes: 12 is even, so a plain 12-point mean would fall between two months; it must be **centred** so that each smoothed value aligns with a month.

Q13. (a) Let the week-4 value be y. Then $\frac{22+y+26}{3}=25$, so $22+y+26=75$ and $y=75-48=27$. (b) Let the day-7 value be z. Then $\frac{45+38+z}{3}=40$, so $45+38+z=120$ and $z=120-83=37$.

Q14. $\frac{30+45+33}{3}=36$, $\frac{45+33+48}{3}=42$, $\frac{33+48+36}{3}=39$, $\frac{48+36+51}{3}=45$, $\frac{36+51+39}{3}=42$, giving 36, 42, 39, 45, 42. The smoothed values rise overall (from 36 to around 42–45), so the trend is increasing, although the raw data alternate up and down.

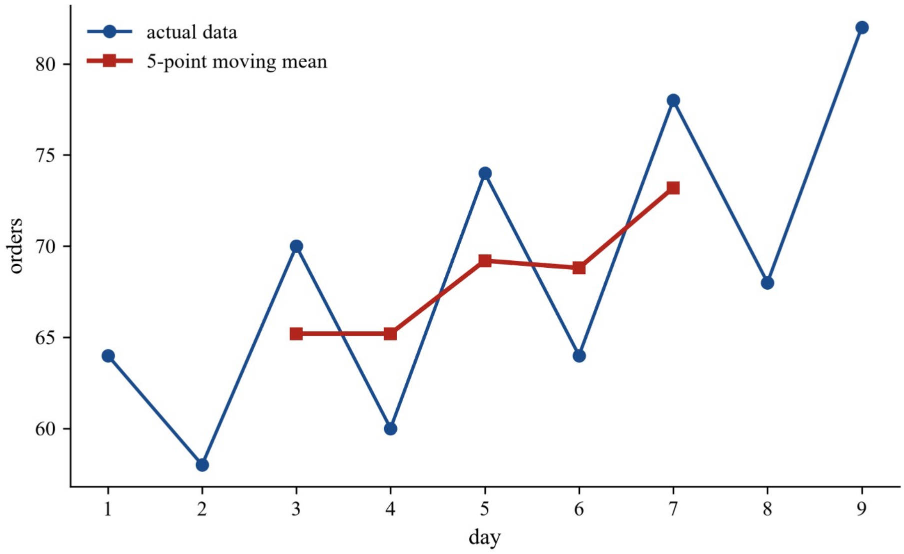
Q15. A 3-point window covers three consecutive values and has a single middle value, so its mean already lines up with a data point — no adjustment is needed. A 4-point window covers four consecutive values and has no single middle point, so its mean falls halfway between the two central time points and does **not** line up with any data value. The **centring** step averages two neighbouring 4-point means, which places the smoothed value back onto a data point.

Q16. A 7-point mean needs three values on each side, so it can be found for months 4 to 8. Month 4: $\frac{22+30+26+34+28+36+30}{7}=\frac{206}{7}\approx 29.4$; month 5: $\frac{30+26+34+28+36+30+40}{7}=\frac{224}{7}=32.0$; month 6: $\frac{26+34+28+36+30+40+34}{7}=\frac{228}{7}\approx 32.6$; month 7: $\frac{34+28+36+30+40+34+44}{7}=\frac{246}{7}\approx 35.1$; month 8: $\frac{28+36+30+40+34+44+38}{7}=\frac{250}{7}\approx 35.7$. Smoothed series 29.4, 32.0, 32.6, 35.1, 35.7; three points are lost at each end.

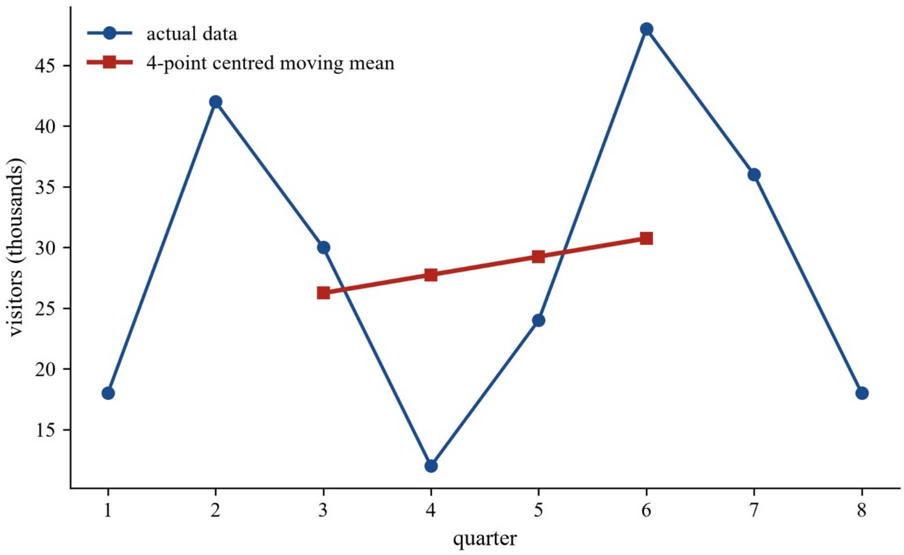
Q17. (a) The smoothed (red) series climbs steadily from left to right, so the long-term trend is **increasing**. The sharp week-4 peak, well above its neighbours, is a one-off **irregular fluctuation** (an outlier) caused by the media mention, not part of the trend. (b) Each 5-point mean averages a value with its two neighbours on each side, so the week-to-week noise largely cancels and the steady rise underneath becomes visible as a smooth line. (c) A moving mean is not resistant to an outlier: the large week-4 value is included in every window that reaches it, so it lifts the smoothed values for weeks 3 to 6 above the true trend and then, once it leaves the window, the smoothed value for week 7 dips. The spike is damped but its influence is spread across several smoothed points rather than removed — a moving **median** would handle it better.

Q18. Using the weighted formula for each quarter, quarter 3: $\frac{200+2(260)+2(180)+2(300)+240}{8}=\frac{1920}{8}=240$; quarter 4: $\frac{260+2(180)+2(300)+2(240)+320}{8}=\frac{2020}{8}=252.5$; quarter 5: $\frac{180+2(300)+2(240)+2(320)+220}{8}=\frac{2120}{8}=265$; quarter 6: $\frac{300+2(240)+2(320)+2(220)+360}{8}=\frac{2220}{8}=277.5$. Smoothed series 240, 252.5, 265, 277.5.

Q19. (a) $\frac{64+58+70+60+74}{5}=\frac{326}{5}=65.2$, $\frac{58+70+60+74+64}{5}=\frac{326}{5}=65.2$, $\frac{70+60+74+64+78}{5}=\frac{346}{5}=69.2$, $\frac{60+74+64+78+68}{5}=\frac{344}{5}=68.8$, $\frac{74+64+78+68+82}{5}=\frac{366}{5}=73.2$, giving 65.2, 65.2, 69.2, 68.8, 73.2. (b) Plotting these on the same axes as the raw series shows a rising trend beneath the day-to-day zig-zag.



Q20. (a) The data are collected each quarter and repeat over a one-year cycle of four quarters. Because 4 is even, a 4-point **centred** moving mean averages exactly one full year (so the seasonal high and low cancel) and lands each smoothed value on a quarter. (b) The 4-point moving means are $\frac{18+42+30+12}{4}=25.5$, $\frac{42+30+12+24}{4}=27$, $\frac{30+12+24+48}{4}=28.5$, $\frac{12+24+48+36}{4}=30$, $\frac{24+48+36+18}{4}=31.5$; centring neighbouring pairs gives quarter 3 $\frac{25.5+27}{2}=26.25$, quarter 4 $\frac{27+28.5}{2}=27.75$, quarter 5 $\frac{28.5+30}{2}=29.25$, quarter 6 $\frac{30+31.5}{2}=30.75$. (c) The smoothed values 26.25, 27.75, 29.25, 30.75 rise steadily, so visitor numbers are trending upward from year to year. This gentle rise is hidden in the raw data by the large seasonal swings — a strong peak in the second quarter of each year (42 and 48) and a deep trough in the fourth (12 and 18) — which the centring has averaged away.



Theory

A time series plot often shows so much short-term **irregular fluctuation** that the **trend** — the long-term direction of the data — is hard to see. **Smoothing** replaces the jagged raw series with a steadier one in which the fluctuations are damped, so that the trend stands out. One way to smooth is with **moving means** (an earlier section). This section develops the other standard method: **moving medians**.

A **moving median** replaces each data value with the **median of a small window of consecutive values centred on it**. Because a median is used, the method requires an **odd number of points** in each window (3, 5, 7, …): with an odd number of values there is a single **middle time period** to attach the smoothed value to, and the median is simply the middle value of the window when its values are put in order.

Three-point moving median.

For a **3-point moving median**, each smoothed value is the median of three consecutive values — the value itself together with its immediate neighbour on each side — and it is recorded against the **middle** of those three periods. For example, for the series 8, 14, 10, 12, … the first smoothed value uses 8, 14, 10; ordered these are 8, 10, 14, so the median is 10, recorded against the **second** period. The window then slides one period to the right and the process repeats.

Five-point moving median.

A **5-point moving median** uses five consecutive values and records their median against the **middle (third)** period of the five. Using more points removes more of the fluctuation, so a 5-point smooth is steadier than a 3-point smooth, but it also loses more of the series at each end.

How many points are lost.

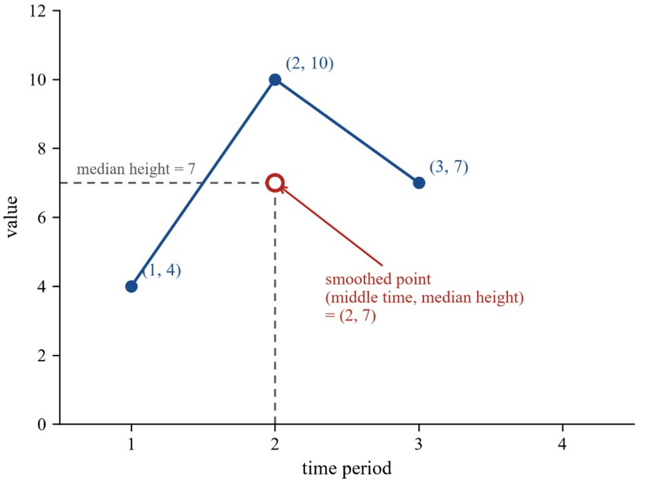
A smoothed value can only be recorded where a full window fits, so the ends of the series are lost. A 3-point moving median loses **one** value at each end (two in total); a 5-point moving median loses **two** at each end (four in total). In general a $(2k+1)$**-point** moving median loses k values at each end, so a series of n values produces $n-2k$ smoothed values.

Reading moving medians off the graph (graphical smoothing).

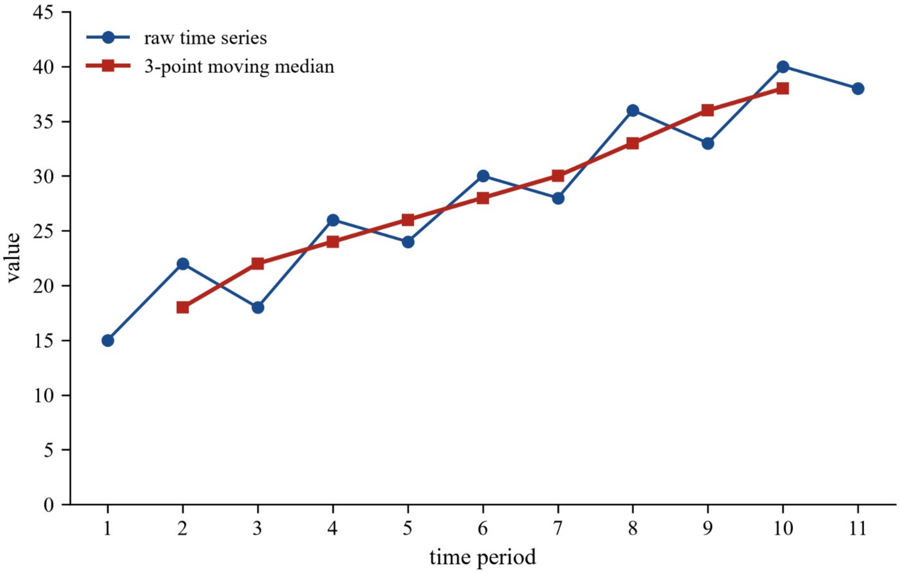
Moving medians are especially useful because they can be found **directly from the time series plot**, with no arithmetic. Consider any window of an odd number of consecutive plotted points. The smoothed point sits at:

- the **middle time** of the window (the horizontal position of the central point), and
- the **median height** of the window (the middle value of the points, read off vertically).

Because the window has an odd number of points, the median height is **one of the actual plotted heights** — so you simply pick out, by eye, the point of middle height and read its value, then plot it at the middle time. The figure below shows the construction for one window of three points: the point at $(2,10)$ is a spike, but the median of the three heights $4,10,7$ is 7, so the smoothed point is plotted at $(2,7)$.



Marking the smoothed point for each window and joining the results with line segments produces the smoothed series. In the plot below, the raw series fluctuates up and down, but its 3-point moving-median smooth rises steadily, revealing a clear upward trend.



Why use a median rather than a mean.

The median of a window depends only on the middle value, so a single extreme value — a **possible outlier** caused by a one-off event — does **not** distort the smoothed value: the moving median is **resistant** to outliers. A moving mean, by contrast, uses every value in the window, so an outlier drags the smoothed value towards it and creates a false bump in the trend. When a time series contains an outlier or very large fluctuations, a moving median usually gives the more faithful picture of the trend. (Moving medians do **not** remove seasonal, calendar-based movements — that is a separate task, addressed later in this chapter.)

Using CAS. General Mathematics is a technology-active, open-book course. In practice the raw values are entered into a CAS calculator or spreadsheet, which can compute the moving medians and plot the smoothed series. The by-hand and graphical methods below let you understand what the smoothing does and check that the technology output is reasonable.

Worked examples

Example 1 — scaffolded

The weekly numbers of items returned to a store over seven weeks were 12, 18, 15, 20, 22, 19, 25. Smooth the series using a 3-point moving median.

Working	Comment
Weeks 1–3: 12, 18, 15 → ordered 12, 15, 18; median $= 15$.	Record 15 against week 2, the middle of the three.
Weeks 2–4: 18, 15, 20 → ordered 15, 18, 20; median $= 18$.	Slide the window one week to the right.
Weeks 3–5: 15, 20, 22 → ordered 15, 20, 22; median $= 20$.	Record against week 4.
Weeks 4–6: 20, 22, 19 → ordered 19, 20, 22; median $= 20$.	Record against week 5.
Weeks 5–7: 22, 19, 25 → ordered 19, 22, 25; median $= 22$.	Record against week 6.

Weeks 1 and 7 have no smoothed value, since a full three-week window does not fit at the ends.

week	value	3-point moving median
1	12	—
2	18	15
3	15	18
4	20	20
5	22	20
6	19	22
7	25	—

The smoothed values 15, 18, 20, 20, 22 increase steadily, so the series has an upward trend.

Example 2 — scaffolded

The values of a time series over nine periods were 10, 14, 12, 18, 16, 22, 20, 26, 24. Smooth the series using a 5-point moving median.

Working	Comment
Periods 1–5: 10, 14, 12, 18, 16 → ordered 10, 12, 14, 16, 18; median $= 14$.	Record 14 against period 3, the middle of the five.
Periods 2–6: 14, 12, 18, 16, 22 → ordered 12, 14, 16, 18, 22; median $= 16$.	Record against period 4.
Periods 3–7: 12, 18, 16, 22, 20 → ordered 12, 16, 18, 20, 22; median $= 18$.	Record against period 5.
Periods 4–8: 18, 16, 22, 20, 26 → ordered 16, 18, 20, 22, 26; median $= 20$.	Record against period 6.
Periods 5–9: 16, 22, 20, 26, 24 → ordered 16, 20, 22, 24, 26; median $= 22$.	Record against period 7.

The 5-point moving median loses **two** values at each end, so only periods 3 to 7 have smoothed values: 14, 16, 18, 20, 22. The smooth is a straight rising line, showing a strong upward trend.

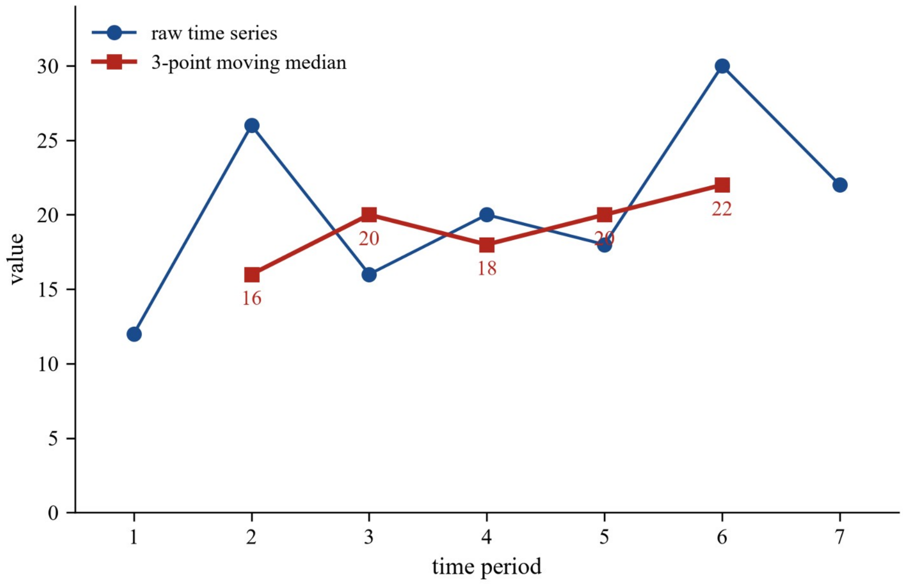
Example 3 — scaffolded

The time series plot below shows the values 12, 26, 16, 20, 18, 30, 22 recorded over seven periods. Use a **graphical** 3-point moving median to smooth the series, reading each smoothed value directly from the plot.

Working	Comment
Periods 1–3 have heights 12, 26, 16; the middle height is	The spike at period 2 (height 26) is ignored; plot 16 at

Working	Comment
16.	period 2.
Periods 2–4 have heights 26, 16, 20; the middle height is 20.	Plot 20 at period 3.
Periods 3–5 have heights 16, 20, 18; the middle height is 18.	Plot 18 at period 4.
Periods 4–6 have heights 20, 18, 30; the middle height is 20.	The spike at period 6 (30) is ignored; plot 20 at period 5.
Periods 5–7 have heights 18, 30, 22; the middle height is 22.	Plot 22 at period 6.

The smoothed points, plotted at the middle times and joined, are 16, 20, 18, 20, 22 for periods 2 to 6. Notice that the two spikes (26 and 30) have been pulled down to the level of their neighbours — the moving median has smoothed them away.



Example 4 — unscaffolded

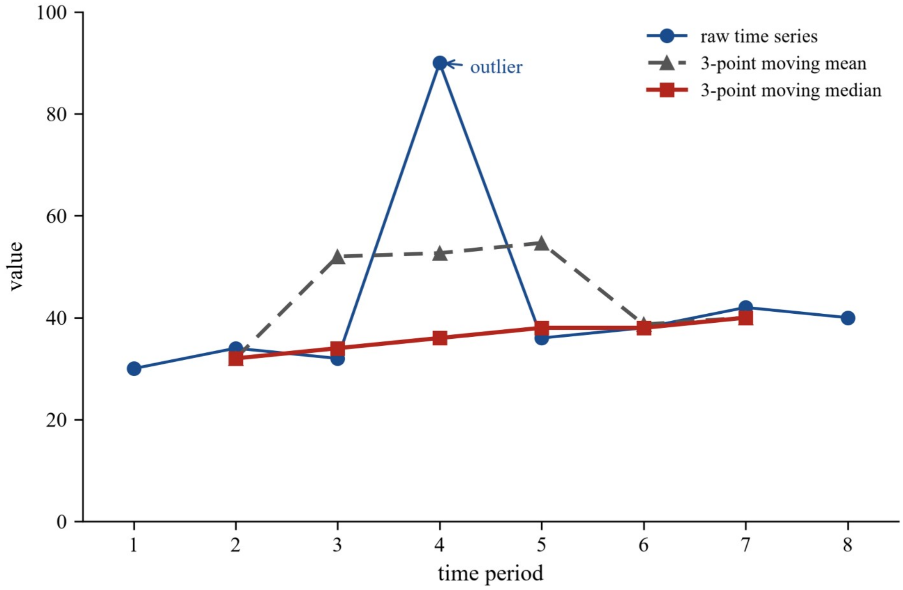
A shop's monthly sales (in \$000) over eight months were 30, 34, 32, 90, 36, 38, 42, 40, where the value in month 4 is an outlier caused by a one-off clearance event. Smooth the series with a 3-point moving mean and with a 3-point moving median, and comment on which better shows the trend.

The two smoothed series are set out below (moving means to one decimal place).

month	value	3-point moving mean	3-point moving median
1	30	—	—
2	34	32.0	32
3	32	52.0	34
4	90	52.7	36
5	36	54.7	38
6	38	38.7	38
7	42	40.0	40
8	40	—	—

The moving **mean** is dragged upwards by the outlier: near month 4 it jumps to about 52–55, inventing a large hump that is not part of the true trend. The moving **median** ignores the outlier — for each window it keeps only the middle value — so it rises smoothly (32, 34, 36, 38, 38, 40) and follows the genuine gentle upward trend.

∴ because the series contains an outlier, the **moving median** gives the more faithful smooth.



Example 5 — unscaffolded

The quarterly values of a time series were 22, 20, 24, 60, 26, 28, 30, 32, where the value 60 is an outlier. Smooth the series with a 5-point moving median, state how many smoothed values there are, and comment on the effect of the outlier.

The 5-point windows and their medians are:

periods	ordered values	median
1–5	20, 22, 24, 26, **60**	24
2–6	20, 24, 26, 28, **60**	26
3–7	24, 26, 28, 30, **60**	28
4–8	26, 28, 30, 32, **60**	30

A 5-point moving median loses two values at each end, so from eight values there are $8 - 4 = 4$ smoothed values, recorded against periods 3 to 6: 24, 26, 28, 30. In every window the outlier 60 is the largest value and so is never the median; it has **no effect at all** on the smoothed series, which rises steadily by 2 each quarter.

∴ the smoothed values are 24, 26, 28, 30 (periods 3–6), and the outlier does not distort them.

Practice questions

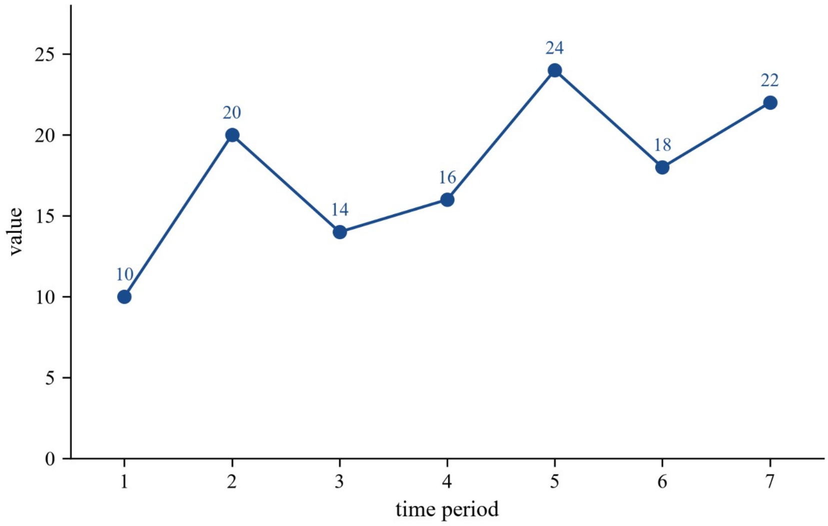
Scaffolded

Q1. The values of a time series over seven periods were 8, 11, 9, 14, 12, 17, 15. (a) Find the 3-point moving median for period 2 (use periods 1, 2, 3). (b) Find the 3-point moving medians for periods 3, 4, 5 and 6. (c) Explain why periods 1 and 7 have no smoothed value. (d) State whether the smoothed series is increasing or decreasing.

Q2. A time series has values 40, 35, 38, 30, 34, 28, 32 over seven periods. (a) Complete a 3-point moving median smooth. (b) State how many values are lost at each end. (c) Describe the trend shown by the smoothed series.

Q3. The values of a time series over nine periods were 5, 9, 7, 13, 11, 17, 15, 21, 19. (a) Find the 5-point moving median for period 3 (use periods 1 to 5). (b) Find the remaining 5-point moving medians. (c) State which periods have a smoothed value and how many values are lost in total.

Q4. The time series plot below shows values recorded over seven periods.



(a) Read the three heights in periods 1–3 and write down their median.

(b) Repeat for the windows 2–4, 3–5, 4–6 and 5–7 to complete a graphical 3-point moving median.

(c) State the two periods whose spikes are smoothed away.

Q5. A time series with a one-off outlier has values 15, 18, 16, 50, 20, 22, 19. (a) Find the 3-point moving **mean** for each of periods 2 to 6 (to one decimal place). (b) Find the 3-point moving **median** for each of periods 2 to 6. (c) State which smooth is affected by the outlier and which is not.

Q6. A time series has 20 values. (a) How many smoothed values result from a 3-point moving median, and how many values are lost? (b) Answer part (a) for a 5-point moving median. (c) Answer part (a) for a 7-point moving median. (d) How many smoothed values result from a 5-point moving median applied to 12 values?

Unscaffolded

Q7. Smooth the time series 23, 27, 25, 31, 29, 35, 33 using a 3-point moving median.

Q8. Smooth the time series 4, 10, 6, 14, 8, 18, 12, 22, 16 using a 5-point moving median.

Q9. A time series has values 50, 44, 47, 41, 44, 38, 41. Find its 3-point moving median and describe the trend.

Q10. The monthly values 60, 64, 62, 110, 66, 68, 70 include an outlier in month 4. Smooth the series with a 3-point moving median and explain what happens to the outlier.

Q11. A time series has values 8, 9, 7, 40, 10, 11, 9. (a) Find the 3-point moving mean (to one decimal place) and the 3-point moving median. (b) State, with a reason, which smooth you would report.

Q12. A time series has values 30, 42, 34, 46, 38, 50, 42, 54 over eight periods. Find its 5-point moving median, state which periods carry a smoothed value, and say how many values are lost.

Q13. Smooth the time series 25, 28, 26, 29, 80, 31, 33, 30 using a 3-point moving median, and comment on the value 80.

Q14. (a) Explain why a moving median must use an odd number of points. (b) Explain why a moving median is described as resistant to outliers.

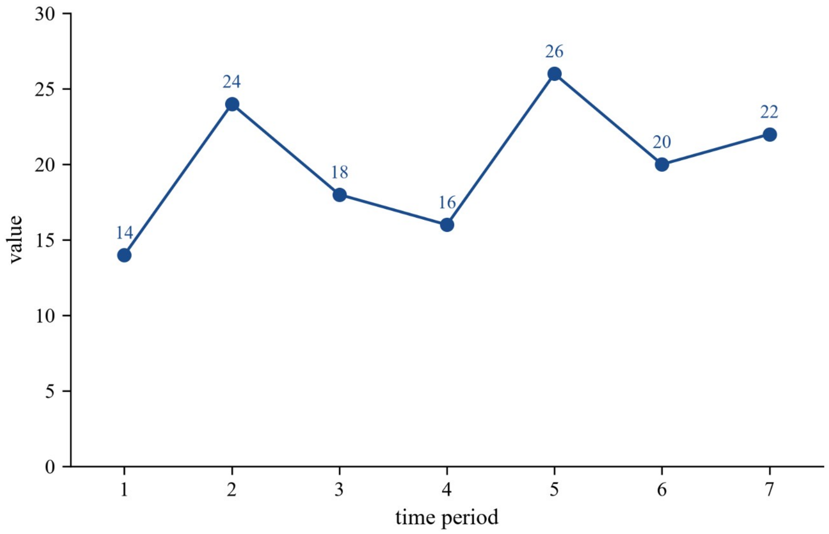
Q15. A time series has values 10, 16, 12, 20, 14, 22, 16, 24, 18. (a) Find the 3-point moving median. (b) Find the 5-point moving median. (c) State which smooth is steadier, and why.

Q16. Part of a 3-point moving median smooth of the series 45, 50, 47, 53, 49, 55, 51 has been completed below. Find the missing smoothed values p, q and r.

period	1	2	3	4	5	6	7
value	45	50	47	53	49	55	51
smoothe	—	47	p	49	q	r	—

period	1	2	3	4	5	6	7
d							

Q17. The time series plot below shows values recorded over seven periods.



Use a graphical 3-point moving median to smooth the series, and describe the trend revealed.

Q18. The monthly memberships of a gym over seven months were 200, 240, 220, 280, 260, 320, 300. Smooth the series with a 3-point moving median and describe the trend.

Q19. A time series has values 12, 18, 14, 16, 22, 18, 20, 26, 22. Find its 5-point moving median and state how many smoothed values there are.

Q20. (a) State one advantage of a moving median over a moving mean for smoothing a series that contains an outlier. (b) State how many values a 3-point moving median and a 5-point moving median each lose at every end of a series.

Answers

Q1. (a) 9; (b) period 3: 11, period 4: 12, period 5: 14, period 6: 15; (c) a full three-period window does not fit at the ends; (d) increasing. **Q2.** (a) 38, 35, 34, 30, 32 (periods 2–6); (b) one value at each end; (c) a downward (decreasing) trend. **Q3.** (a) 9; (b) 11, 13, 15, 17; (c) periods 3–7 have smoothed values; four values are lost (two at each end). **Q4.** (a) 14; (b) 14, 16, 16, 18, 22 (periods 2–6); (c) the spikes at periods 2 (height 20) and 5 (height 24). **Q5.** (a) 16.3, 28.0, 28.7, 30.7, 20.3; (b) 16, 18, 20, 22, 20; (c) the moving mean is affected by the outlier (50), the moving median is not. **Q6.** (a) 18 smoothed, 2 lost; (b) 16 smoothed, 4 lost; (c) 14 smoothed, 6 lost; (d) 8 smoothed values. **Q7.** 25, 27, 29, 31, 33 (periods 2–6). **Q8.** 8, 10, 12, 14, 16 (periods 3–7). **Q9.** 47, 44, 44, 41, 41 (periods 2–6); a downward trend. **Q10.** 62, 64, 66, 68, 68 (months 2–6); the outlier 110 is never the middle value, so it is smoothed away. **Q11.** (a) mean 8.0, 18.7, 19.0, 20.3, 10.0; median 8, 9, 10, 11, 10; (b) the median, because the value 40 is an outlier that distorts the mean. **Q12.** 38, 42, 42, 46 (periods 3–6); four values lost (two at each end). **Q13.** 26, 28, 29, 31, 33, 31 (periods 2–7); the outlier 80 is never the median, so it does not affect the smooth. **Q14.** (a) so that each window has a single middle period to record the median against; (b) the median depends only on the middle value, so an outlier in the window does not change it. **Q15.** (a) 12, 16, 14, 20, 16, 22, 18 (periods 2–8); (b) 14, 16, 16, 20, 18 (periods 3–7); (c) the 5-point smooth is steadier, because a larger window removes more of the fluctuation. **Q16.** $p=50$, $q=53$, $r=51$. **Q17.** 18, 18, 18, 20, 22 (periods 2–6); a rising trend (after a flat start). **Q18.** 220, 240, 260, 280, 300 (months 2–6); a steady upward trend. **Q19.** 16, 18, 18, 20, 22 (periods 3–7); five smoothed values. **Q20.** (a) the moving median is resistant, so an outlier does not distort the smoothed trend; (b) a 3-point loses one at each end, a 5-point loses two at each end.

Fully-worked solutions

Q1. (a) Periods 1–3 are 8, 11, 9; ordered 8, 9, 11, so the median is 9. (b) Periods 2–4: 11, 9, 14 → 9, 11, 14, median 11; periods 3–5: 9, 14, 12 → 9, 12, 14, median 12; periods 4–6: 14, 12, 17 → 12, 14, 17, median 14; periods 5–7: 12, 17, 15 → 12, 15, 17, median 15. (c) At period 1 there is no earlier value and at period 7 no later value, so a full three-period window cannot be formed. (d) The smoothed values 9, 11, 12, 14, 15 increase, so the trend is increasing.

Q2. (a) Periods 1–3: 40, 35, 38 → 35, 38, 40, median 38; periods 2–4: 35, 38, 30 → 30, 35, 38, median 35; periods 3–5: 38, 30, 34 → 30, 34, 38, median 34; periods 4–6: 30, 34, 28 → 28, 30, 34, median 30; periods 5–7: 34, 28, 32 → 28, 32, 34, median 32. The smooth is 38, 35, 34, 30, 32 for periods 2–6. (b) One value is lost at each end. (c) The smoothed values fall from 38 to about 30, so the trend is downward.

Q3. (a) Periods 1–5 are 5, 9, 7, 13, 11; ordered 5, 7, 9, 11, 13, so the median is 9. (b) Periods 2–6: 9, 7, 13, 11, 17 → 7, 9, 11, 13, 17, median 11; periods 3–7: 7, 13, 11, 17, 15 → 7, 11, 13, 15, 17, median 13; periods 4–8: 13, 11, 17, 15, 21 → 11, 13, 15, 17, 21, median 15; periods 5–9: 11, 17, 15, 21, 19 → 11, 15, 17, 19, 21, median 17. (c) The smoothed values 9, 11, 13, 15, 17 sit against periods 3–7; two values are lost at each end, four in total.

Q4. Reading heights 10, 20, 14, 16, 24, 18, 22 from the plot: (a) periods 1–3 are 10, 20, 14; the median is 14. (b) Periods 2–4: 20, 14, 16 → median 16; periods 3–5: 14, 16, 24 → median 16; periods 4–6: 16, 24, 18 → median 18; periods 5–7: 24, 18, 22 → median 22. The graphical smooth is 14, 16, 16, 18, 22 for periods 2–6. (c) The spikes at period 2 (height 20) and period 5 (height 24) are each higher than both neighbours, so they are never the median and are smoothed away.

Q5. (a) Moving means: $(15+18+16)/3=16.3$, $(18+16+50)/3=28.0$, $(16+50+20)/3=28.7$, $(50+20+22)/3=30.7$, $(20+22+19)/3=20.3$. (b) Moving medians: median$(15,18,16)=16$, median $(18,16,50)=18$, median$(16,50,20)=20$, median$(50,20,22)=22$, median$(20,22,19)=20$; that is 16, 18, 20, 22, 20. (c) The moving mean jumps to about 28–31 around the outlier 50, so it is affected by it; the moving median (16, 18, 20, 22, 20) never uses the outlier as its middle value, so it is unaffected.

Q6. A $(2k+1)$-point moving median loses k values at each end. (a) A 3-point median has $k=1$: $20-2=18$ smoothed values, 2 lost. (b) A 5-point median has $k=2$: $20-4=16$ smoothed values, 4 lost. (c) A 7-point median has $k=3$: $20-6=14$ smoothed values, 6 lost. (d) A 5-point median on 12 values gives $12-4=8$ smoothed values.

Q7. Periods 1–3: 23, 27, 25 → median 25; 2–4: 27, 25, 31 → median 27; 3–5: 25, 31, 29 → median 29; 4–6: 31, 29, 35 → median 31; 5–7: 29, 35, 33 → median 33. The 3-point moving median is 25, 27, 29, 31, 33 (periods 2–6).

Q8. Periods 1–5: 4, 10, 6, 14, 8 → ordered 4, 6, 8, 10, 14, median 8; 2–6: 10, 6, 14, 8, 18 → 6, 8, 10, 14, 18, median 10; 3–7: 6, 14, 8, 18, 12 → 6, 8, 12, 14, 18, median 12; 4–8: 14, 8, 18, 12, 22 → 8, 12, 14, 18, 22, median 14; 5–9: 8, 18, 12, 22, 16 → 8, 12, 16, 18, 22, median 16. The 5-point moving median is 8, 10, 12, 14, 16 (periods 3–7).

Q9. Periods 1–3: 50, 44, 47 → median 47; 2–4: 44, 47, 41 → median 44; 3–5: 47, 41, 44 → median 44; 4–6: 41, 44, 38 → median 41; 5–7: 44, 38, 41 → median 41. The smooth is 47, 44, 44, 41, 41 (periods 2–6), which falls steadily — a downward trend.

Q10. Months 1–3: 60, 64, 62 → median 62; 2–4: 64, 62, 110 → median 64; 3–5: 62, 110, 66 → median 66; 4–6: 110, 66, 68 → median 68; 5–7: 66, 68, 70 → median 68. The smooth is 62, 64, 66, 68, 68 (months 2–6). In every window the outlier 110 is the largest value, so it is never the median; the moving median therefore smooths it away and shows a steady upward trend.

Q11. (a) Moving means: $(8+9+7)/3=8.0$, $(9+7+40)/3=18.7$, $(7+40+10)/3=19.0$, $(40+10+11)/3=20.3$, $(10+11+9)/3=10.0$. Moving medians: median$(8,9,7)=8$, median$(9,7,40)=9$, median$(7,40,10)=10$, median$(40,10,11)=11$, median$(10,11,9)=10$; that is 8, 9, 10, 11, 10. (b) The value 40 is an outlier. It inflates the moving mean to about 19–20 in the middle, creating a false hump, but leaves the moving median unchanged, so report the **moving median** 8, 9, 10, 11, 10.

Q12. Periods 1–5: 30, 42, 34, 46, 38 → ordered 30, 34, 38, 42, 46, median 38; 2–6: 42, 34, 46, 38, 50 → 34, 38, 42, 46, 50, median 42; 3–7: 34, 46, 38, 50, 42 → 34, 38, 42, 46, 50, median 42; 4–8: 46, 38, 50, 42, 54 → 38, 42, 46, 50, 54, median 46. The 5-point moving median is 38, 42, 42, 46, recorded against periods 3–6; four values are lost (two at each end).

Q13. Periods 1–3: 25, 28, 26 → median 26; 2–4: 28, 26, 29 → median 28; 3–5: 26, 29, 80 → median 29; 4–6: 29, 80, 31 → median 31; 5–7: 80, 31, 33 → median 33; 6–8: 31, 33, 30 → median 31. The smooth is 26, 28, 29, 31, 33, 31 (periods 2–7). The value 80 is an outlier; in each window that contains it, it is the largest value and so never the median, so it does not affect the smoothed series.

Q14. (a) The smoothed value for a window is placed against the window's middle period. With an **odd** number of points there is exactly one middle period; with an even number the middle would fall between two periods, so there would be nowhere to record the value (an even-order smooth requires an extra centring step, which belongs with moving means). (b) The median of a window is its middle value once the values are ordered. Changing an extreme value — an outlier — does not change which value is in the middle, so the median, and hence the moving-median smooth, is resistant to outliers.

Q15. (a) 3-point: median$(10,16,12)=12$, median$(16,12,20)=16$, median$(12,20,14)=14$, median$(20,14,22)=20$, median$(14,22,16)=16$, median$(22,16,24)=22$, median$(16,24,18)=18$; that is 12, 16, 14, 20, 16, 22, 18 (periods 2–8). (b) 5-point: median$(10,16,12,20,14)=14$, median$(16,12,20,14,22)=16$, median$(12,20,14,22,16)=16$, median$(20,14,22,16,24)=20$, median$(14,22,16,24,18)=18$; that is 14, 16, 16, 20, 18 (periods 3–7). (c) The 5-point smooth varies less from value to value, so it is steadier; a larger window averages out more of the short-term fluctuation.

Q16. Each smoothed value is the median of three consecutive raw values. p (period 3) is median$(50,47,53)=50$; q (period 5) is median$(53,49,55)=53$; r (period 6) is median$(49,55,51)=51$. So $p=50$, $q=53$, $r=51$.

Q17. Reading heights 14, 24, 18, 16, 26, 20, 22 from the plot: periods 1–3: 14, 24, 18 → median 18; 2–4: 24, 18, 16 → median 18; 3–5: 18, 16, 26 → median 18; 4–6: 16, 26, 20 → median 20; 5–7: 26, 20, 22 → median 22. The graphical smooth is 18, 18, 18, 20, 22 (periods 2–6): after a flat start it rises, revealing a gently increasing trend once the spikes at periods 2 and 5 are removed.

Q18. Months 1–3: 200, 240, 220 → median 220; 2–4: 240, 220, 280 → median 240; 3–5: 220, 280, 260 → median 260; 4–6: 280, 260, 320 → median 280; 5–7: 260, 320, 300 → median 300. The 3-point moving median is 220, 240, 260, 280, 300 (months 2–6) — a steady upward trend of about 20 new members per month.

Q19. Periods 1–5: 12, 18, 14, 16, 22 → ordered 12, 14, 16, 18, 22, median 16; 2–6: 18, 14, 16, 22, 18 → 14, 16, 18, 18, 22, median 18; 3–7: 14, 16, 22, 18, 20 → 14, 16, 18, 20, 22, median 18; 4–8: 16, 22, 18, 20, 26 → 16, 18, 20, 22, 26, median 20; 5–9: 22, 18, 20, 26, 22 → 18, 20, 22, 22, 26, median 22. The 5-point moving median is 16, 18, 18, 20, 22 (periods 3–7): five smoothed values.

Q20. (a) The moving median uses only the middle value of each window, so a single outlier — for example a one-off spike — is never used and does not distort the smoothed trend; a moving mean uses every value and is pulled towards the outlier. (b) A 3-point moving median loses one value at each end (two in total); a 5-point moving median loses two values at each end (four in total).

Theory

Many time series repeat a **seasonal pattern**: the figures rise and fall in a regular way that is tied to the calendar — the four quarters of a year, the twelve months, or the seven days of a week. Retail sales peak before the December holidays, electricity demand climbs every summer, and a café is busiest each weekend. To compare one period fairly with another, and to see the underlying **trend** beneath the repeating wave, we measure the size of each season's effect with a **seasonal index**.

What a seasonal index measures.

A **seasonal index** is the ratio of a season's typical value to the average value across all the seasons of one full cycle:

$$\text{seasonal index} = \frac{\text{average value for that season}}{\text{average value over all seasons}}.$$

An index of exactly 1 means the season sits *at* the average. An index **above 1** means the season typically runs **above** average; an index **below 1** means it runs **below** average. How far the index is from 1, written as a percentage, is how far above or below average that season runs:

- an index of **1.20** means the season is **20%** above the average season;
- an index of **0.85** means the season is **15%** below the average season, since $1 - 0.85 = 0.15$.

Because every index compares a season with the all-season average, the indices for one full cycle **add to the number of seasons** — 4 for quarterly data, 12 for monthly data, 7 for daily data — and so they **average 1**. This total is a built-in check: a set of seasonal indices that does not add to the number of seasons has not been calculated (or adjusted) correctly.

Calculating seasonal indices from data.

When the data cover several years, we compute the indices using **seasonal and yearly means**, dividing by each year's own mean so that year-to-year growth does not distort the seasonal picture:

1. For each **year** (one full cycle), find the **yearly mean** — the average of that year's seasonal figures.
2. Divide each seasonal figure by **its own year's mean**. This gives a seasonal index for every season of every year.
3. For each season, **average** the indices found for it across the years. These averages are the seasonal indices.
4. **Check and adjust.** The indices should add to the number of seasons. If a set does not (whether from rounding or from a preliminary calculation), multiply every index by the **correction factor**

$$\text{correction factor} = \frac{\text{number of seasons}}{\text{sum of the calculated indices}}.$$

When only a single year of data is available, steps 1–3 reduce to dividing each seasonal figure by that one year's mean.

Deseasonalising.

Once the seasonal indices are known, we can strip the seasonal effect out of an actual figure to reveal the underlying level. The **deseasonalised** (seasonally adjusted) value is

$$\text{deseasonalised value} = \frac{\text{actual value}}{\text{seasonal index}}.$$

Dividing by an index above 1 scales a busy season *down*; dividing by an index below 1 scales a quiet season *up*. Deseasonalised figures can be compared directly from one season to the next, because the seasonal swing has been removed — this is exactly what is meant by a "seasonally adjusted" statistic in a news report.

Reseasonalising.

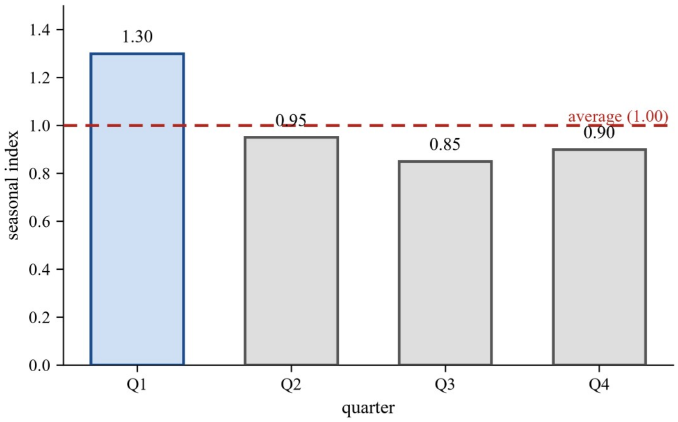
To put the seasonal effect back — for instance, to turn a deseasonalised prediction into an actual forecast for a particular season — multiply by that season's index:

$$\text{actual value} = \text{deseasonalised value} \times \text{seasonal index}.$$

The two operations are exact inverses: deseasonalising and then reseasonalising the same value returns the original figure. (Fitting a trend line to the deseasonalised series to *make* such forecasts is developed later in this chapter; here the deseasonalised value is given.)

Using CAS. General Mathematics is technology-active. In practice you enter the series into a spreadsheet or CAS list and compute the yearly means, the indices and the deseasonalised column at once. Learn the by-hand steps below so that you can interpret and check that output. Throughout this section seasonal indices are quoted correct to two decimal places.

The bar chart below shows a set of quarterly seasonal indices. The dashed line at 1.00 marks the average season; bars reaching above it are peak seasons and bars below it are quiet seasons.



Worked examples

Example 1 — scaffolded

A surf shop records its swimwear sales, in thousands of dollars, over the four quarters of one year: Q1 (summer) 130, Q2 95, Q3 (winter) 85, Q4 90. Calculate the seasonal index for each quarter, and interpret the quarter 1 index.

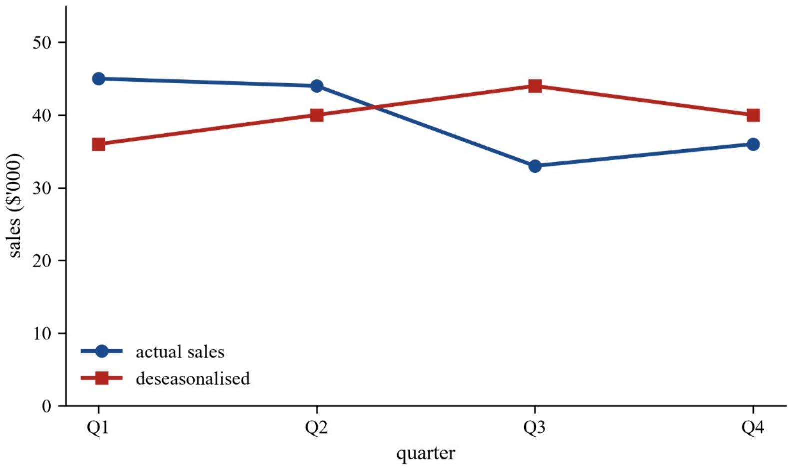
Working	Comment
Yearly mean $= \dfrac{130+95+85+90}{4} = \dfrac{400}{4} = 100.$	Average of the four quarters.
Q1 index $= \dfrac{130}{100} = 1.30.$	Seasonal index = season value ÷ yearly mean.
Q2 index $= \dfrac{95}{100} = 0.95.$	Below 1, so a below-average quarter.
Q3 index $= \dfrac{85}{100} = 0.85.$	The quietest quarter.
Q4 index $= \dfrac{90}{100} = 0.90.$	
Check: $1.30+0.95+0.85+0.90 = 4.00.$	The four indices add to the number of seasons (4).
Interpret Q1: $1.30-1 = 0.30.$	Summer sales run **30%** above the average quarter.

Example 2 — scaffolded

A bookshop's quarterly sales, in thousands of dollars, for one year were Q1 45, Q2 44, Q3 33, Q4 36. The seasonal indices are Q1 1.25, Q2 1.10, Q3 0.75, Q4 0.90. Deseasonalise each quarter, confirm that reseasonalising quarter 3 recovers its original figure, and state which quarter was strongest once the seasonal effect is removed.

Working	Comment
Q1: $\dfrac{45}{1.25}=36.$	Deseasonalised = actual ÷ seasonal index.
Q2: $\dfrac{44}{1.10}=40.$	The busy-season figure is scaled down.
Q3: $\dfrac{33}{0.75}=44.$	The quiet-season figure is scaled up.
Q4: $\dfrac{36}{0.90}=40.$	
Reseasonalise Q3: $44\times 0.75=33.$	Actual = deseasonalised × index — the original returns.
Strongest quarter: Q3, deseasonalised 44.	Q3 had the **lowest** raw sales (33) yet the **highest** underlying sales.

Once deseasonalised, the four quarters (36, 40, 44, 40) no longer swing with the seasons, so they can be compared directly. The line graph shows the raw sales zig-zagging steeply, while the deseasonalised series swings far less and rises to a clear peak in quarter 3 — the underlying strength that the raw figures hide.



Example 3 — unscaffolded

The quarterly revenue, in thousands of dollars, of an air-conditioning installer over three years is shown below (quarter 1 is summer). Calculate the four seasonal indices.

Year	Q1	Q2	Q3	Q4
1	52	38	30	44
2	58	42	34	48
3	64	46	38	54

The yearly means are

$$\bar{y}_1=\frac{52+38+30+44}{4}=41,\ \bar{y}_2=\frac{58+42+34+48}{4}=45.5,\ \bar{y}_3=\frac{64+46+38+54}{4}=50.5.$$

Dividing each figure by its own year's mean gives a seasonal index for every quarter of every year (to three decimal places):

Year	Q1	Q2	Q3	Q4
1	1.268	0.927	0.732	1.073
2	1.275	0.923	0.747	1.055
3	1.267	0.911	0.752	1.069

Averaging each column over the three years,

$$Q1=\frac{1.268+1.275+1.267}{3}\approx 1.27\,,Q2=\frac{0.927+0.923+0.911}{3}\approx 0.92\,,$$

$$Q3=\frac{0.732+0.747+0.752}{3}\approx 0.74\,,Q4=\frac{1.073+1.055+1.069}{3}\approx 1.07\,.$$

The indices add to $1.27+0.92+0.74+1.07=4.00$, the number of seasons, so no adjustment is needed.

∴ the seasonal indices are Q1 1.27, Q2 0.92, Q3 0.74 and Q4 1.07 — summer runs about 27% above average and winter about 26% below.

Example 4 — unscaffolded

A department store's sales follow a strong yearly cycle. A statistical model estimates that, with the seasonal effect removed, the store's **deseasonalised** sales for next December will be $\$80\,000$. The seasonal index for December is 1.45 and for June it is 0.70. Forecast the actual December sales, interpret the two indices, and — given that last June's actual sales were $\$42\,000$ — compare June's underlying level with December's.

Reseasonalising the December estimate,

$$\text{actual December}=80\,000\times 1.45=\$116\,000\,.$$

The index 1.45 means December sales run 45% above the average month (the pre-Christmas peak), while 0.70 means June sales run 30% below the average month.

Deseasonalising last June's figure,

$$\text{deseasonalised June}=\frac{42\,000}{0.70}=\$60\,000\,.$$

∴ the December forecast is $\$116\,000$; and because December's underlying level ($\$80\,000$) still exceeds June's underlying level ($\$60\,000$) after the seasonal swing is removed, the gap between the two months is more than just a seasonal effect.

Practice questions

Scaffolded

Q1. The quarterly attendance (in hundreds) at an outdoor pool for one year was Q1 132, Q2 120, Q3 52, Q4 96. (a) Find the yearly mean. (b) Find the seasonal index for each quarter. (c) Check that the four indices add to 4. (d) Interpret the quarter 3 index.

Q2. A shop's quarterly sales ($'000) for one year were Q1 72, Q2 69, Q3 30, Q4 45, and the seasonal indices are Q1 1.20, Q2 1.15, Q3 0.75, Q4 0.90. Deseasonalise each quarter using $\text{deseasonalised}=\frac{\text{actual}}{\text{seasonal index}}$.

Q3. A business expects its deseasonalised sales to be $\$40\,000$ in every quarter of next year. The seasonal indices are Q1 1.25, Q2 1.10, Q3 0.80, Q4 0.85. Reseasonalise to forecast the **actual** sales in each quarter, using $\text{actual}=\text{deseasonalised}\times\text{seasonal index}$.

Q4. The twelve monthly seasonal indices for an ice-cream van are Jan 1.45, Feb 1.35, Mar 1.10, Apr 0.95, May 0.80, Jun 0.65, Jul 0.60, Aug 0.70, Sep 0.85, Oct 1.00, Nov 1.20, Dec 1.35. (a) Which month is busiest, and which is quietest? (b) By what percentage are July's sales below the average month? (c) What does October's index of 1.00 tell you? (d) Confirm that the twelve indices add to 12.

Q5. A preliminary calculation produced the quarterly seasonal indices 1.24, 1.06, 0.78, 0.88. (a) Show that these do not add to 4. (b) Find the correction factor $\frac{4}{\text{their sum}}$. (c) Multiply each index by the correction factor to obtain the adjusted indices (to two decimal places), and check that they now add to 4.

Q6. The quarterly sales ($'000) of a nursery over two years were Year 1: 80, 110, 130, 80 and Year 2: 96, 120, 168, 96. (a) Find the mean of each year. (b) Divide each figure by its own year's mean to find that year's seasonal indices. (c) Average the two years to find the seasonal index for each quarter. (d) Check that the four indices add to 4.

Unscaffolded

Q7. The quarterly number of covers (in hundreds) served by a restaurant in one year was Q1 110, Q2 145, Q3 80, Q4 65. Find the four seasonal indices and state which quarter is the busiest.

Q8. The seasonal indices for a firm's four quarters are 1.30, 1.05, 0.75 and 0.90. In one year the actual sales ($'000) were Q1 65, Q2 63, Q3 30, Q4 54. Deseasonalise each quarter.

Q9. A model estimates the deseasonalised summer sales of a swimwear label to be 120 thousand units. The summer seasonal index is 1.35. Forecast the actual summer sales.

Q10. A toy shop's seasonal index for December is 1.60, and its actual December sales were $\$96\,000$. Find the deseasonalised December sales and interpret the index.

Q11. The quarterly ticket sales (in thousands) at a tourist railway over three years were Year 1: 48, 32, 28, 44; Year 2: 54, 38, 30, 50; Year 3: 60, 40, 36, 52. Calculate the four seasonal indices.

Q12. A preliminary calculation gave the quarterly seasonal indices 1.18, 1.12, 0.86 and 0.80. Adjust them with a correction factor so that they add to 4.

Q13. A ski resort's winter seasonal index is 1.25. (a) Last winter it had $75\,000$ visitors. Find the deseasonalised figure. (b) The deseasonalised forecast for next winter is $68\,000$ visitors. Forecast the actual number of winter visitors.

Q14. A market stall's seasonal index for Tuesday is 0.60. (a) By what percentage are Tuesday's takings below the daily average? (b) Last Tuesday the stall took $84. Find the deseasonalised takings.

Q15. Three of the four quarterly seasonal indices for a company are 1.30, 0.95 and 0.85. Find the fourth index, and state what it tells you about that quarter.

Q16. Eleven of the twelve monthly seasonal indices for a business add to 11.15. Find the twelfth index.

Q17. A café forecasts its deseasonalised quarterly sales ($'000) for next year to be 200, 210, 220 and 230. The seasonal indices are Q1 1.20, Q2 0.90, Q3 0.80, Q4 1.10. Forecast the actual sales in each quarter.

Q18. In one year a store's quarter 1 sales were $\$39\,000$ (seasonal index 1.30) and its quarter 3 sales were $\$30\,000$ (seasonal index 0.75). (a) Deseasonalise each quarter. (b) Which quarter had the stronger underlying sales? Explain how the raw figures were misleading.

Q19. A firm's actual quarterly sales ($'000) for one year were 80, 77, 54 and 72, and the seasonal indices are Q1 1.25, Q2 1.10, Q3 0.75, Q4 0.90. Deseasonalise the four quarters and describe what the deseasonalised series shows about the underlying trend.

Q20. (a) What does a seasonal index of exactly 1.00 tell you about that season? (b) Why are deseasonalised figures more useful than raw figures for judging whether sales are genuinely growing? (c) A student's four quarterly seasonal indices add to 4.35. Explain what must have gone wrong.

Answers

Q1. (a) 100; (b) Q1 1.32, Q2 1.20, Q3 0.52, Q4 0.96; (c) $1.32+1.20+0.52+0.96=4.00$; (d) quarter 3 attendance is 48% below the average quarter. **Q2.** Deseasonalised ($'000): Q1 60, Q2 60, Q3 40, Q4 50. **Q3.** Actual ($): Q1 50 000, Q2 44 000, Q3 32 000, Q4 34 000. **Q4.** (a) busiest January (1.45), quietest July (0.60); (b) 40% below; (c) October is an exactly average month; (d) the twelve indices add to 12.00. **Q5.** (a) $1.24+1.06+0.78+0.88=3.96\neq 4$; (b) correction factor $=\dfrac{4}{3.96}\approx 1.0101$; (c) adjusted 1.25, 1.07, 0.79, 0.89 (sum 4.00). **Q6.** (a) 100 and 120; (b) Year 1 0.80, 1.10, 1.30, 0.80 and Year 2 0.80, 1.00, 1.40, 0.80; (c) Q1 0.80, Q2 1.05, Q3 1.35, Q4 0.80; (d) sum $=4.00$. **Q7.** Q1 1.10, Q2 1.45, Q3 0.80, Q4 0.65; quarter 2 is the busiest. **Q8.** Deseasonalised ($'000): Q1 50, Q2 60, Q3 40, Q4 60. **Q9.** $120\times 1.35=162$ thousand units. **Q10.** Deseasonalised $=\dfrac{96\,000}{1.60}=\$60\,000$; December sales run 60% above the average month. **Q11.** Q1 1.27, Q2 0.86, Q3 0.73, Q4 1.14 (sum 4.00). **Q12.** Correction factor $\dfrac{4}{3.96}\approx 1.0101$; adjusted 1.19, 1.13, 0.87, 0.81 (sum 4.00). **Q13.** (a) $\dfrac{75\,000}{1.25}=60\,000$; (b) $68\,000\times 1.25=85\,000$ visitors. **Q14.** (a) 40% below; (b) $\dfrac{84}{0.60}=\$140$. **Q15.** $4-(1.30+0.95+0.85)=0.90$; that quarter runs 10% below the average quarter. **Q16.** $12-11.15=0.85$. **Q17.** Actual ($'000): Q1 240, Q2 189, Q3 176, Q4 253. **Q18.** (a) Q1 $\dfrac{39\,000}{1.30}=30\,000$, Q3 $\dfrac{30\,000}{0.75}=40\,000$; (b) quarter 3 was stronger underlying — its low raw figure is only because Q3 is a quiet season. **Q19.** Deseasonalised ($'000): 64, 70, 72, 80 — a steadily rising series, so underlying sales are trending upward even though the raw sales zig-zag. **Q20.** (a) that season is exactly average; (b) deseasonalising removes the seasonal swing, so a rise in the deseasonalised figures reflects genuine growth, not the season; (c) the indices must add to 4 — a total of 4.35 means an arithmetic error, so the indices need re-checking or adjusting.

Fully-worked solutions

Q1. The yearly mean is $\dfrac{132+120+52+96}{4}=\dfrac{400}{4}=100$. Dividing each quarter by 100 gives the indices $\dfrac{132}{100}=1.32$, $\dfrac{120}{100}=1.20$, $\dfrac{52}{100}=0.52$ and $\dfrac{96}{100}=0.96$. Their sum is $1.32+1.20+0.52+0.96=4.00$, as required. The quarter 3 index 0.52 is $1-0.52=0.48$ below 1, so winter attendance is 48% below the average quarter.

Q2. Deseasonalise each quarter with actual ÷ index: $\dfrac{72}{1.20}=60$, $\dfrac{69}{1.15}=60$, $\dfrac{30}{0.75}=40$ and $\dfrac{45}{0.90}=50$. The deseasonalised sales are $\$60\,000$, $\$60\,000$, $\$40\,000$ and $\$50\,000$.

Q3. Reseasonalise each quarter with deseasonalised × index: $40\,000\times 1.25=50\,000$, $40\,000\times 1.10=44\,000$, $40\,000\times 0.80=32\,000$ and $40\,000\times 0.85=34\,000$. The forecast actual sales are $\$50\,000$, $\$44\,000$, $\$32\,000$ and $\$34\,000$. Even though the underlying level is the same each quarter, the actual figures differ because of the seasonal effect.

Q4. (a) The largest index is January's 1.45, so January is busiest; the smallest is July's 0.60, so July is quietest. (b) July's index is $1-0.60=0.40$ below 1, so its sales are 40% below the average month. (c) An index of 1.00 means October sits exactly at the average month — no seasonal effect. (d) Adding all twelve, $1.45+1.35+1.10+0.95+0.80+0.65+0.60+0.70+0.85+1.00+1.20+1.35=12.00$, the number of seasons.

Q5. (a) $1.24+1.06+0.78+0.88=3.96$, which is not 4, so the indices must be adjusted. (b) The correction factor is $\dfrac{4}{3.96}\approx 1.0101$. (c) Multiplying each index by 1.0101: $1.24\times 1.0101\approx 1.25$, $1.06\times 1.0101\approx 1.07$, $0.78\times 1.0101\approx 0.79$ and $0.88\times 1.0101\approx 0.89$. The adjusted indices $1.25+1.07+0.79+0.89=4.00$ now add to 4.

Q6. (a) Year 1 mean $=\frac{80+110+130+80}{4}=\frac{400}{4}=100$; Year 2 mean $=\frac{96+120+168+96}{4}=\frac{480}{4}=120$. (b) Year 1 indices: $\frac{80}{100},\frac{110}{100},\frac{130}{100},\frac{80}{100}=0.80,1.10,1.30,0.80$; Year 2 indices: $\frac{96}{120},\frac{120}{120},\frac{168}{120},\frac{96}{120}=0.80,1.00,1.40,0.80$. (c) Averaging the two years quarter by quarter: $\frac{0.80+0.80}{2}=0.80$, $\frac{1.10+1.00}{2}=1.05$, $\frac{1.30+1.40}{2}=1.35$ and $\frac{0.80+0.80}{2}=0.80$. (d) The sum is $0.80+1.05+1.35+0.80=4.00$.

Q7. The yearly mean is $\frac{110+145+80+65}{4}=\frac{400}{4}=100$. The indices are $\frac{110}{100}=1.10$, $\frac{145}{100}=1.45$, $\frac{80}{100}=0.80$ and $\frac{65}{100}=0.65$. Quarter 2, with the largest index (1.45), is the busiest — about 45% above the average quarter.

Q8. Deseasonalise with actual ÷ index: $\frac{65}{1.30}=50$, $\frac{63}{1.05}=60$, $\frac{30}{0.75}=40$ and $\frac{54}{0.90}=60$. The deseasonalised sales are $\$50\,000$, $\$60\,000$, $\$40\,000$ and $\$60\,000$.

Q9. Reseasonalise the deseasonalised estimate: $120\times1.35=162$. The forecast actual summer sales are 162 thousand units.

Q10. The deseasonalised December figure is $\frac{96\,000}{1.60}=\$60\,000$. An index of 1.60 is $1.60-1=0.60$ above 1, so December sales run 60% above the average month.

Q11. The yearly means are $\frac{48+32+28+44}{4}=38$, $\frac{54+38+30+50}{4}=43$ and $\frac{60+40+36+52}{4}=47$. Dividing each figure by its year's mean and then averaging each quarter over the three years gives

$$Q1=\frac{1.263+1.256+1.277}{3}\approx1.27,\ Q2=\frac{0.842+0.884+0.851}{3}\approx0.86,$$

$$Q3=\frac{0.737+0.698+0.766}{3}\approx0.73,\ Q4=\frac{1.158+1.163+1.106}{3}\approx1.14.$$

The indices are Q1 1.27, Q2 0.86, Q3 0.73, Q4 1.14, which add to 4.00.

Q12. The preliminary indices add to $1.18+1.12+0.86+0.80=3.96$, so the correction factor is $\frac{4}{3.96}\approx1.0101$. Multiplying: $1.18\times1.0101\approx1.19$, $1.12\times1.0101\approx1.13$, $0.86\times1.0101\approx0.87$ and $0.80\times1.0101\approx0.81$. The adjusted indices $1.19+1.13+0.87+0.81=4.00$ add to 4.

Q13. (a) Deseasonalised $=\frac{75\,000}{1.25}=60\,000$ visitors. (b) Reseasonalise the forecast: $68\,000\times1.25=85\,000$ visitors.

Q14. (a) The index 0.60 is $1-0.60=0.40$ below 1, so Tuesday's takings are 40% below the daily average. (b) Deseasonalised $=\frac{84}{0.60}=\$140$.

Q15. The four indices must add to 4, so the fourth is $4-(1.30+0.95+0.85)=4-3.10=0.90$. Since 0.90 is 0.10 below 1, that quarter runs 10% below the average quarter.

Q16. The twelve monthly indices must add to 12, so the missing index is $12-11.15=0.85$.

Q17. Reseasonalise each quarter with deseasonalised × index: $200\times1.20=240$, $210\times0.90=189$, $220\times0.80=176$ and $230\times1.10=253$. The forecast actual sales are $\$240\,000$, $\$189\,000$, $\$176\,000$ and $\$253\,000$.

Q18. (a) Deseasonalising: quarter 1 $\frac{39\,000}{1.30}=30\,000$ and quarter 3 $\frac{30\,000}{0.75}=40\,000$. (b) Once the seasonal effect is removed, quarter 3 ($\$40\,000$) is the stronger of the two, even though its raw sales ($\$30\,000$) were lower than quarter 1's ($\$39\,000$). The raw figures were misleading because quarter 1 is a busy season (index 1.30 inflates it) while quarter 3 is a quiet season (index 0.75 deflates it).

Q19. Deseasonalise with actual ÷ index: $\frac{80}{1.25}=64$, $\frac{77}{1.10}=70$, $\frac{54}{0.75}=72$ and $\frac{72}{0.90}=80$. The deseasonalised sales ($'000) are 64, 70, 72, 80. Although the raw sales fall then rise, the deseasonalised series increases steadily from 64 to 80, so once the seasonal swing is removed the underlying sales are trending upward.

Q20. (a) A seasonal index of exactly 1.00 means that season equals the all-season average — it has no seasonal effect. (b) Raw figures rise and fall with the season, so a high figure may just be a busy season rather than real growth; deseasonalising divides out the seasonal index, leaving figures that can be compared fairly, so a genuine upward movement in the deseasonalised series signals real growth. (c) Four quarterly seasonal indices must add to 4; a total of 4.35 shows an arithmetic slip in the calculation, and the indices would need re-checking (or, if only rounding is at fault, rescaling by the correction factor $\frac{4}{4.35}$).

Chapter 5 Investigating and modelling time series — 5E Fitting a trend line and forecasting

Theory

A time series is a set of measurements of one variable recorded at successive, equally spaced points in time — each year, each quarter, each month, and so on. When such a series shows a clear long-term **trend** (a steady rise or fall), we can model that trend with a straight line and then use the line to **forecast** values at future times. This section shows how to fit that line, how to read it, and — just as importantly — how far it can be trusted.

Time as the explanatory variable.

To fit a least-squares line we need a numerical explanatory variable. For a time series the natural explanatory variable is **time itself**: the response variable depends on *when* the measurement was taken. Because the observations are equally spaced, we simply **number the time periods** $1, 2, 3, \ldots$ in order. If a series records yearly figures for 2019, 2020, 2021, …, we label them $t=1, 2, 3, \ldots$; if it records quarterly figures, the first quarter is $t=1$, the second $t=2$, and so on. The value of t carries all the timing information the line needs.

Fitting the trend line.

With the periods numbered, (t, y) becomes an ordinary bivariate data set and we fit the **least-squares line** in the usual way. Writing the line as

$$y = a + bt,$$

the slope and intercept are

$$b = r\frac{s_y}{s_t}, a = \bar{y} - b\bar{t},$$

where r is the correlation coefficient between t and y, and s_t and s_y are the standard deviations of the period numbers and of the values — the same least-squares formulas used for any bivariate data set, with time as the explanatory variable. In practice you enter the period numbers and the values into your CAS, run the **least-squares regression**, and read off a and b — exactly as for any other regression. Because it models the underlying direction of the series, the fitted line $y = a + bt$ is called the **trend line**.

- The **slope** b is the average change in the response variable **per time period**. A positive b signals an upward trend, a negative b a downward trend.
- The **intercept** a is the value the model gives at $t=0$ — the period *just before* the first recorded one.

Quote the coefficients to a sensible accuracy (this book rounds them to two decimal places), but keep the full-precision values stored in your calculator to use in any forecast.

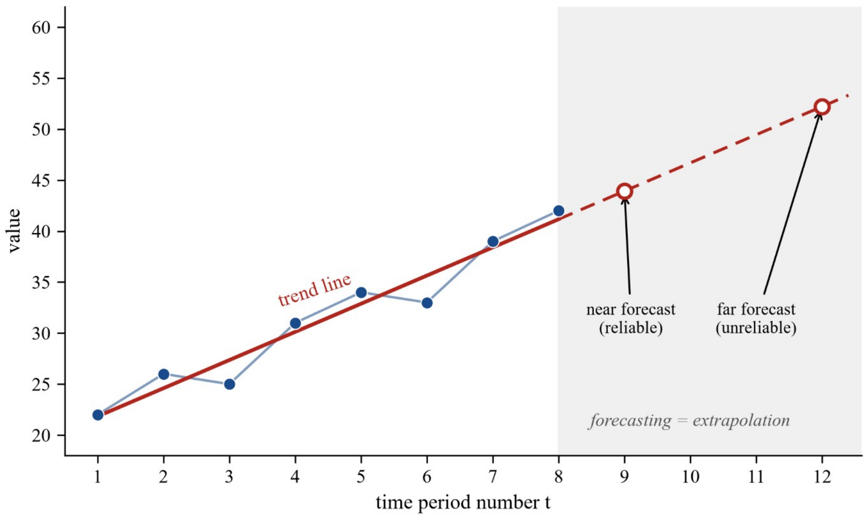
Forecasting.

To **forecast** a future value, work out the period number t of the future time and substitute it into the trend line. For yearly data numbered $t=1, 2, \ldots, 6$, the next year is $t=7$, the year after that $t=8$, and so on. Every forecast is an **extrapolation** — a prediction *beyond* the range of the data — so the reliability of a forecast falls away the further ahead we go:

- a forecast **one or two periods** past the data is usually reasonable, provided the trend continues;
- a forecast **far** into the future is unreliable, because the trend may flatten, reverse, or be disrupted by a **structural change** (a new competitor, a policy change, a change in technology), and a straight line eventually predicts values that are impossible in context (such as a negative population or an ever-rising temperature).

If a question asks *when* the trend reaches a given value, substitute that value for y and solve for t. The answer is rarely a whole number, and forecasts are only made at whole periods, so **round up**: the first period at which the trend has reached the value is the next whole period after the solution.

Always state a forecast together with a comment on how far ahead it reaches.



Forecasting from seasonal data.

Many real series are **seasonal**: they carry a repeating within-year pattern on top of the trend, so a line fitted to the raw figures is pulled up and down by the seasons and models the trend poorly. The remedy is to remove the seasonal pattern first, fit the trend to the smoothed series, and then put the seasonal pattern back into the forecast. This gives a four-step pipeline:

1. **Deseasonalise** the data. In an earlier section we found the **seasonal indices**. Recall that a seasonal index measures how one season compares with the average season: an index of 1.20 means that season runs $20\,\%$ **above** the average (deseasonalised) level, and 0.85 means $15\,\%$ **below** it. For a cycle of L seasons the indices add to L, so they average 1. To remove the seasonal effect, divide each value by its season's index:
$$\text{deseasonalised value} = \frac{\text{actual value}}{\text{seasonal index}}.$$
2. **Fit the trend line** to the *deseasonalised* values against the period number t, using least squares as above.
3. **Forecast** the deseasonalised value at the required future period by substituting its t into the trend line.
4. **Reseasonalise** the forecast — put the season back — by multiplying by the seasonal index of that season:
$$\text{forecast (actual) value} = \text{deseasonalised forecast} \times \text{seasonal index}.$$

Deseasonalising divides out the season; reseasonalising multiplies it back in. The trend line lives entirely in the deseasonalised world; the seasonal index converts a trend forecast into a forecast of the actual figure you would expect to see.

Using CAS. General Mathematics is a technology-active course. In practice you let the CAS do the regression: enter the period numbers t as the explanatory list and the values (the deseasonalised values, for seasonal data) as the response list, and read off a and b. The by-hand structure below shows what the calculator is doing so that you can set out the deseasonalise–fit–forecast–reseasonalise steps correctly and check that its answer is reasonable.

Worked examples

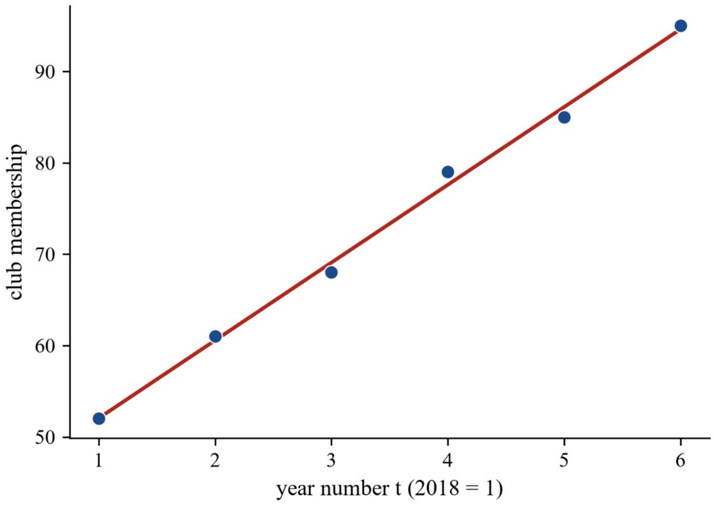
Example 1 — scaffolded

A coding club recorded its membership at the start of each year from 2018 to 2023: 52, 61, 68, 79, 85, 95. Fit a least-squares trend line, interpret its slope, and forecast the membership at the start of 2024.

Working	Comment
Number the years $t = 1, 2, 3, 4, 5, 6$ for 2018–2023.	Time is the explanatory variable; equally spaced periods are numbered in order.
Least-squares regression of y on t gives $a = 43.53$,	Enter t and y into the CAS and read off the intercept

Working	Comment
$b = 8.51$.	and slope (2 d.p.).
Trend line: $y = 43.53 + 8.51t$.	y is the membership, t the year number.
Slope $b = 8.51$: membership grows by about **8.5 members per year** on average.	The slope is the average change per time period.
2024 is $t = 7$. Using the stored (unrounded) coefficients, $y = 43.5333\ldots + 8.5142\ldots \times 7 = 103.13\ldots$, so $y \approx 103$.	Substitute the future period number into the *stored* line, not the rounded one; round to a whole number of members.
$t = 7$ is one year past the data, so the forecast is reasonable if the trend holds.	Modest extrapolation — treat a forecast for, say, 2035 with far more caution.

∴ the trend line is $y = 43.53 + 8.51t$ and the membership is forecast to be about 103 at the start of 2024.



Example 2 — scaffolded

A company that supplies home heating recorded its quarterly sales (in thousands of dollars) over two years:

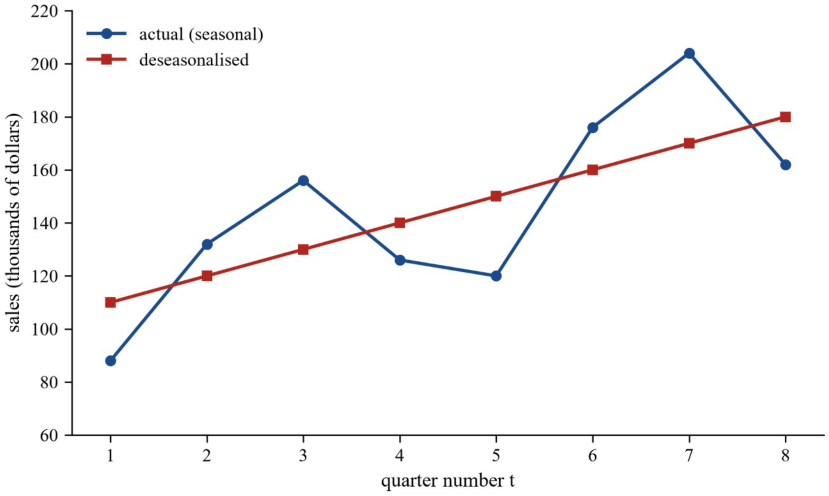
Quarter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Sales	88	132	156	126	120	176	204	162

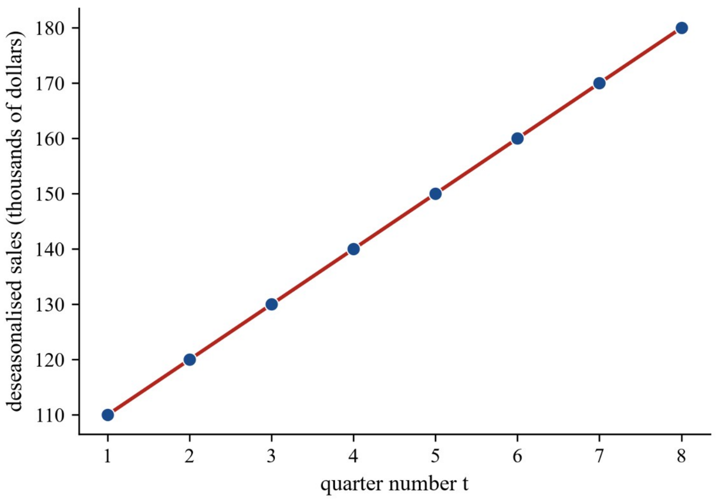
The seasonal indices found earlier are Q1 $= 0.8$, Q2 $= 1.1$, Q3 $= 1.2$, Q4 $= 0.9$. Deseasonalise the data, fit a trend line, and forecast the sales in each quarter of the third year.

Working	Comment
Number the quarters $t = 1, 2, \ldots, 8$.	Eight equally spaced periods.
Deseasonalise: divide each figure by its quarter's index, e.g. $88 \div 0.8 = 110$, $132 \div 1.1 = 120$, $156 \div 1.2 = 130$, $126 \div 0.9 = 140$.	$\text{deseasonalised} = \text{actual} \div \text{index}$; removes the seasonal swing.
Deseasonalised series: 110, 120, 130, 140, 150, 160, 170, 180.	A smooth, steadily rising series — the trend laid bare.
Least-squares regression on the deseasonalised values gives $a = 100$, $b = 10$.	Fit t against the deseasonalised figures, not the raw ones.
Trend line: $y = 100 + 10t$.	Deseasonalised sales rise by about \$10 thousand per quarter.
Year 3 is $t = 9, 10, 11, 12$. Trend forecasts: $100 + 10 \times 9 = 190$; then 200, 210, 220.	Substitute each future period number.
Reseasonalise (multiply by that quarter's index): Q1	$\text{forecast} = \text{deseasonalised forecast} \times \text{index}$; puts the

Working	Comment
$190 \times 0.8 = 152$; Q2 $200 \times 1.1 = 220$; Q3 $210 \times 1.2 = 252$; Q4 $220 \times 0.9 = 198$.	season back.

∴ the third-year sales are forecast to be about \$152, \$220, \$252 and \$198 thousand for Q1–Q4 respectively.





Example 3 — unscaffolded

A café's quarterly customer numbers were deseasonalised and the least-squares trend line was found to be $y = 200 + 15t$, where t is the quarter number ($t = 1$ is the first quarter of 2021, and the data run to $t = 8$, the last quarter of 2022). The seasonal indices are Q1 $= 1.3$, Q2 $= 1.1$, Q3 $= 0.7$, Q4 $= 0.9$. Forecast the number of customers in the first quarter of 2024, and comment on the reliability of the forecast.

The first quarter of 2024 is five quarters after the last data point (end of 2022), so its period number is $t = 13$. The trend line gives the deseasonalised forecast

$$y = 200 + 15 \times 13 = 395.$$

The first quarter has seasonal index 1.3, so the reseasonalised forecast is

$$395 \times 1.3 = 513.5 \approx 514 \text{ customers}.$$

Because $t = 13$ lies well beyond the data (which stop at $t = 8$), this is a substantial extrapolation: the linear trend may not continue that far, and the estimate should be treated with caution.

∴ the café is forecast to have about 514 customers in the first quarter of 2024, but the forecast is unreliable because it extrapolates five quarters past the data.

Example 4 — unscaffolded

The number of subscribers (in thousands) to a print magazine from 2020 to 2024 was 88, 84, 79, 73, 66. Fit a trend line, forecast the number of subscribers in 2025, and determine the year in which the trend predicts the number will fall to 50 thousand. Comment on the model's long-term behaviour.

Numbering the years $t=1,2,3,4,5$, least-squares regression gives the trend line

$$y=94.5-5.5t.$$

The slope -5.5 means the magazine is losing about 5.5 thousand subscribers per year on average. For 2025, $t=6$, so

$$y=94.5-5.5\times 6=61.5 \text{ thousand subscribers.}$$

To find when the trend reaches 50 thousand, solve $50=94.5-5.5t$:

$$5.5t=44.5, t=\frac{44.5}{5.5}\approx 8.1,$$

so the line passes 50 between $t=8$ and $t=9$. Forecasts are made at whole periods, so take the first whole year at or past the crossing: $t=9$, the year **2028**. (At $t=8$, the year 2027, the model still predicts $94.5-5.5\times 8=50.5$ thousand, which is above 50.) Extending the line further, it crosses $y=0$ at $t=94.5\div 5.5\approx 17.2$, so from $t=18$ (the year 2037) the model predicts a negative number of subscribers — an impossibility. This shows why a linear trend cannot be trusted for forecasts far into the future.

∴ the trend line is $y=94.5-5.5t$; the 2025 forecast is about 61.5 thousand subscribers, the number is predicted to fall to 50 thousand in 2028, and the model breaks down (predicting impossible negative values) if extended much beyond the data.

Practice questions

Scaffolded

Q1. The weekly attendance at a new fitness class over its first five weeks was 12, 15, 18, 21, 24. (a) Number the weeks $t=1,2,3,4,5$. (b) A CAS gives the least-squares line as $y=a+bt$. State a and b. (c) Write down the trend line. (d) Interpret the slope in context. (e) Forecast the attendance in week 8.

Q2. A shop's quarterly sales ($000) for one year were Q1 96, Q2 63, Q3 81, Q4 120. The seasonal indices are Q1 $=1.2$, Q2 $=0.9$, Q3 $=0.9$, Q4 $=1.0$. (a) Explain what the Q1 index of 1.2 tells you about first-quarter sales. (b) Deseasonalise each quarter's sales. (c) State which quarter had the strongest sales once the season is removed.

Q3. For a quarterly series the deseasonalised trend line is $y=50+4t$, where t is the quarter number. The seasonal indices are Q1 $=0.8$, Q2 $=1.1$, Q3 $=1.3$, Q4 $=0.8$. (a) Find the deseasonalised forecast for $t=9$. (b) The period $t=9$ is a first quarter. Reseasonalise the forecast. (c) Repeat parts (a)–(b) for $t=10$ (a second quarter).

Q4. The monthly number of items sold by an online store over six months was 28, 30, 37, 45, 49, 57. (a) Numbering the months $t=1,\ldots,6$, a CAS gives the trend line $y=20+6t$. Use it to forecast sales in month 8. (b) Forecast sales in month 15. (c) Which of the two forecasts is more reliable, and why?

Q5. A trend line applied to deseasonalised quarterly data gives the following forecasts for next year's four quarters: 160, 168, 176, 184. The seasonal indices are Q1 $=0.8$, Q2 $=1.1$, Q3 $=1.2$, Q4 $=0.9$. Reseasonalise each forecast to predict the actual sales in each quarter.

Q6. A ski resort's half-yearly visitor numbers (thousands) over three years were 60, 44, 72, 52, 84, 60, where the two seasons each year have indices H1 $=1.2$ and H2 $=0.8$. (a) Deseasonalise the six figures. (b) A CAS gives the trend line $y=45+5t$ for the deseasonalised data. State the slope and interpret it. (c) Forecast the visitor numbers in the first half (H1) of year 4. (d) Forecast the visitor numbers in the second half (H2) of year 4.

Unscaffolded

Q7. A town's recycling tonnage over five years, numbered $t=1,\ldots,5$, was 216, 228, 247, 258, 276. The least-squares trend line is $y=200+15t$. (a) Interpret the slope and the intercept in context. (b) Forecast the tonnage in year 7.

Q8. A magazine's subscriber numbers (thousands) follow the trend line $y=94.5-5.5t$, where t numbers the years and the data run to $t=5$. Determine the year in which the trend predicts subscribers will fall to 40 thousand, and explain why this prediction should be treated with caution.

Q9. (a) Explain why, when fitting a trend line to a time series, time is used as the explanatory variable and the periods are numbered $1,2,3,\ldots$. (b) Explain why a forecast made 15 years beyond the data is far less trustworthy than one made a single year ahead.

Q10. The quarterly sales of a clothing store over two years were 92, 76, 64, 88, 115, 95, 80, 110. The seasonal indices are Q1 $=1.15$, Q2 $=0.95$, Q3 $=0.80$, Q4 $=1.10$. (a) Deseasonalise the eight figures. (b) Describe what the deseasonalised series reveals about the underlying trend.

Q11. A juice bar recorded quarterly revenue ($000) over two years: 108, 70, 99, 144, 156, 98, 135, 192, with seasonal indices Q1 $=1.2$, Q2 $=0.7$, Q3 $=0.9$, Q4 $=1.2$. (a) Deseasonalise the data. (b) Fit the trend line to the deseasonalised values (a CAS gives $y=80+10t$). (c) Forecast the revenue in each of the four quarters of the third year.

Q12. The mean monthly maximum temperature (°C) at a site over six months was 12, 15, 17, 21, 23, 26. (a) Numbering the months $t=1,\ldots,6$, the trend line is $y=9.2+2.8t$. Forecast the temperature in month 7. (b) Forecast the temperature in month 12. (c) Explain why the month-12 forecast is not sensible.

Q13. A start-up's quarterly revenue ($000) over five quarters was 120, 138, 161, 179, 202. The trend line is $y=98.5+20.5t$. (a) Forecast the revenue when $t=8$. (b) Find the first quarter in which the trend predicts revenue will exceed $300 thousand.

Q14. For a quarterly series the deseasonalised trend line is $y=300+12t$, and the seasonal indices are Q1 $=1.3$, Q2 $=1.1$, Q3 $=0.7$, Q4 $=0.9$. Forecast the actual values for the four quarters of the year numbered $t=9,10,11,12$.

Q15. An ice-cream vendor's deseasonalised quarterly sales trend line has slope 4.2 ($000 per quarter). (a) Interpret this slope. (b) The vendor uses the line to forecast sales five years ahead. Give one reason the forecast may turn out to be badly wrong.

Q16. A department store's December sales have a seasonal index of 1.45. (a) Interpret this index. (b) December sales one year were $87 000. Find the deseasonalised value. (c) The deseasonalised trend line forecasts $64 000 for next December. Find the reseasonalised (actual) forecast.

Q17. A gym's quarterly membership income ($000) over two years was 61.2, 98.8, 92.4, 64.4, 90, 140.4, 127.6, 86.8, with seasonal indices Q1 $=0.9$, Q2 $=1.3$, Q3 $=1.1$, Q4 $=0.7$. (a) Deseasonalise the data. (b) A CAS gives the trend line $y=60+8t$. Forecast the deseasonalised income for the next two quarters ($t=9$ and $t=10$). (c) Reseasonalise both forecasts ($t=9$ is a first quarter, $t=10$ a second quarter).

Q18. A regional newspaper's circulation (thousands) over six years was 142, 131, 123, 110, 101, 92. The least-squares trend line is $y=151.8-10.09t$. (a) Forecast the circulation in year 7. (b) The trend line reaches zero at about $t=15$. Explain why this is not a sensible prediction of the circulation in year 15.

Q19. A theme park's half-yearly attendance (thousands) over three years was 41.4, 57.2, 52.2, 70.4, 63, 83.6, with half-year indices H1 $=0.9$ and H2 $=1.1$. (a) Deseasonalise the six figures. (b) The trend line for the deseasonalised data is $y=40+6t$. Forecast the attendance in the first half (H1) of year 4. (c) Forecast the attendance in the second half (H2) of year 4.

Q20. A bakery's quarterly profit ($000) over two years was 19.2, 24.3, 33, 39.6, 28.8, 35.1, 46.2, 54, with seasonal indices Q1 $=0.8$, Q2 $=0.9$, Q3 $=1.1$, Q4 $=1.2$. (a) Deseasonalise the data and confirm that the deseasonalised series rises steadily. (b) A CAS gives the trend line $y=21+3t$. Forecast the actual profit in the first two quarters of the third year. (c) Would you trust this trend line to forecast the profit three years beyond the data? Explain.

Answers

Q1. (a) $t=1,2,3,4,5$; (b) $a=9$, $b=3$; (c) $y=9+3t$; (d) attendance rises by 3 per week on average; (e) $y=9+3\times 8=33$. **Q2.** (a) first-quarter sales run 20% above the average quarter; (b) 80, 70, 90, 120; (c) Q4 (deseasonalised 120). **Q3.** (a) $y=50+4\times 9=86$; (b) $86\times 0.8=68.8$; (c) $y=90$, reseasonalised $90\times 1.1=99$. **Q4.** (a) $y=20+6\times 8=68$; (b) $y=20+6\times 15=110$; (c) the month-8 forecast — it is far closer to the data, so extrapolation is smaller. **Q5.** $128, 184.8, 211.2, 165.6$. **Q6.** (a) 50, 55, 60, 65, 70, 75; (b) slope 5 — deseasonalised visitors rise by 5 thousand per half-year; (c) $t=7$: $80\times 1.2=96$ thousand; (d) $t=8$: $85\times 0.8=68$ thousand. **Q7.** (a) slope 15: tonnage rises by 15 per year on average; intercept 200: the model's tonnage in year 0 (the year before the data); (b) $y=200+15\times 7=305$. **Q8.** $40=94.5-5.5t \Rightarrow t\approx 9.9$, during year 10; unreliable because it extrapolates about five years beyond the data, where the trend may not hold. **Q9.** (a) the value depends on *when* it is measured, and equal spacing lets the periods be numbered $1,2,3,\ldots$ as a numerical explanatory variable; (b) forecasting far ahead is a large extrapolation — the trend may change or a structural change may occur, so error grows with distance from the data. **Q10.** (a) 80, 80, 80, 80, 100, 100, 100, 100; (b) once the season is removed the level is steady within each year and jumps from about 80 to about 100 — a clear upward trend between the two years. **Q11.** (a) 90, 100, 110, 120, 130, 140, 150, 160; (b) $y=80+10t$; (c) $t=9$–12: trend 170, 180, 190, 200; reseasonalised $204, 126, 171, 240$. **Q12.** (a) $y=9.2+2.8\times 7=28.8$ °C; (b) $y=9.2+2.8\times 12=42.8$ °C; (c) month 12 is a large extrapolation and 42.8 °C is unrealistically high — temperature does not keep rising linearly. **Q13.** (a) $y=98.5+20.5\times 8=262.5$; (b) solve $98.5+20.5t>300 \Rightarrow t>9.83$, so the 10th quarter. **Q14.** $t=9$: $408\times 1.3=530.4$; $t=10$: $420\times 1.1=462$; $t=11$: $432\times 0.7=302.4$; $t=12$: $444\times 0.9=399.6$. **Q15.** (a) deseasonalised sales rise by about \$4.2 thousand per quarter on average; (b) five years ahead is a large extrapolation — a new competitor, changed weather patterns or shifting tastes (a structural change) could break the trend. **Q16.** (a) December sales run 45% above the average month; (b) $87\,000\div 1.45=60\,000$; (c) $64\,000\times 1.45=92\,800$. **Q17.** (a) 68, 76, 84, 92, 100, 108, 116, 124; (b) $t=9$: 132; $t=10$: 140; (c) $132\times 0.9=118.8$; $140\times 1.3=182$. **Q18.** (a) $y=151.8-10.09\times 7\approx 81$ thousand; (b) $t=15$ is far beyond the data and a circulation of zero is implausible — the decline would level off long before then, so the linear model cannot be extrapolated that far. **Q19.** (a) 46, 52, 58, 64, 70, 76; (b) $t=7$: $82\times 0.9=73.8$ thousand; (c) $t=8$: $88\times 1.1=96.8$ thousand. **Q20.** (a) 24, 27, 30, 33, 36, 39, 42, 45 — rising by 3 each quarter; (b) $t=9$: $48\times 0.8=38.4$; $t=10$: $51\times 0.9=45.9$; (c) no — three years ahead is $t=20$, far past the data ($t=8$), so the forecast is an unreliable extrapolation.

Fully-worked solutions

Q1. (a) The five weeks are numbered $t=1,2,3,4,5$. (b) The values increase by a constant 3 each week, so the least-squares line fits them exactly: $a=9$, $b=3$. (c) The trend line is $y=9+3t$. (d) The slope 3 means attendance increases by 3 people per week on average. (e) Week 8 is $t=8$, so $y=9+3\times 8=33$ people.

Q2. (a) A seasonal index of 1.2 means the first quarter runs 20% above the average quarter for the year. (b) Dividing each figure by its index: $96\div 1.2=80$, $63\div 0.9=70$, $81\div 0.9=90$, $120\div 1.0=120$. (c) Once the seasonal effect is removed, Q4 has the highest deseasonalised value (120), so it was the strongest quarter.

Q3. (a) At $t=9$, $y=50+4\times 9=86$. (b) The ninth period is a first quarter (indices repeat every four quarters), so the reseasonalised forecast is $86\times 0.8=68.8$. (c) At $t=10$, $y=50+4\times 10=90$; the tenth period is a second quarter, so the reseasonalised forecast is $90\times 1.1=99$.

Q4. (a) Month 8 is $t=8$: $y=20+6\times 8=68$ items. (b) Month 15 is $t=15$: $y=20+6\times 15=110$ items. (c) The month-8 forecast is more reliable: it lies only two months beyond the data, whereas month 15 is nine months beyond, a far larger extrapolation over which the trend is less likely to hold.

Q5. Reseasonalise each forecast by multiplying by that quarter's index: $160\times 0.8=128$, $168\times 1.1=184.8$, $176\times 1.2=211.2$, $184\times 0.9=165.6$.

Q6. (a) Divide each figure by its half-year index (1.2 for H1, 0.8 for H2): $60\div 1.2=50$, $44\div 0.8=55$, $72\div 1.2=60$, $52\div 0.8=65$, $84\div 1.2=70$, $60\div 0.8=75$. (b) The slope is 5: the deseasonalised visitor numbers rise by about 5 thousand per half-year. (c) H1 of year 4 is $t=7$: the trend gives $45+5\times 7=80$, and reseasonalising gives $80\times 1.2=96$ thousand. (d) H2 of year 4 is $t=8$: the trend gives $45+5\times 8=85$, and $85\times 0.8=68$ thousand.

Q7. (a) The slope 15 means the recycling tonnage increases by about 15 tonnes per year on average; the intercept 200 is the model's tonnage at $t=0$, the year immediately before the data began. (b) Year 7 is $t=7$: $y=200+15\times7=305$ tonnes.

Q8. Set the trend line equal to 40: $40=94.5-5.5t$, so $5.5t=54.5$ and $t=54.5\div5.5\approx9.9$. This lies in the tenth year. The prediction should be treated with caution because $t\approx9.9$ is about five years past the last data point ($t=5$); over such a long extrapolation the downward trend may slow, stop or reverse, so the estimate is unreliable.

Q9. (a) In a time series the response is measured at successive times, so its value depends on *when* it was recorded — time is therefore the explanatory variable. Because the recordings are equally spaced, the periods can be labelled $1,2,3,\ldots$, giving a numerical explanatory variable to regress against. (b) Every forecast is an extrapolation beyond the data. One year ahead is a small step, so if the trend holds the error is small; fifteen years ahead is a large step over which the trend may flatten or reverse, or a structural change (new technology, new competitor, policy change) may occur — so the uncertainty grows rapidly with the distance forecast.

Q10. (a) Divide each quarter by its index. Year 1: $92\div1.15=80$, $76\div0.95=80$, $64\div0.80=80$, $88\div1.10=80$. Year 2: $115\div1.15=100$, $95\div0.95=100$, $80\div0.80=100$, $110\div1.10=100$. (b) The deseasonalised figures are constant within each year (80 in year 1, 100 in year 2) and step up between the years, revealing a clear upward trend once the seasonal swing is removed.

Q11. (a) Divide each figure by its quarter's index: $108\div1.2=90$, $70\div0.7=100$, $99\div0.9=110$, $144\div1.2=120$, $156\div1.2=130$, $98\div0.7=140$, $135\div0.9=150$, $192\div1.2=160$. (b) The deseasonalised values rise by a constant 10, so the least-squares line is $y=80+10t$. (c) Year 3 is $t=9,10,11,12$, giving trend values $170,180,190,200$. Reseasonalising with the quarter indices: Q1 $170\times1.2=204$, Q2 $180\times0.7=126$, Q3 $190\times0.9=171$, Q4 $200\times1.2=240$.

Q12. (a) Month 7 is $t=7$: $y=9.2+2.8\times7=28.8$ °C. (b) Month 12 is $t=12$: $y=9.2+2.8\times12=42.8$ °C. (c) Month 12 is six months beyond the data — a large extrapolation — and 42.8 °C is an unrealistically high monthly maximum. A temperature series is seasonal and cannot keep rising along a straight line, so the linear model must not be pushed that far.

Q13. (a) At $t=8$: $y=98.5+20.5\times8=98.5+164=262.5$, i.e. \$262.5 thousand. (b) Solve $98.5+20.5t>300$: $20.5t>201.5$, so $t>9.83$. The first whole quarter beyond this is $t=10$, the tenth quarter.

Q14. Substitute each period into the trend line, then multiply by the season's index. $t=9$ (Q1): $300+12\times9=408$, $408\times1.3=530.4$. $t=10$ (Q2): $300+120=420$, $420\times1.1=462$. $t=11$ (Q3): $300+132=432$, $432\times0.7=302.4$. $t=12$ (Q4): $300+144=444$, $444\times0.9=399.6$.

Q15. (a) The slope 4.2 means the deseasonalised sales increase by about \$4.2 thousand per quarter on average. (b) A forecast five years (20 quarters) ahead is a large extrapolation, so any change in conditions — a new competitor opening nearby, unusual weather, or a shift in consumer tastes (a structural change) — could break the trend and make the forecast badly wrong.

Q16. (a) The index 1.45 means December sales are typically $45\,\%$ above the average month. (b) Deseasonalised value $=87\,000\div1.45=60\,000$. (c) Reseasonalised forecast $=64\,000\times1.45=92\,800$, i.e. \$92 800.

Q17. (a) Divide each figure by its quarter's index: $61.2\div0.9=68$, $98.8\div1.3=76$, $92.4\div1.1=84$, $64.4\div0.7=92$, $90\div0.9=100$, $140.4\div1.3=108$, $127.6\div1.1=116$, $86.8\div0.7=124$. (b) At $t=9$: $y=60+8\times9=132$; at $t=10$: $y=60+8\times10=140$. (c) The ninth period is a first quarter, so $132\times0.9=118.8$; the tenth is a second quarter, so $140\times1.3=182$.

Q18. (a) Year 7 is $t=7$: $y=151.8-10.09\times7=151.8-70.63\approx81.2$, i.e. about 81 thousand. (b) At $t=15$ the line gives $y\approx0$, but a circulation of zero is not plausible — in reality the decline would slow and level off well before then. Because $t=15$ is nine years past the data, this is an extreme extrapolation and the linear trend cannot be relied on that far ahead.

Q19. (a) Divide each figure by its half-year index (0.9 for H1, 1.1 for H2): $41.4\div0.9=46$, $57.2\div1.1=52$, $52.2\div0.9=58$, $70.4\div1.1=64$, $63\div0.9=70$, $83.6\div1.1=76$. (b) H1 of year 4 is $t=7$: the trend gives

$40+6\times7=82$, and reseasonalising, $82\times0.9=73.8$ thousand. (c) H2 of year 4 is $t=8$: the trend gives $40+6\times8=88$, and $88\times1.1=96.8$ thousand.

Q20. (a) Dividing each profit by its quarter's index gives $19.2\div0.8=24$, $24.3\div0.9=27$, $33\div1.1=30$, $39.6\div1.2=33$, $28.8\div0.8=36$, $35.1\div0.9=39$, $46.2\div1.1=42$, $54\div1.2=45$ — a series rising by a constant 3 each quarter. (b) At $t=9$ the trend is $21+3\times9=48$; reseasonalised (Q1), $48\times0.8=38.4$. At $t=10$ the trend is $21+3\times10=51$; reseasonalised (Q2), $51\times0.9=45.9$ — about \$38 400 and \$45 900. (c) No. Three years past the two-year data set is period $t=20$, far beyond the last observed period ($t=8$); such a long extrapolation is unreliable, since the trend may change or a structural change may occur.

Topic 1 Test A — Multiple-choice

One bound reference and one approved technology (CAS calculator or software) are permitted; a scientific calculator may also be used. Reading time: 3 minutes. Writing time: 20 minutes. 12 multiple-choice questions, 12 marks. Choose the response that is ***correct*** *for the question.*

Question 1

A market researcher records the following four variables for each shopper: their postcode, the number of items purchased, their preferred payment method, and the total amount spent (in dollars). How many of these four variables are numerical?

A. 1

B. 2

C. 3

D. 4

Question 2

For a data set, the first quartile is $Q_1 = 18$ and the third quartile is $Q_3 = 30$. Using the $1.5 \times \text{IQR}$ rule, the upper fence for identifying outliers is

A. 48

B. 42

C. 66

D. 36

Question 3

The times taken by swimmers to complete a lap are approximately normally distributed with a mean of 50 seconds and a standard deviation of 4 seconds. Using the 68–95–99.7% rule, the percentage of swimmers who take between 46 and 58 seconds is closest to

A. 68%

B. 95%

C. 99.7%

D. 81.5%

Question 4

The two-way frequency table below shows whether 200 residents own a pet, classified by their type of dwelling.

Owns a pet	House	Apartment
Yes	90	32
No	30	48

The percentage of apartment residents who own a pet is

A. 40%

B. 16%

C. 60%

D. 72%

Question 5

A scatterplot of two numerical variables is approximately linear, and the correlation coefficient is $r = -0.68$. This correlation is best described as

A. strong and negative

B. weak and negative

C. moderate and negative

D. moderate and positive

Question 6

The least-squares line relating the price of a used car, price (in thousands of dollars), to its age (in years) is

$$\text{price} = 28.4 - 2.6 \times \text{age}.$$

On average, for each additional year of age, the price of the car

A. decreases by $2.60

B. increases by $2600

C. decreases by $28 400

D. decreases by $2600

Question 7

In a study, the least-squares regression of a student's exam mark on the number of hours studied gives $r = 0.9$. The coefficient of determination tells us that

A. 90% of the variation in exam mark is explained by the variation in hours studied

B. 81% of the variation in hours studied is explained by the variation in exam mark

C. 81% of the variation in exam mark is explained by the variation in hours studied

D. 81% of all exam marks are predicted correctly

Question 8

A least-squares line is $y = 4.5 + 1.8x$. For a data point with $x = 10$, the actual value is $y = 25$. The residual for this point is

A. 22.5

B. 2.5

C. -2.5

D. 47.5

Question 9

A scatterplot of a population size P against time t is increasing at an increasing rate, consistent with exponential growth. The transformation most likely to linearise this association is

A. $\log_{10}(t)$

B. $\frac{1}{P}$

C. $\log_{10}(P)$

D. t^2

Question 10

A log transformation produces the least-squares line

$$\log_{10}(y) = 1.2 + 0.05x.$$

Using this model, the value of y predicted when $x = 20$ is closest to

A. 2.2

B. 13

C. 1585

D. 158

Question 11

The seasonal index for summer sales at a store is 1.25. The actual sales in summer were \$60 000. The deseasonalised summer sales figure is

A. \$75 000

B. \$48 000

C. \$61 250

D. \$60 000

Question 12

The numbers of visitors to a gallery on six consecutive days were

$$20, 28, 24, 30, 26, 34.$$

The three-point moving mean centred on the third day (24 visitors) is closest to

A. 27.3

B. 24

C. 26

D. 82

End of Topic 1 Test A

Test A — answers and solutions

Q	1	2	3	4	5	6	7	8	9	10	11	12
Answer	B	A	D	A	C	D	C	B	C	D	B	A

Q1. B. Number of items and amount spent are numerical; postcode and payment method are categorical, so 2 variables are numerical.

Q2. A. $\mathrm{IQR} = 30 - 18 = 12$, so the upper fence is $Q_3 + 1.5 \times \mathrm{IQR} = 30 + 1.5 \times 12 = 48$.

Q3. D. $46 = 50 - 1\,\mathrm{sd}$ and $58 = 50 + 2\,\mathrm{sd}$; the percentage between is $68/2 + 95/2 = 34 + 47.5 = 81.5\,\%$.

Q4. A. There are $32 + 48 = 80$ apartment residents, of whom 32 own a pet: $32/80 = 40\,\%$.

Q5. C. $|r| = 0.68$ lies in $[0.5, 0.75)$ (moderate) and the sign is negative.

Q6. D. The slope -2.6 is in thousands of dollars per year, so the price falls by $2.6 \times \$1000 = \$2600$ per year.

Q7. C. $r^2 = 0.9^2 = 0.81$; 81% of the variation in the response (exam mark) is explained by the explanatory variable (hours studied).

Q8. B. Predicted $y = 4.5 + 1.8 \times 10 = 22.5$; residual = actual − predicted $= 25 - 22.5 = 2.5$.

Q9. C. Exponential growth is linearised by compressing the response axis, i.e. a $\log_{10}(P)$ transformation.

Q10. D. $\log_{10}(y) = 1.2 + 0.05 \times 20 = 2.2$, so $y = 10^{2.2} \approx 158$.

Q11. B. Deseasonalised value = actual ÷ index $= 60\,000 \div 1.25 = \$48\,000$.

Q12. A. The three-point moving mean is $\dfrac{28 + 24 + 30}{3} = \dfrac{82}{3} \approx 27.3$.

Topic 1 Test B — Short-answer

One bound reference and one approved technology (CAS calculator or software) are permitted; a scientific calculator may also be used. Reading time: 5 minutes. Writing time: 40 minutes. Total 40 marks. In all questions where a numerical answer is required, an exact value must be given unless otherwise specified. Where more than one mark is available, appropriate working must be shown.

Question 1 (7 marks)

A café owner records the number of customers served each day for 15 days. The values, arranged in order, are

$$15, 18, 20, 22, 23, 25, 26, 28, 30, 31, 33, 35, 38, 40, 62.$$

a. Determine the five-number summary for this data set. 2 marks

b. Using the $1.5 \times \text{IQR}$ rule, show that 62 is an outlier. 2 marks

c. Construct a boxplot for the data, displaying the outlier separately. 2 marks

d. Describe the shape of the distribution, referring to your boxplot. 1 mark

Question 2 (6 marks)

A researcher surveyed 250 people, recording each person's age group (under 30, or 30 or over) and whether or not they exercise regularly. The results are shown below.

	Under 30	30 or over	Total
Exercises regularly	96	27	123
Does not exercise regularly	64	63	127
Total	160	90	250

a. What percentage of people under 30 exercise regularly, and what percentage of people aged 30 or over exercise regularly? 2 marks

b. Do these data support the contention that there is an association between age group and regular exercise? Justify your answer by quoting the percentages from part **a.** 2 marks

c. Identify the explanatory variable and the response variable in this investigation, and state the type (categorical or numerical) of each. 2 marks

Question 3 (8 marks)

The table shows the maximum daily temperature, t (°C), and the number of ice-creams sold, n, at a kiosk on seven days. Temperature is the explanatory variable.

Temperature t (°C)	16	18	20	22	24	26	28
Ice-creams sold n	22	26	33	38	44	48	55

a. Determine the equation of the least-squares regression line, giving the coefficients correct to two decimal places. 2 marks

b. Interpret the slope of the regression line in terms of temperature and ice-creams sold. 1 mark

c. Write down the value of the coefficient of determination, correct to four decimal places, and interpret it in this context. 2 marks

d. Use the regression line to predict the number of ice-creams sold when the maximum temperature is 25 °C. State whether this is an interpolation or an extrapolation, and why. 2 marks

e. Calculate the residual for the day when the temperature was 20 °C. 1 mark

Question 4 (7 marks)

The number of bacteria, y (in thousands), in a culture is recorded each hour for six hours.

Time t (hours)	0	1	2	3	4	5
Bacteria y (thousands)	6	11	19	34	65	114

A scatterplot of y against t is curved, so a $\log_{10}$ transformation is applied to the response variable y.

a. Explain why a least-squares line fitted to the raw data would not be appropriate. 1 mark

b. Complete a table of $\log_{10}(y)$ values, correct to two decimal places. 2 marks

c. Determine the least-squares line for $\log_{10}(y)$ against t, giving the coefficients correct to three significant figures. 2 marks

d. Use your model to predict the number of bacteria after 7 hours, correct to the nearest ten thousand. 2 marks

Question 5 (8 marks)

A swimwear retailer records its quarterly sales (in thousands of dollars) for one year.

Quarter	Q1	Q2	Q3	Q4
Sales ($'000)	120	105	75	100

a. Calculate the four seasonal indices for this year, correct to two decimal places. 3 marks

b. Interpret the seasonal index for Quarter 1. 1 mark

c. In the following year, the actual Quarter 3 sales were $66 000. Calculate the deseasonalised Quarter 3 sales figure. 2 marks

d. The deseasonalised quarterly sales are modelled by the trend line $\hat{s} = 90 + 5t$, where $\hat{s}$ is in thousands of dollars and t is the quarter number ($t = 1$ for Quarter 1 of the first year). Forecast the actual sales for Quarter 1 of the fourth year ($t = 13$). 2 marks

Question 6 (4 marks)

The weights of apples from an orchard are approximately normally distributed with a mean of 150 grams and a standard deviation of 20 grams.

a. Between what two weights do approximately 95% of the apples lie? 1 mark

b. Using the 68–95–99.7% rule, find the percentage of apples that weigh more than 170 grams. 2 marks

c. Calculate the standardised value (z-score) of an apple that weighs 190 grams. 1 mark

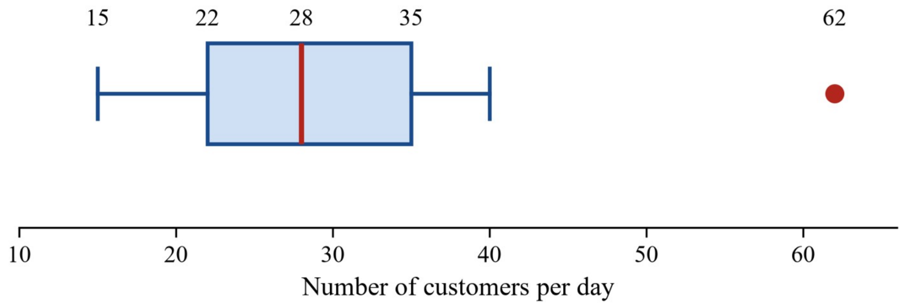
End of Topic 1 Test B

Test B — answers and solutions

Q1. (a) $n = 15$, so the median is the 8th value, $M = 28$. The lower half is the first 7 values, giving $Q_1 = 22$; the upper half is the last 7 values, giving $Q_3 = 35$. With $min = 15$ and $max = 62$, the five-number summary is

$$15, 22, 28, 35, 62.$$

(b) $\text{IQR} = Q_3 - Q_1 = 35 - 22 = 13$. The upper fence is $Q_3 + 1.5 \times \text{IQR} = 35 + 1.5 \times 13 = 54.5$. Since $62 > 54.5$, the value 62 is an outlier. (The lower fence is $22 - 19.5 = 2.5$, and $15 > 2.5$, so there is no low outlier.) (c) The box spans $Q_1 = 22$ to $Q_3 = 35$ with the median at 28; the outlier 62 is plotted separately and the upper whisker stops at the largest non-outlier, 40.



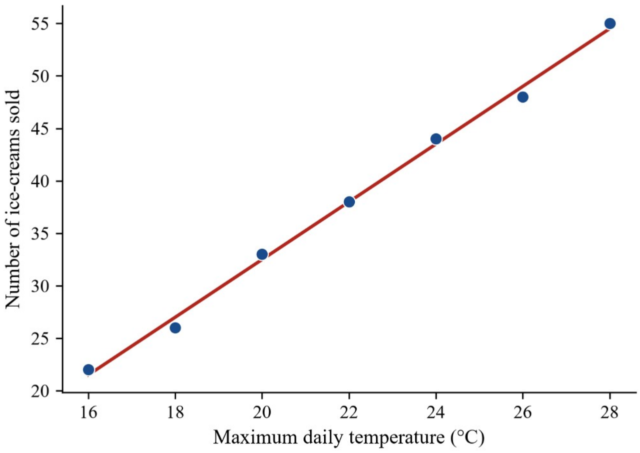
(d) The distribution is positively skewed: the mean (≈ 29.7) exceeds the median (28), and there is a single high outlier stretching the upper tail.

Q2. (a) Under 30: $\dfrac{96}{160} = 60\,\%$ exercise regularly. Aged 30 or over: $\dfrac{27}{90} = 30\,\%$ exercise regularly. (b) Yes. The percentage who exercise regularly differs markedly between the two age groups (60% for those under 30 versus 30% for those 30 or over). Because the percentages are not equal, the data support an association between age group and regular exercise. (c) The explanatory variable is age group and the response variable is whether or not a person exercises regularly. Both variables are categorical.

Q3. (a) Using technology, the least-squares line is

$$n = -22.50 + 2.75\,t.$$

(b) On average, the number of ice-creams sold increases by 2.75 for each 1 °C increase in the maximum daily temperature. (c) $r^2 = 0.9965$. About 99.65% (approximately 99.7%) of the variation in the number of ice-creams sold is explained by the variation in the maximum daily temperature. (d) $n = -22.50 + 2.75 \times 25 = 46.25$, so about 46 ice-creams. Since 25 °C lies within the range of the data (16 °C to 28 °C), this is an interpolation and is therefore reliable. (e) Predicted at $t = 20$: $n = -22.50 + 2.75 \times 20 = 32.5$. The actual value is 33, so the residual is $33 - 32.5 = 0.5$.



Q4. (a) The scatterplot of the raw data is curved (it increases at an increasing rate), so a straight line does not fit the pattern; the residuals would show a clear curved pattern rather than random scatter. (b) The transformed values, correct to two decimal places, are

t	0	1	2	3	4	5
$\log_{10}(y)$	0.78	1.04	1.28	1.53	1.81	2.06

(c) Fitting the least-squares line to $\log_{10}(y)$ against t gives

$$\log_{10}(y)=0.777+0.256t\left(r^2=0.9996\right).$$

(d) At $t=7$: $\log_{10}(y)=0.777+0.256\times 7=2.569$, so $y=10^{2.569}\approx 371$ (thousand). To the nearest ten thousand, the model predicts about **370 000 bacteria**.

Q5. (a) The mean quarterly sales are $\dfrac{120+105+75+100}{4}=\dfrac{400}{4}=100$. Each seasonal index is the quarter's sales divided by this mean:

$$\text{Q1}:120/100=1.20,\text{Q2}:105/100=1.05,\text{Q3}:75/100=0.75,\text{Q4}:100/100=1.00.$$

(The four indices sum to 4, as required.) (b) The Quarter 1 index of 1.20 means that Quarter 1 sales are typically 20% above the average quarterly sales for the year. (c) Deseasonalised value = actual ÷ seasonal index $=\dfrac{66\,000}{0.75}=\$88\,000$. (d) The trend value at $t=13$ is $\hat{s}=90+5\times 13=155$ (thousand dollars, deseasonalised). Reseasonalising with the Quarter 1 index gives $155\times 1.20=186$, i.e. a forecast of $\$186\,000$.

Q6. (a) 95% of values lie within two standard deviations of the mean: $150\pm 2\times 20$, i.e. between **110 grams and 190 grams**. (b) $170=150+20$ is one standard deviation above the mean. Since 68% lie within one standard deviation, 32% lie outside it, split equally between the two tails, so the percentage above 170 grams is $32/2=16\%$. (c) $z=\dfrac{190-150}{20}=\dfrac{40}{20}=2$.