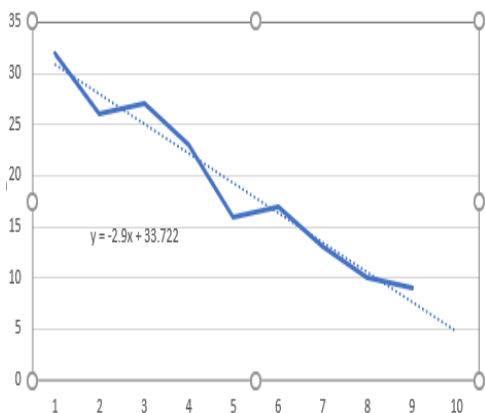


Goals

This brief:

- construct time series plots
- describe time series plots by identifying features such as trend (long term direction), seasonality (systematic, calendar-related movements), and irregular fluctuations (unsystematic, short term fluctuations), and recognise when there are outliers, for example, one-off unanticipated events
- smooth time series data by using a simple moving average, including the use of spreadsheets to implement this process
- calculate seasonal indices by using the average percentage method
- fit a least-squares line to model long-term trends in time series data



Theoretical Components

Resources:

For this week the theory work is in the *PDF file*:
Week 1 and 2 Notes & Exercises

Finding the linear regression equation using Google Sheets:

<https://www.youtube.com/watch?v=0DuvyOKAc1I>

An interesting exercise in using a trend line:

<https://www.khanacademy.org/math/probability/scatter-plots-a1/estimating-trend-lines/v/smoking-1945-extrapolation>

An introduction to smoothing using the mean. Note: we do not use 4 point mean or median smoothing. You can skip these sections:

<https://www.youtube.com/watch?v=U0dF8uJs3mE>

Knowledge Checklist

- Time series
- Trends, outliers, cycles
- Secular, seasonal, cyclic, random
- Trend line
- Forecasting
- Linear regression equation
- Smoothing out fluctuations
- Moving-average smoothing
- Predictions
- Median point smoothing
- Using a spreadsheet

Practical Components

There are questions to be answered in the booklet *Week 1 and 2 Notes & Exercises*

You will need a ruler and access to Excel or Google Sheets.

Investigation

See the end of the exercises

2021 MATHEMATICAL APPLICATIONS 4

Week 1 and 2 Notes and Exercises

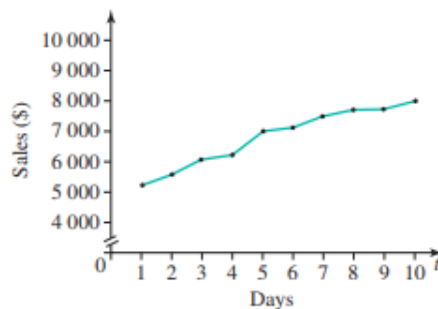
Graphing over time

A time series is a special kind of graph that records successive measurements made over a period of time. It may be a series of seconds, minutes, hours, days, weeks, months, or years, just to name a few. We often (but not always) draw graphs of time series data as line graphs that can be used to make predictions and draw conclusions. We read a time series data graph just like we would with a line graph.

The main purpose of a time series is to see how some quantity varies with time. For example, a company may wish to record its daily sales figures over a 10-day period.

Time	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Sales (\$)	5200	5600	6100	6200	7000	7100	7500	7700	7700	8000

We could also make a graph of this time series and as seen from this graph, there seems to be a trend upwards – clearly, this company is increasing its revenues. **Note:** time is always the x axis.



Other examples of time series data include:

- Fuel prices – they tend to rise and fall in a cyclical pattern according to the day of the week
- Air conditioner sales – they do not remain constant all year, but sales increase and decrease according to the time of year
- Ocean tides – their height is cyclical over a 24-hour period
- Number of cold and flu patients – they also do not remain constant all year, number of cold and flu cases increase and decrease according to the time of year

Types of trends

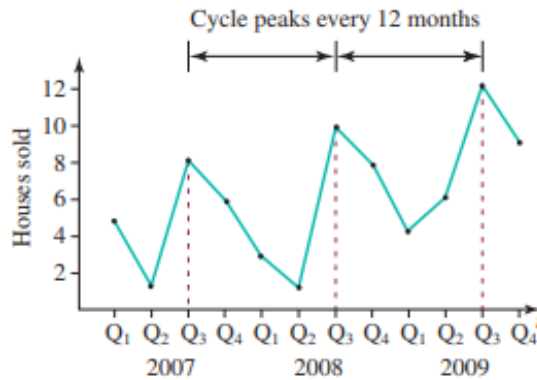
Although many types of trends exist, in this topic we shall be looking at trends that are classified as secular, seasonal, cyclic, and random.

Secular trends

If over a reasonably long period of time a trend appears to be either increasing or decreasing steadily, with no major changes of direction, then it is called a **secular trend**. It is important to look at the data over a long period. If the trend in the graph above continued for 30 days then we could safely conclude that the company was indeed becoming profitable. What appears to be a steady increase over a short term – stock market share prices – can turn out to be something quite different over the long run.

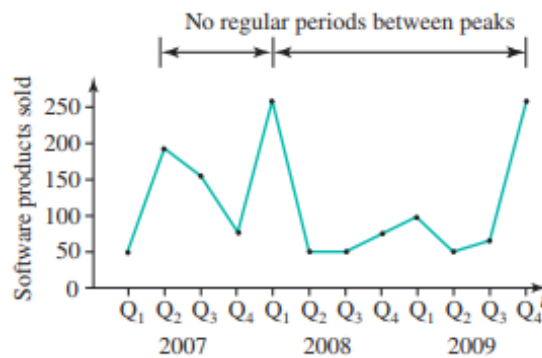
Season trends

Certain data seem to fluctuate during the year, as the seasons change. Consequently, this is termed a seasonal trend. The most obvious example of a seasonal trend would be total rainfall during summer, autumn, winter, and spring in a year. The name seasonal is not specific to the seasons of a year. It could also be related to other constant periods of highs and lows. For example, sales figures at a fast food store could be consistently higher on Saturdays and Sundays and drop off during the weekdays. here the seasons are days of the week and repeat once every week.



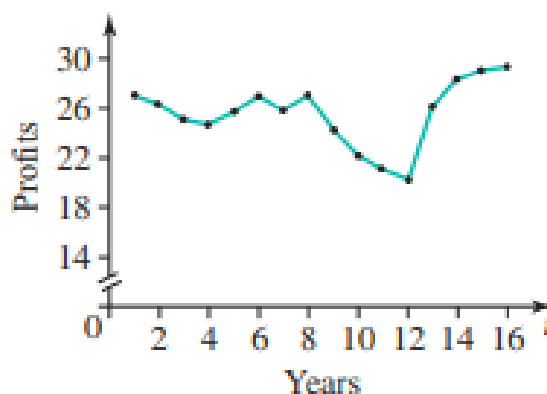
Cyclic trends

Like seasonal trends, cyclic trends show fluctuations upwards and downwards, but not according to season. Businesses often have cycles where at times profits increase, then decline, then increase again.



Random trends

Trends may seem to occur at random. This can be caused by external events such as floods, wars, new technologies or inventions, or anything else that results from random causes. There is no obvious way to predict the direction of the trend or even when it changes direction.



Graphs and language of time series data

When we examine a graph of time series data, there are many noteworthy features.

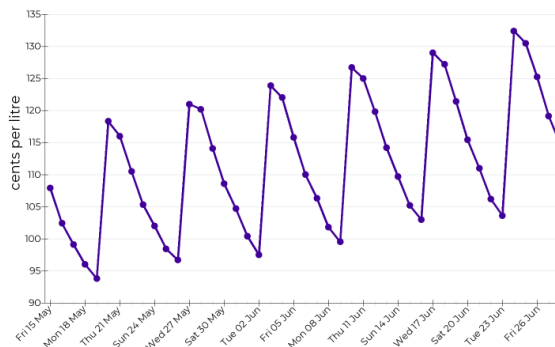
Trend – by looking at the graph, we can see overall whether the response variable is increasing (positive trend) or decreasing (negative trend).

Seasonality – by following the rise and fall, or cyclical nature of the data, we can determine how many seasons or time periods constitute one cycle of the data. This will be helpful to use later.

Irregular Fluctuations – once we have observed the first two features, we will be interested to see whether the data for some seasons or for entire cycles are experiencing irregular fluctuations and unexpected results.

Outliers – by observing the trends and seasonality, we can easily be able to determine an outlier.

The following graph is from the data collected by the Australian Competition and Consumer Commission and it represents the fuel price cycle for Perth in mid-2020.



The overall trend of the data is an increasing trend as we see the peak prices rise. This means petrol is getting more expensive to buy in Perth.

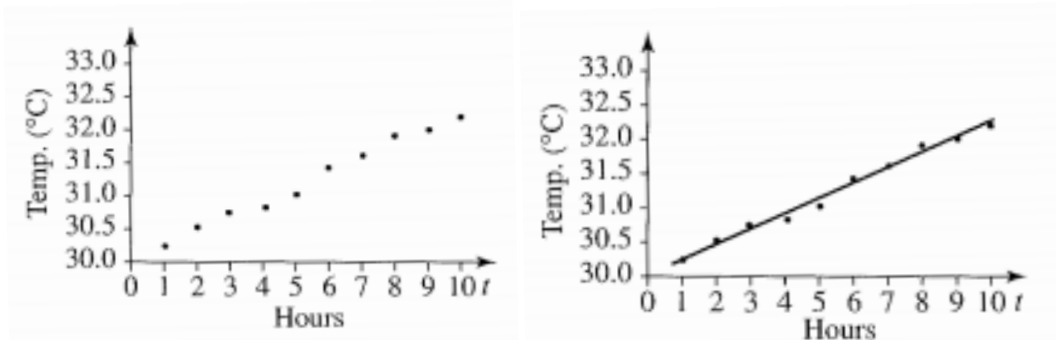
When we count the points in each cycle, we clearly see there are 7 points in each cycle. The easiest way to count these is from peak (highest point) to trough (lowest point) or vice versa. Thus, a cycle lasts a week and there are seven 'seasons', days or time periods per cycle.

It does not appear to have any start fluctuations or outliers.

The trendline

If we want to predict the future values of a trend, it is important to be able to fit a straight line to the data we already have. We could fit the line 'by eye' or 'equal number of points' technique, however, the best method is to use the least squares regression line.

The first graph shows the body temperature of a patient with appendicitis, taken every hour. The second graph shows the trendline fitted.



It is unlikely that the temperature will continue to rise indefinitely, but the line may be significant over a short period of time.

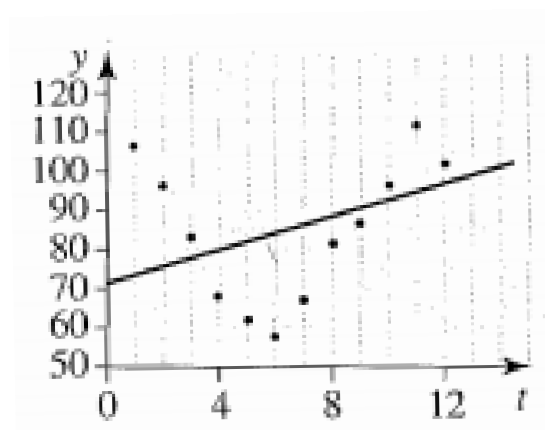
Limitations of fitting trendlines

It is important to note that these techniques are limited to the case where the trend is clearly linear, they cannot be applied effectively to cyclical or seasonal trends.

Consider the following data set which represents sales of swimsuits at a popular clothing store.

Month (x)	1	2	3	4	5	6	7	8	9	10	11	12
Sales (y)	106	96	83	68	62	58	67	81	87	96	111	102

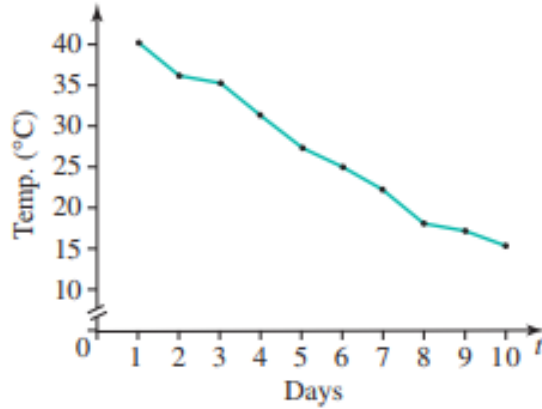
With the data plotted and a trendline fitted, we can see that the trendline is a poor predictor, due to the cycle nature of the trend.



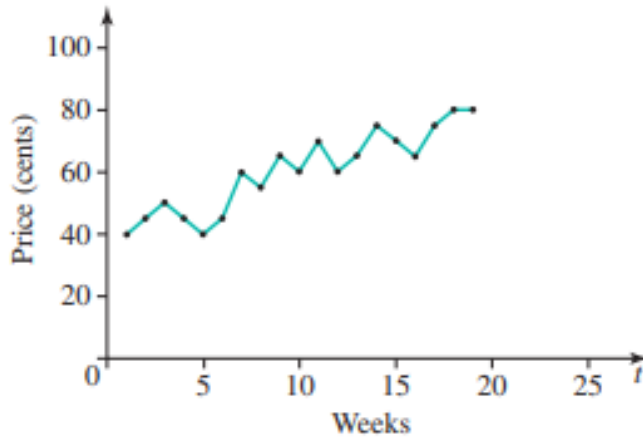
Exercise 1

1. Identify whether the following trends are likely to be secular, seasonal, cyclic, or random for:
 - a. The amount of rainfall, per month, in Western Victorian
 - b. The number of soldiers in the United States army, measured annually
 - c. The number of people living in Australia, measure annually
 - d. The share price of BHP Billiton, measured monthly
 - e. The number of seats held by the Labor Party in the Federal Parliament
 - f. The number of confirmed cases of COVID-19 in Australia

2. Fit a trend line by eye to the data in the graph below.



3. Consider the data in the figure below which represents the price of oranges over a 19-week period. Fit a straight trendline to the data. Use the trendline to predict the price in week 25.



4. A wildlife park ranger is travelling on safari towards the centre of a wildlife park. Each day (t), he records the number of sightings (y) of zebra he sees. He draws up the table shown.

Day(t)	1	2	3	4	5	6	7	8	9	10	11	12
Sightings(y)	6	9	13	8	9	14	15	17	14	11	15	19

Graph the data by hand and fit a trendline. What type of trend is best reflected by these data?

5. Plot the following monthly sales data for umbrellas and fit a trendline. Discuss the type of trend that is best reflected by the data and the limitations of your trendline.

Month(x)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sales(y)	5	10	15	40	70	95	100	90	60	35	20	10

Fitting trendlines and forecasting

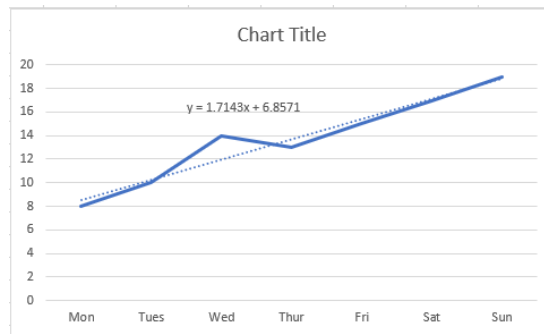
Using our eyes to fit a straight line to a set of data or to predict values can be an inadequate mathematical technique. In the next section, we shall look at using the least squares regression to find the equation of a trendline.

It is important to remember the techniques for fitting trendlines are used on the original data where the trend is clearly linear; that is, random or secular. These techniques cannot be applied effectively to cyclical or seasonal trends. We will be using spreadsheets or Google sheets to plot these graphs. The spreadsheet will also generate the linear regression equation which can be used for predictions.

The data is entered into a spreadsheet, like this:

	A	B
1	Mon	8
2	Tues	10
3	Wed	14
4	Thur	13
5	Fri	15
6	Sat	17
7	Sun	19

The associated graph, trendline and linear regression equation is shown below. See video link on learning brief for step-by-step instructions.



For the sake of predictions, Mon can be considered as 1, Tues as 2, etc. Thus, to predict the value for the first Monday not on the chart, we would use $x = 8$, which gives us $y = 1.7143 \times 8 + 6.8571$, which equals 20.6.

Exercise 2

- The Teeny-Tiny-Tot Company has started to make prams. Its sales figures for the first 8 months are given in the table below.

Date (x)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Sales (y)	65	95	130	115	145	170	190	220

- Use a spreadsheet to plot the data points. Fit a trendline and generate the least squares regression equation. Make sure you attach your graph. .
- Use least squares regression equation to predict the company's sales for December.
- Comment on the suitability of the trendline as a predictor of future trends.

2. The following time series shows the number of new confirmed cases of COVID-19 in the ACT over the following 10 days after restrictions were in place.

Day (x)	1	2	3	4	5	6	7	8	9	10
Number of Cases (y)	9	8	6	1	3	4	4	3	2	3

- a. Plot the data and comment on the linearity of the data.
- b. Fit a trendline and comment on the suitability of the trendline as a predictor of future trends.

3. A mathematics teacher gives her students a test each month for 10 months, and the class average is recorded. The tests are carefully designed to be of similar difficulty.

Test	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Mark (%)	57	63	62	67	65	68	70	72	74	77

- a. Plot the data, fit a trendline and generate the linear regression equations. Attach this graph below.

- b. Use the trendline equation to predict the results for the last exam in December.

- c. Comment on the suitability of the trendline as a predictor.

4. A farmer's yield of soybeans per hectare has been monitored over the last 8 years. By using modern farming methods, the yield has increased most years. There was a drought in 2016, which caused a bad yield. Yields are measured in tonnes per hectare.

Year	2011	2012	2013	2014	2015	2016	2017	2018
Yield	1.2	1.7	2.1	2.3	2.6	1.1	2.5	2.9

- a. Plot the data, fit a trendline and generate the linear regression equations. Attach this below.
- b. Use the trendline equation to predict the yield for 2020.
- c. Predict the yield for 2008.
5. The following table represents the number of cars remaining to be completed on an assembly line.

Time (hours)	1	2	3	4	5	6	7	8	9
Cars remaining	32	26	27	23	16	17	13	10	9

- a. Plot the data, fit a trendline and generate the linear regression equations. Attach this below:

- b. Predict the number of cars remaining to be completed after 11 hours.
- c. At what rate is the number of cars on the assembly line being reduced by?
How do you know?
- d. From the equation of the trendline, we can predict when there are no cars left on the assembly line. This is done by finding the value of x (time) which makes $y = 0$. Using the equation from part a, find the time when there will be no cars left on the assembly line (rearrange the equation).

Smoothing time series

Smoothing time series is a technique used to level out fluctuations and to show a clearer picture of the overall trend. This works best if the data trend is random (rather than seasonal or cyclical). There are several ways to smooth time series.

Moving average smoothing

This technique relies on the principle that averages of data can be used to represent the original data. When applied to time series, a number of the data points are averaged, then we move on to another group of data points in a systematic fashion and average them, and so on. It is generally quite simple. Consider the following example:

Time (t)	Day (y)	Moving average
1	12	
2	10	$\frac{12 + 10 + 15}{3} = 12.3$
3	15	$\frac{10 + 15 + 13}{3} = 12.7$
4	13	$\frac{15 + 13 + 16}{3} = 14.7$
5	16	$\frac{13 + 16 + 13}{3} = 14.0$
6	13	$\frac{16 + 13 + 18}{3} = 15.7$
7	18	$\frac{13 + 18 + 21}{3} = 17.3$
8	21	$\frac{18 + 21 + 19}{3} = 19.3$
9	19	

Notice how the third column in the table is computed from the first two.

1. Take the first three t points (1, 2, 3) and find their average (2); take the first three y points (12, 10, 15) and find their average (12.3). 12.3 is written in the second (2) row.
2. Take the next three t points (2, 3, 4) and find their average (3); take the next three y points (10, 15, 13) and find their average (12.7).
3. Repeat until you reach the last three t points.
4. Take the last three t points (7, 8, 9) and find their average (8); take the last three y points (18, 21, 19) and find their average (19.3).

As we use three points to average, moving down the table from top to bottom, this is called a 3-point moving average smoothing.

This method for smoothing with an odd number (3, 5, ...) is quite simple, and can be done in a vertical table form. It is crucial that the time values are equally spaced, but they do not have to be in the sequence 1, 2, 3. **Note:** there are fewer smoothed points than original ones. For a 3-point smooth, 1 point at either end is 'lost'; while for a 5-point smooth, 2 points at either end are 'lost'.

The main reason for using a smoothing technique is to remove irregularities or wild variations in our time series.

Example

The temperature of a sick patient is measured every 2 hours and the results are recorded.

- a. Create a 3-point moving average smoothing of the data.
- b. Plot both original and smoothed data.
- c. Predict the temperature for 18 hours using the last smoothed value.

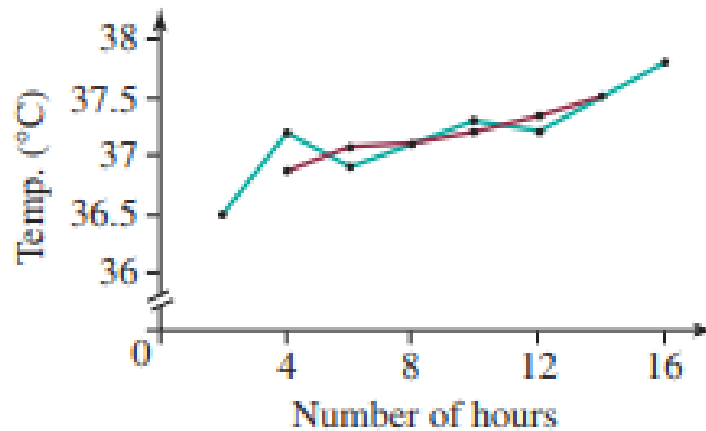
Time (hours)	2	4	6	8	10	12	14	16
Temp (°C)	36.5	37.2	36.9	37.1	37.3	37.2	37.5	37.8

Solution

- a. Create a 3-point moving average smoothing of the data.

Time (hours)	Temp (°C)	Smoothing temp (°C)
2	36.5	
4	37.2	$\frac{36.5 + 37.2 + 36.9}{3} = 36.87$
6	36.9	$\frac{37.2 + 36.9 + 37.1}{3} = 37.07$
8	37.1	$\frac{36.9 + 37.1 + 37.3}{3} = 37.10$
10	37.3	$\frac{37.1 + 37.3 + 37.2}{3} = 37.20$
12	37.2	$\frac{37.3 + 37.2 + 37.5}{3} = 37.33$
14	37.5	$\frac{37.2 + 37.5 + 37.8}{3} = 37.50$
16	37.8	

- b. Plot both original and smoothed data.



- c. Predict the temperature for 18 hours using the last smoothed value.
The smoothed line has removed some of the fluctuation of the original time series and now has a secular trend in temperature. The last smoothed data point is 37.50 and the temperature at 18 hours is predicted to be 37.5°C.

Prediction using moving point averages

The moving average does not generate a single linear equation, there are limited possibilities for using the resultant smoothed data for prediction. However, there are two things that can be done.

1. Predict the next value – use the last smoothed value to predict the next time point. In the example above, our prediction for $t = 18$ is 37.50°C. This is not necessarily an accurate prediction, but it is the best we can do without a linear trend equation.
2. Fit a single straight line to the smoothed data – using the least squares techniques, one could find a single equation for the smoothed data points. This is often the preferred technique.

It is always preferable to use an odd value of points (in the above example, 3 points), regardless of whether the data points are even or odd. The larger the value of points averages, the smoother the trendline of the resulting data becomes, this means more of the fluctuations will be removed. However, you can go too far.

Exercise 3

1. Use the table to complete a 3-point moving average on the following data.

Week	Sales	Smoothed Data	Week	Sales	Smoothed Data
1	34				
2	27		12	44	
3	31		13	47	
4	37		14	49	
5	41		15	41	
6	29		16	52	
7	32		17	48	
8	37		18	44	
9	47		19	49	
10	38		20	56	
11	41		21	54	

2. The following table represents sales of a textbook.

Year (t)	2012	2013	2014	2015	2016	2017	2018	2019
Sales (y)	2250	2600	2400	2750	2900	2450	3100	3400

- a. Create a 3-point moving average of the data using the table. You can use 1 for 2012, 2 for 2013, etc.

- b. Plot both the original and smoothed data on the same axis. Make sure you attach your graph below.

- c. Predict the sales for 2020 using the last smoothed value.

3. The sales of a certain car seem to be declining in the recent months. The management wishes to find out if this is the case.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sales	120	70	100	110	90	80	70	90	80	100	60	60

- a. Using a 3-point moving average, smooth the data and comment on the results. You can use 1 for Jan, 2 for Feb, etc. Use the table for calculating the smoothed data, then use a spreadsheet for the graph. Make sure you attach your graphs.

b. Use the least squares method, find the equation for the smoothed data.
The spreadsheet will calculate this for you.

c. Use the equation to predict the number of sales for March next year.
Comment on the predictions.

4. Consider the quarterly rainfall data below. Rainfall has been measured over a 3-year period. Since the data is seasonal, perform a 3-point moving average and comment on whether there is a trend other than the seasonal one. You will need a graph to determine this. You can use a spreadsheet.

Time (t)	Spring 2006	Summer 2006	Autumn 2007	Winter 2007	Spring 2007	Summer 2007	Autumn 2008	Winter 2008	Spring 2008	Summer 2008	Autumn 2009	Winter 2009
Rainfall (mm)	100	50	65	120	90	50	60	110	85	40	50	100

Use 1 for Spring 2006, 2 for Summer 2006, etc.

Moving average smoothing with a spreadsheet

A spreadsheet can be used to calculate the average data values and then the new set of smoothed points plotted on a graph. Shown is a section of the spreadsheet for the **example about the temperature of a sick patient used previously**. The graph is shown in the example.

	A	B	C	D
1	Time	Temp	Moving average	
2	2	36.5		
3	4	37.2	36.86666667	
4	6	36.9	37.06666667	
5	8	37.1	37.1	
6	10	37.3	37.2	
7	12	37.2	37.33333333	
8	14	37.5	37.5	
9	16	37.8		
10				

Shown below are the formulas used. Note the row and column numbers carefully. It should be clear how to turn this into a 5-point or 7-point smooth but since there are only 8 points, the results would be meaningless.

	A	B	C	D
1	Time	Temp	Moving average	
2	2	36.5		
3	4	37.2	=SUM(B2:B4)/3	
4	6	36.9	=SUM(B3:B5)/3	
5	8	37.1	=SUM(B4:B6)/3	
6	10	37.3	=SUM(B5:B7)/3	
7	12	37.2	=SUM(B6:B8)/3	
8	14	37.5	=AVERAGE(B7:B9)	
9	16	37.8		
10				

Exercise 4

1. The attendance at the Football Club games was recorded over 10 years. The management wants to determine if there is a trend. Use a spreadsheet to perform a 3-point moving average on the data and comment on the results. Make sure you attach your graphs to your booklet.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Attendance (thousands)	75	72	69	74	66	72	61	64	69	65

2. The sales of a new car can vary because of advertising and promotion. The sales figures for a car dealer's new sedan are shown in the table.

Month	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sales	141	270	234	357	267	387	288	303	367	465	398

Use a spreadsheet to:

- a. Perform a 3-point moving average to smooth the data.

- b. Plot both the original and smoothed data on the same axis. Make sure you attach your graph below.

- c. Find the least squares regression equation and use it to predict sales for the next month.

3. A large building site requires varying numbers of workers. The weekly employment figures over the last 7 weeks have been recorded.

Week	1	2	3	4	5	6	7
Employees	67	78	54	82	69	88	94

Use a spreadsheet to:

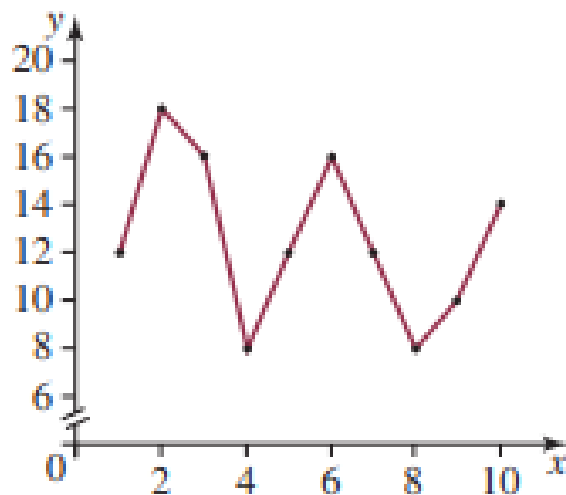
- a. Perform a 3-point moving average to smooth the data.
- b. Plot both the original and smoothed data on the same axis. Make sure you attach your graph below.
- c. Find the least squares regression equation and use it to predict the number of employees for the next week.

Median point smoothing

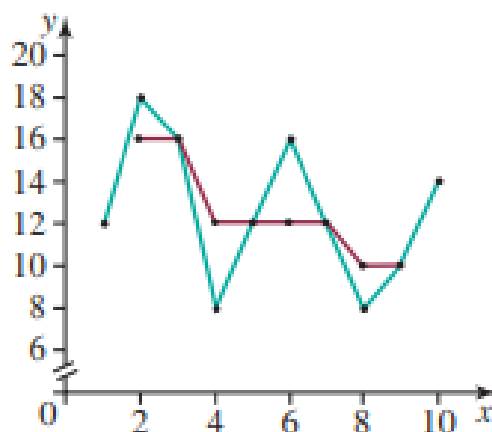
An alternative to moving average smoothing is to replace the averaging of a group of points with the median of each group. Although no particular mathematical advantage is gained, it is a faster technique requiring no calculations (provided you use odd point median smoothing). Often it can be done directly on a graph of a time series.

Example

Perform a 3-point median smoothing on the graph of a time series.



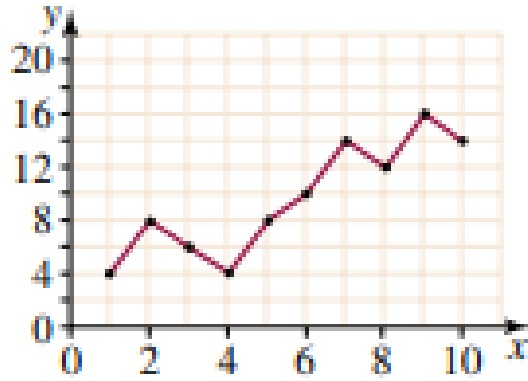
The 1st data points are 12, 18, 16 – median = 16.
The 2nd data points are 18, 16, 8 – median = 16.
The 3rd data points are 16, 8, 12 – median = 12.
Continue until you have used up all of the data points.
Plot the medians on the graph.



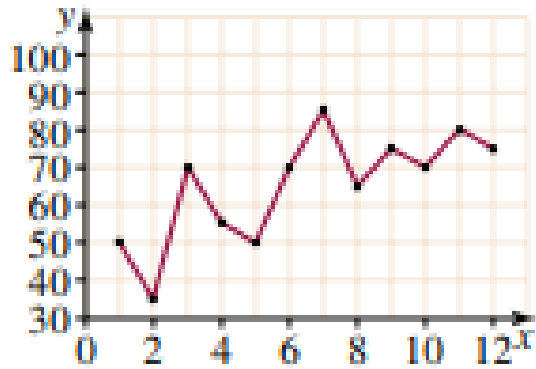
The median smoothing has indicated a downward trend that is probably not in the real time series. This indicates that moving average smoothing would be the preferred option.

Exercise 5

1. Perform a 3-point median smooth on the graphical time series. Comment on the effectiveness of the results.



2. Perform a 3-point median smooth on the graphical time series. Comment on the effectiveness of the results.



2021 MA4 Week 1/2 Investigation

Perform a 5-point moving average to smooth the data in the following table, which represents the number of new cases of COVID-19 in Australia between March 10 and April 18, 2020.

Graph the original and smoothed data on the same axis.

Comment on the nature of the trend and provide one interesting observation.

Day	Cases	Day	Cases	Day	Cases	Day	Cases
1	21	11	167	21	265	31	90
2	16	12	196	22	310	32	100
3	28	13	281	23	304	33	89
4	42	14	364	24	272	34	21
5	50	15	430	25	222	35	46
6	50	16	285	26	195	36	41
7	78	17	374	27	135	37	47
8	77	18	374	28	108	38	21
9	113	19	460	29	113	39	55
10	142	20	346	30	105	40	42

Marking Rubric

Week 1 and 2

Name:

CRITERIA	EXPECTATIONS	POSS	MULT	GIVEN	TOTAL
Practical	Student completes practical work, including exercises and any Mathspace and/or other tasks, of the brief to an acceptable standard set by the teacher.	2	3		/6
Investigation Task	Student completes the investigation task of the week to an acceptable standard set by the teacher.	2	2		/4
Reasoning and Communications	Student responses are accurate and appropriate in presentation of mathematical ideas, with clear and logical working out shown.	4	-		/4
Concepts and Techniques	Student submitted work selects and applies appropriate mathematical techniques to solve practical problems and demonstrates proficiency in the use of mathematical facts, techniques, and formulae.	4	-		/4
	Submission Guidelines				
Timeliness	Student submits the practical work, including exercises and any Mathspace and/or other tasks, and investigation by the set deadline. See scoring guidelines for specific details.	2	-		/2
				FINAL	/20

Student Reflection: How did you go with this week's work? What did you learn? What did you find easy?
 What do you need to work on?