General Maths Unit 3

Core Topic: Data Analysis
Chapter 5
Investigating and Modelling Time Series

Area of Study 1 – Data Analysis

- Qualitative features of time series plots; recognition of features such as trend (long-term direction), seasonality (systematic, calendar related movements) and irregular fluctuations (unsystematic, short-term fluctuations); possible outliers and their sources, including one-off real world events, and signs of structural change such as a discontinuity in the time series

- numerical smoothing of time series data using moving means with consideration of the number of terms required (using centring when appropriate) to help identify trends in time series plot with large fluctuations

- graphical smoothing of time series plots using moving medians (involving an odd number of points only) to help identify long-term trends in time series with large fluctuations

- seasonal adjustment including the use and interpretation of seasonal indices and their calculation using seasonal and yearly means

- modelling trend by fitting a least squares line to a time series with time as the explanatory variable (data de-seasonalised where necessary), and the use of the model to make forecasts (with re-seasonalisation where necessary) including consideration of the possible limitations of fitting a linear model and the limitations of extending into the future.

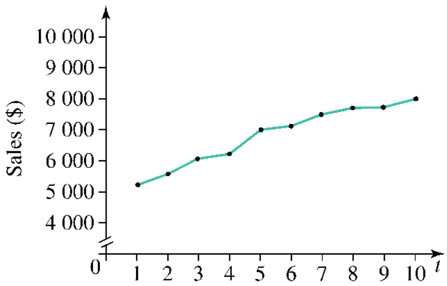
Time series and trend lines – Ex 5a

- When the x-variable (explanatory) is time, and generally, where time goes up in even increments such as hours, days, weeks or years, we have what is called a **time series.**
- The main purpose of time series 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.

Example

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 as shown below.



Looking for patterns in time series plots

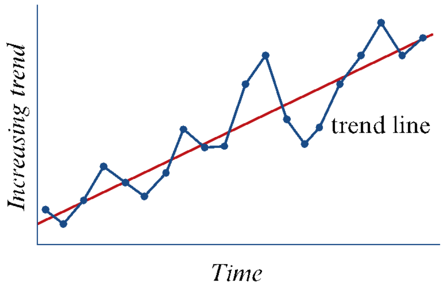
The features we look for in a time series are:

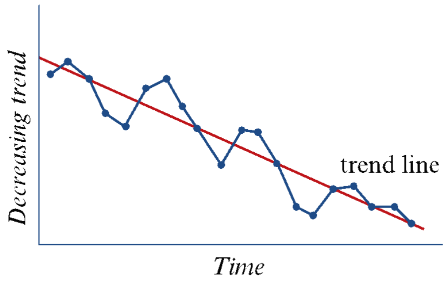
- Trend
- Cycles
- Seasonality
- Structural change
- Possible outliers
- Irregular (random) fluctuations

Trend

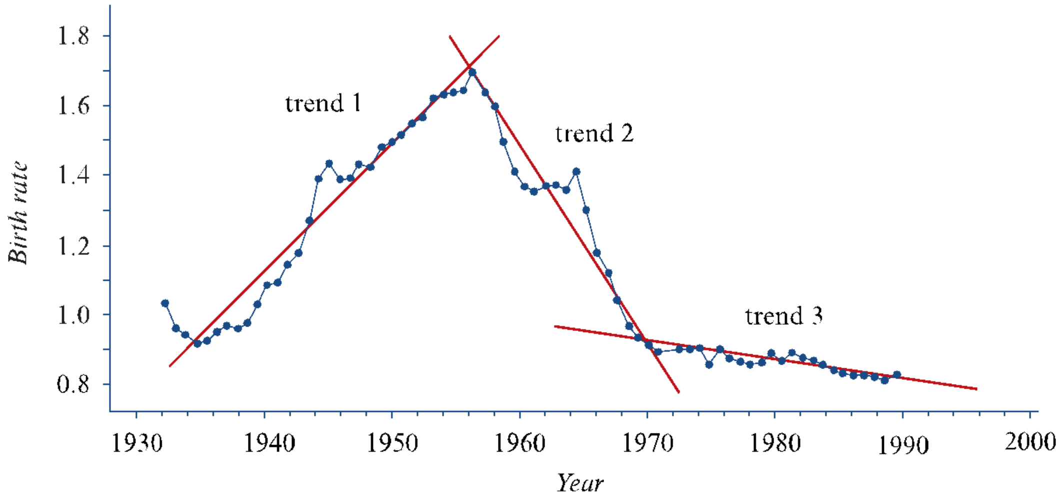
The trendency for values in a time series to generally **increase** or **decrease** over a significant period of time is called a trend

***Note – describe trend as increasing/decreasing not positive/negative*



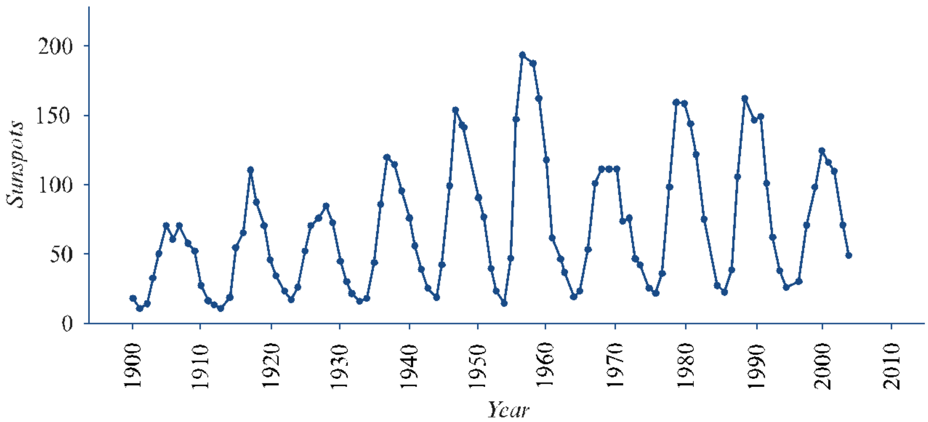


Sometimes different trends are apparent



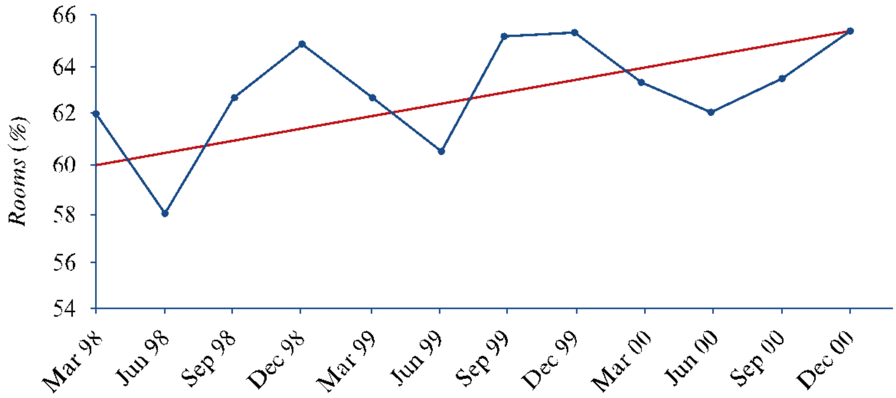
Cycles

Cycles are periodic movements in a time series, but over a period greater then 1 year. (e.g. interest rates)



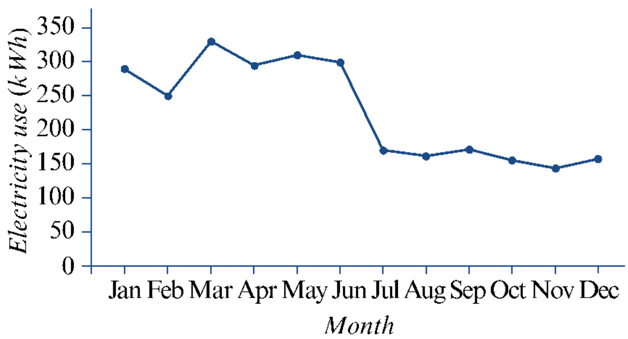
Seasonality
Seasonality is present when there is a periodic movement in a time series that has a calendar related period – for example a year, a month or a week.

The plot below shows the total percentage of rooms occupied in hotels, motels and other accommodation in Australia by quarter, over the years 1998–2000.



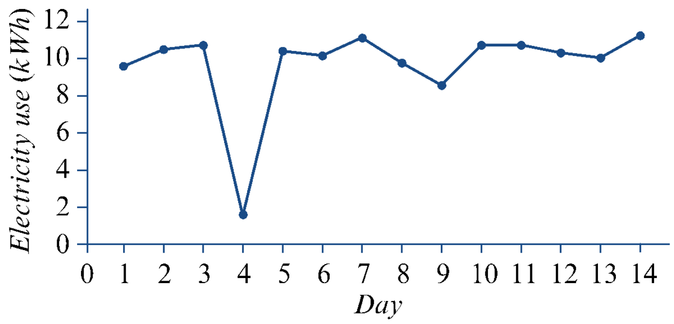
Structural Change
Structural change is present when there is a sudden change in the established pattern of a time series plot.

The time series plot below shows the power bill for a rental house (in kWh) for the 12 months of a year.



Outliers
Outliers are present when there are individual values that stand out from the general body of data.

The time series plot below shows the daily power bill for a house (in kWh) for a fortnight.



Irregular (random) fluctuations

Irregular fluctuations include all the variations in a time series that we cannot reasonably attribute to systematic changes like trend, cycles, seasonality and structural change or an outlier.

Plotting time-series data

- It is useful to plot the data in order to make judgements.
- Similar to a scatterplot but with some differences –
  - The x-variable (explanatory) is always time: e.g. days, weeks, months, quarters
  - The x variables may be labels (e.g day 2) and not numerical
  - The points are connected as they occur in chronological order. Joining them allows us to see the type of trend or time-series pattern.

****If using CAS an **association table** may need to be used so that time periods can be converted to numerals

Association Tables

Example 1

Week 1 Mon.	Week 1 Tues.	Week 1 Wed.	Week 1 Thurs.	Week 1 Fri.	Week 1 Sat.	Week 1 Sun.	Week 2 Mon.	Week 2 Tues.
1	2	3	4	5	6	7	8	9

Example 2

Jan. 2009	Feb. 2009	Mar. 2009	Apr. 2009	May 2009	June 2009	July 2009	Aug. 2009
1	2	3	4	5	6	7	8

Constructing a time series plot using CAS

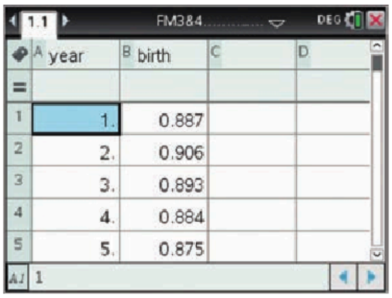
Construct a time series plot for the data presented below. The years have been recoded as 1, 2, . . . , 12, as is common practice.

2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1	2	3	4	5	6	7	8	9	10	11	12
0.887	0.906	0.893	0.884	0.875	0.861	0.855	0.848	0.846	0.844	0.833	0.848

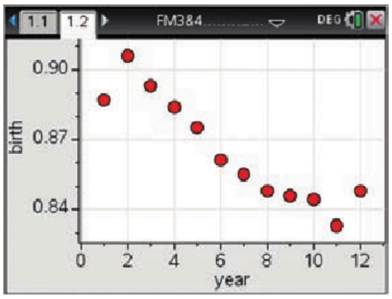
Steps

1 Start a new document by pressing ctrl + N.

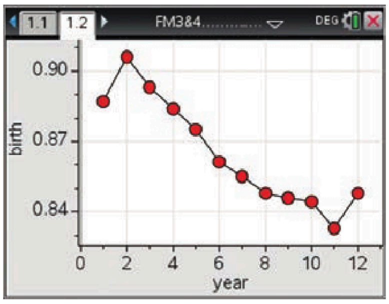
2 Select **Add Lists & Spreadsheet.** Enter the data into lists named *year* and *birth*.



3 Press ctrl + I and select **Add Data & Statistics.** Construct a scatterplot of *birth* against *year*.
Let *year* be the explanatory variable and *birth* the response variable.



4 To display as a connected time series plot, move the cursor to the main graph area and press ctrl + menu>**Connect Data Points.** Press enter.



Ex 5a

Smoothing Time Series – Ex 5b

When data fluctuates a lot, it is often hard to see the underlying trend. In order to reveal the trend, we may need to try and remove some of the fluctuations before attempting to fit the trend line. This process is referred to as smoothing.

Smoothing is removing fluctuations

Two basic types for smoothing random or cyclical variation

- **Median smoothing**
    - Preferred where there are small data sets
    - Can be done graphically on a time series plot
    - Can be used when there are many outliers as the median is not affected by outliners
- **Moving average smoothing.**
    - This option is preferred for data sets with few random fluctuations

Moving Average Smoothing

Averages of data are used to represent the original data.
A number of data points are averaged, then we move on to another group of data points in a systematic way and avereage them and so on
Example

Time (t)	Data (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	

*This is an example of **3-point-moving** average smoothing as we are using 3 points
*The number of points used can vary, e.g. **4-point, 5-point etc.**
*It is preferential to use an odd number, however an even number can be used the method is just a little different.
*There are fewer smoothed points than original points.

Example

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

a **Use a 3-point moving-average technique to smooth the data.**

b **Plot both original and smoothed data on the same set of axes.**

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

Step 1

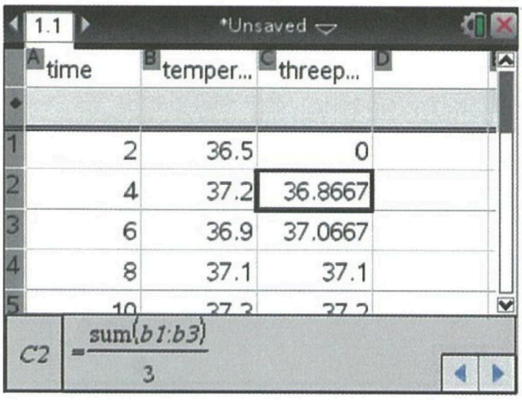
Put the data in a new table to "smooth"

Time	Temp	Smoothed Temp

b) Use CAS

a

1. On a Lists & Spreadsheet page, enter time values into column A and temperature values into column B. Label each accordingly.
2. Label column C as *threepointsmooth.* To enter the rule for the 3-point moving average, that is:
$\dfrac{t_1 + t_2 + t_3}{3}$,
start in cell C2 and complete the entry line as:
$= \dfrac{sum(b1:b3)}{3}$
Fill down until the end of the data list. Check against the spreadsheet formula on the template presented earlier.



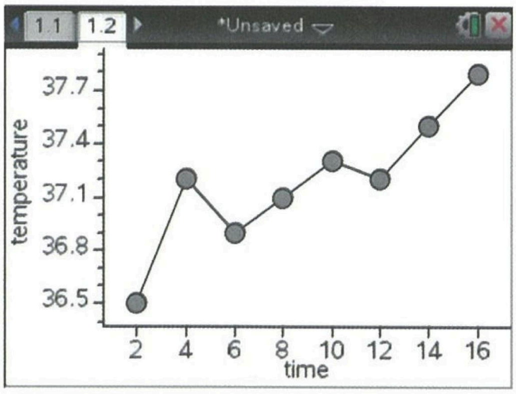
Fill down:
Menu → Data
→ Fill
→ Drag
down

c) To plot…..

1 Add a Data & Statistics page. Click on the graph to add time to the *x*-axis and both *temperature* and *threepointsmooth* on the *y*-axis. To add a second *y*-variable, press:

- MENU [menu]
- 2: Plot Properties [2]
- 8: Add Y Variable [8]

b

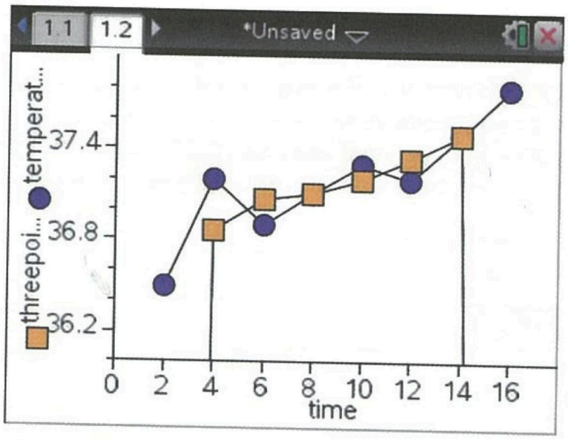


The time-series plot should appear as shown.

If the data points are not joined, press:

- MENU [menu]
- 2: Plot Properties [2]
- 1: Connect Data Points [1]

Note: The vertical lines on the smoothed time series are to be ignored as the first and last points are lost when performing a 3-point smooth.



How many points should be in the moving average???

1. For small data sets, the value of p should be *considerably* smaller than n. For example, if $n = 7$, p should be no more than about 4.
2. The most common practice if there is a cyclic variation that we want to remove, is to let $p \approx$ length of cycle. If the data shows seasonal variation, p should be proportional to the number of seasons. For example, if there are quarterly data with seasonal variation, use $p = 4$.

   Other preferred choices for number of points used in the smoothing procedures include:

Monthly sales figures	use 12 point centred
Daily sales for a store open each day of the week	use 7 point
Daily sales for a store open from Monday to Friday only	use 5 point
Quarterly electricity consumption figures	use 4 point centred
3. It is *always* preferable to use an odd value of p, regardless of whether n is even or odd.
4. The larger the value of p, the smoother the trend line of the resulting data becomes. More of the fluctuations will be removed. However, you can go too far as the bigger p is, the more data points are lost: when $p = 3$, we lose 2 points, if $p = 5$ – 4 points, if $p = 7$ – 6 points etc. Therefore if the set is small, we may lose nearly all data by selecting large value of p!

Example 2

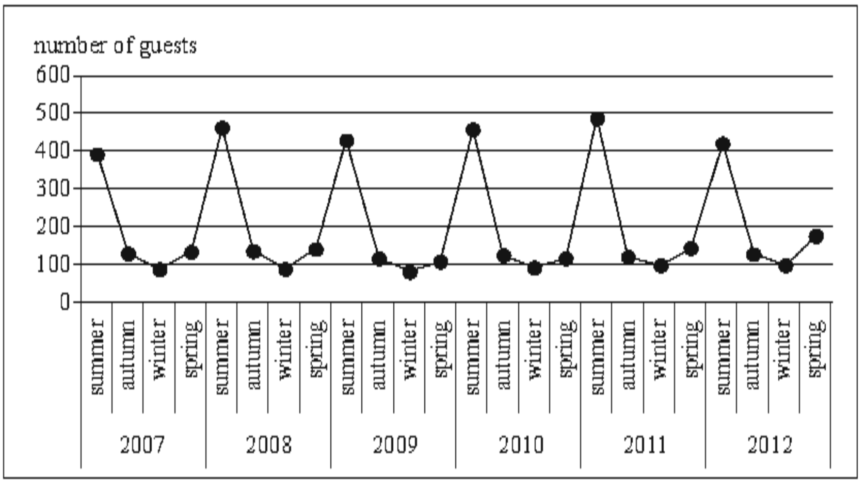
The sales of a new car can vary due to the effect of advertising and promotion. The sales figures for Nassin Motor Company's new sedan are shown in the table. Use 5-point moving averages to smooth the data. Plot the data, and use the last smoothed value to predict sales for the next month.

Month	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Sales	141	270	234	357	267	387	288	303	367	465	398

Exam Questions - 2013

Use the following information to answer Questions 12 and 13.

The time series plot below displays the number of guests staying at a holiday resort during summer, autumn, winter and spring for the years 2007 to 2012 inclusive.



Question 12

Which one of the following best describes the pattern in the time series?

A. random variation only

B. decreasing trend with seasonality

C. seasonality only

D. increasing trend only

E. increasing trend with seasonality

Youtube Clip

https://www.youtube.com/watch?v=L6A3dzWUtWM

Smoothing with an even number of points

- Preferable to use an odd number of points
- When using an odd number the points are automatically "centred" to match up with a time value
- A little more work needs to be done for an even amount of points!!!
- We add an extra column which is called "centring"

Example

		4-point average (smoothed value)		4-point average after centring	
Time	***y*-value**	**Calculation**	**Result**	**Calculation**	**Result**
2006	6				
2007	10				
		$(6 + 10 + 14 + 12) \div 4$	10.5		
2008	14			$(10.5 + 11.75) \div 2$	11.125
		$(10 + 14 + 12 + 11) \div 4$	11.75		
2009	12			$(11.75 + 13) \div 2$	12.375
		$(14 + 12 + 11 + 15) \div 4$	13		
2010	11			$(13 + 13.5) \div 2$	13.25
		$(12 + 11 + 15 + 16) \div 4$	13.5		
2011	15				
2012	16				

Note

- The first result, 10.5 is not lined up with a time value

Plan for smoothing with an even number

1. Average the first 4 amounts
2. Find the average of the first two smoothed points and align it with the 3rd point
3. Find the average of the next two smoothed points and align it with the 4th time point
4. Repeat to the end!

You tube clips

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

Ex 5B

Median Smoothing – Ex 5c

- An alternative to moving-average smoothing
- We find the median of each group rather than averaging them
- Faster technique with no calculations
- Often it can be done directly on a graph of time series

Median smoothing from a table

- Look at each group of 3 points and choose the middle value
- Progress through the table one point at a time
- Points will be lost at the beginning and end of the table

Example

Perform a 3-point median smoothing on the data in the table below. The table shows the cost of an airline ticket between Perth and Melbourne over an 8-month period. Construct a time-series plot of the original data and smoothed data on the same set of axis.

Time	1	2	3	4	5	6	7	8
Cost ($)	340	350	320	340	300	330	350	310

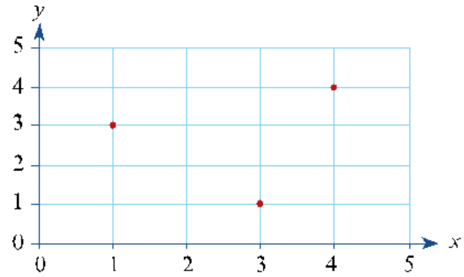
3-Median Smoothing

Time									
Cost									

Median Smoothing from a graph
Locating Medians Graphically

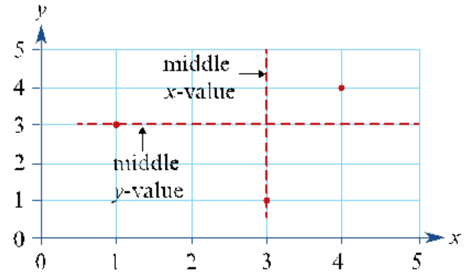
Locating medians graphically

The graph opposite shows three data points plotted on set of coordinate axes. The task is to locate the median of these three points. The median will be a point somewhere on this set of coordinate axes. To locate this point we proceed as follows.



Step 1

Identify the middle data point moving in the x-direction. Draw a vertical line through this value as shown.



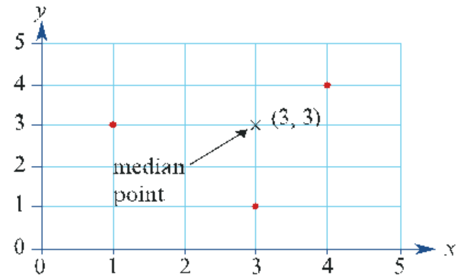
Step 2

Identify the middle data point moving in the y-direction. Draw a horizontal line through this value as shown.

Step 3

The median value is where the two lines intersect – in this case, at the point (3, 3).

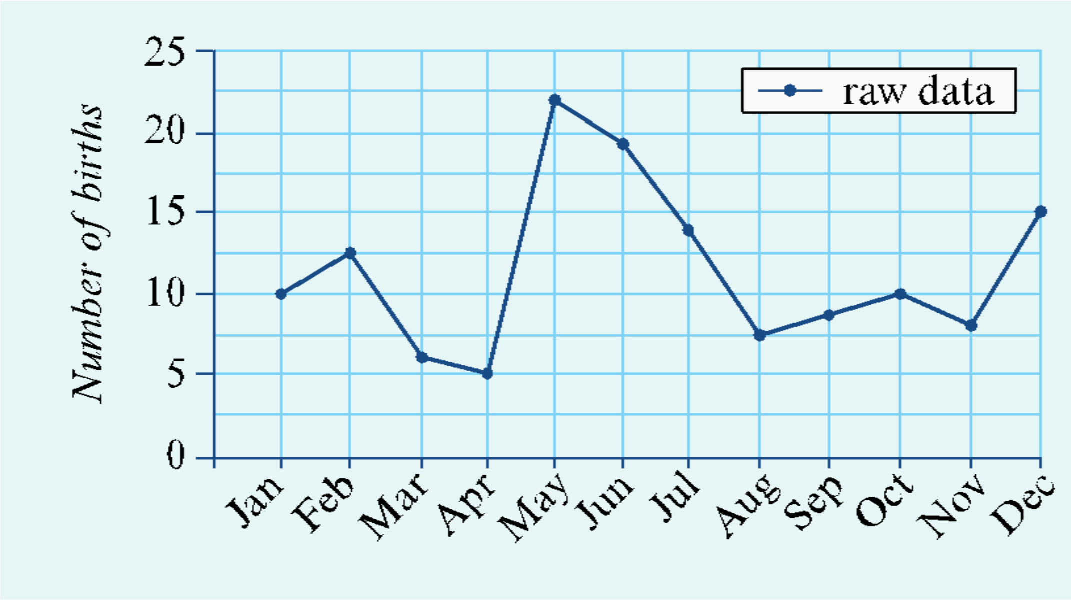
Mark this point with a cross (×).



The process of graphically smoothing a time series plot requires no more than repeating the above process for each group of three or five data points in the plot as required. The following worked examples demonstrate the process.

Example

Construct a five-median smoothed plot of the time series shown below



Ex 5c

Seasonal Indices – Ex 5d

- Seasonal trend is similar to a cyclical trend where there are defined peaks and troughs in the time series data, except for one notable difference -
- Seasonal trends have a fixed and regular period of time between one peak and the next peak in the data values.
- Conversely, there is a fixed and regular period of time between one trough and the next trough.

Important points…

- It is difficult to fit a straight line to seasonal data.
- Seasonal data does not lend itself to "smoothing"
- We may have to accept that the data varies from season to season and treat each record individually.

****Deseasonalising a time series involves replacing the original time series with another one where most or all of the seasonal variation is removed.**

The concept of a seasonal index

Consider the (hypothetical) monthly seasonal indices for unemployment given in the table.

Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sept*	*Oct*	*Nov*	*Dec*	*Total*
1.1	1.2	1.1	1.0	0.95	0.95	0.9	0.9	0.85	0.85	1.1	1.1	12.0

Key fact 1

Seasonal indices are calculated so that their *average* is *1*. This means that the *sum* of the seasonal indices equals the *number of seasons*.

Thus, if the seasons are months, the seasonal indices add to 12. If the seasons are quarters, then the seasonal indices would to 4, and so on.

Key fact 2

Seasonal indices tell us how a particular season (generally a day, month or quarter) compares to the *average season*.

For example:

- seasonal index for unemployment for the month of February is 1.2 or 120%.
  This tells us that February unemployment figures tend to be 20% *higher* than the monthly average. Remember, the average seasonal index is 1 or 100%.
- seasonal index for August is 0.90 or 90%.
  This tells us that the August unemployment figures tend to be only 90% of the monthly average. Alternatively, August unemployment figures are 10% *lower* than the monthly average.

Calculating Seasonal Indices

1. Calculate **seasonal indices.**
   - Average over all seasons for each year – these are the yearly averages
   - Divide each point in the original time series by its corresponding yearly average
   - Using this new series, average over all years for each season – these are the seasonal indicies.

2. **Deseasonalise** the data by dividing each point in the original time series by its corresponding seasonal index.

$$\text{Deseasonalised figure or value} = \frac{\text{actual original figure or value}}{\text{seasonal index}}$$

- To seasonalise (predicted) figures:

  Seasonalised figure or value = deseasonalised figure or value × seasonal index.
- The sum of the seasonal indices is equal to the number of seasons.

Explaining seasonal index (plural indicies)

Seasonal index = 1.3 means that season is 1.3 times the average season (that is, the figures for this season are 30% above the seasonal average). It is a peak or high season.

Seasonal index = 0.7 means that season is 0.7 times the average season (that is the figures for this season are 30% below the seasonal average). It is a trough or low season.

Seasonal index = 1.0 means that season is the same as the average season or neither a peak nor a trough.

Example 1

The seasonal index (SI) for cold drink sales for summer is SI = 1.33.

Last summer a beach kiosk's actual cold drink sales totalled \$15 653.

What were the *deseasonalised* sales?

Reseasonalising Data

Reseasonalising data

Time series data are reseasonalised using the rule:

$$\text{actual figure} = \text{deseasonalised figure} \times \text{seasonal index}$$

Example 2

The seasonal index for cold drink sales for spring is SI = 0.85.

Last spring a beach kiosk's deseasonalised cold drink sales totalled $10 870.

What were the *actual* sales?

Example

The quarterly sales figures for Mikki's shop over a 3-year period are given below.

Year	*Summer*	*Autumn*	*Winter*	*Spring*
1	920	1085	1241	446
2	1035	1180	1356	541
3	1299	1324	1450	659

Use the seasonal indices shown to deseasonalise these sales figures. Write answers correct to the nearest whole number.

Summer	*Autumn*	*Winter*	*Spring*
1.03	1.15	1.30	0.52

Deseasonalised sales figures

Year	Summer	Autumn	Winter	Spring
1				
2				
3				

Calculating seasonal indicies

Example 10 Calculating seasonal indices (1 year's data)

Mikki runs a shop and she wishes to determine quarterly seasonal indices based on last year's sales, which are shown in the table opposite.

Summer	*Autumn*	*Winter*	*Spring*
920	1085	1241	446

Seasonal Indices

Summer	Autumn	Winter	Spring

Example

Unemployment figures have been collected over a 5-year period and presented in this table. It is difficult to see any trends, other than seasonal ones.

a Calculate the seasonal indices.

b Deseasonalise the data using the seasonal indices.

c Plot the original and deseasonalised data.

d Comment on your results, supporting your statements with mathematical evidence.

eBook plus

TUTORIAL
eles-1266
Worked example 9

Season	2005	2006	2007	2008	2009
Summer	6.2	6.5	6.4	6.7	6.9
Autumn	8.1	7.9	8.3	8.5	8.1
Winter	8.0	8.2	7.9	8.2	8.3
Spring	7.2	7.7	7.5	7.7	7.6

Part 1 – Calculating Seasonal Indices

Step 1 – Find the yearly average

Year	2005	2006	2007	2008	2009
Average					

Step 2

Divide each term in the original time series by its yearly average

Year	2005	2006	2007	2008	2009
Summer					
Autumn					
Winter					
Spring					

Step 3

To find the seasonal indices, calculate the average for each season

Year	Summer	Autumn	Winter	Spring
Seasonal Index				

**Note the seasonal indicies should always add up to the total amount of seasons, in this case 4.

i.e. if adjusting months the seasonal indicies should add to 12

Part 2 – Deseasonalize the Data (Take out the fluctuations)

Step 4

Divide each term in the original series by its seasonal index. That is, dived all summer figures by the summer seasonal index, and do the same for the other seasons

Year	2005	2006	2007	2008	2009
Summer					
Autumn					
Winter					
Spring					

Step 5

Plot the orginal and the deseasonalized data.

What do you notice???

Correcting for seasonality (Important)

Also, using the rule

$$\text{deseasonalised figure} = \frac{\text{actual figure}}{\text{seasonal index}}$$

we can work out how much we need to increase or decrease the actual sales figures to correct for seasonality.

For example, we see that for winter:

$$\text{deseasonalised figure} = \frac{\text{actual figure}}{1.30}$$
$$= 0.769\ldots \times \text{actual figure} \approx 77\% \text{ of the actual figures}$$

Thus, to correct the seasonality in winter, we need to decrease the actual sales by about 23%.

Similarly we can show that, to correct for seasonality in spring ($SI_{spring} = 0.52$), we need to increase the actual spring sales figure by around 92% ($\frac{1}{0.52} \approx 1.92$).

Example (2011 VCAA)

Question 12

The seasonal index for headache tablet sales in summer is 0.80.

To correct for seasonality, the headache tablet sales figures for summer should be

A. reduced by 80%

B. reduced by 25%

C. reduced by 20%

D. increased by 20%

E. increased by 25%

Ex 5d

Fitting a trend line and forecasting – Ex 5e

Fitting a trend line

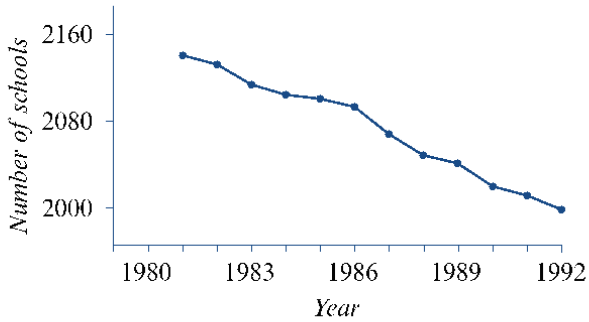
Fit a trend line to the data in the following table, which shows the number of government schools in Victoria over the period 1981–92, and interpret the slope.

Year	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Number	2149	2140	2124	2118	2118	2114	2091	2064	2059	2038	2029	2013

Solution

1 Construct a time series plot of the data to confirm that the trend is linear.

Note: For convenience we let 1981 = 1, 1982 = 2 and so on when entering the data into a calculator.



2 Fit a least squares line to the data with *year* as the EV. Write down its equation.

Number of schools = 2169 − 12.5 × *year*

3 Write down the slope and interpret.

Slope = −12.5

Over the period 1981–92 the number of schools in Victoria decreased at an average rate of 12.5 schools per year.

Forecasting

- Using a trend line fitted to a time series plot to make predictions about the future is known as trend line forecasting

Forecasting with seasonal time series

A regression line can be used to make predictions and forecast using the deseasonallised data.

However the new value will be deseasonalised or smoothed value. To get back to an original value we use the following rule:

Seasonalised figure or value = deseasonalised figure or value x seasonal index

Example

The *deseasonalised* quarterly sales data from Mikki's shop are shown below.

Quarter	1	2	3	4	5	6	7	8	9	10	11	12
Sales	893	943	955	858	1005	1026	1043	1040	1261	1151	1115	1267

Fit a trend line and interpret the slope.

Making predictions with deseasonalised data

> What sales do we predict for Mikki's shop in the winter of year 4? (Because many items have to be ordered well in advance, retailers often need to make such decisions.)

Seasonal indicies

As stated before, the sum of the seasonal indicies is equal to the number of seasons

Example

A fast food store that is open seven days a week has the following seasonal indices.

Season	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Index	0.5	0.2	0.5	0.6		2.2	1.1

The index for Friday has not been recorded. Calculate the missing index.

Exam Questions

2010

Question 13

A garden supplies outlet sells water tanks. The monthly seasonal indices for the revenue from the sale of water tanks are given below.

The seasonal index for September is missing.

Seasonal index											
Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1.26	0.96	0.74	0.48	0.31	0.39	0.75	1.55		1.32	1.25	1.58

The revenue from the sale of water tanks in September 2009 was $104 500.

The deseasonalised revenue for September 2009 is closest to

A. $42 800

B. $74 100

C. $104 500

D. $141 000

E. $147 300

2012

Use the following information to answer Questions 11 and 12.

The table below shows the long-term average rainfall (in mm) for summer, autumn, winter and spring. A shown are the seasonal indices for summer and autumn. The seasonal indices for winter and spring are missing.

	Season			
	Summer	**Autumn**	**Winter**	**Spring**
Long-term average rainfall (mm)	52.0	54.5	48.8	61.3
Seasonal index	0.96	1.01		

Question 11

The seasonal index for spring is closest to

A. 0.90

B. 1.03

C. 1.13

D. 1.15

E. 1.17

Question 12

In 2011, the rainfall in autumn was 48.9 mm.

The deseasonalised rainfall (in mm) for autumn is closest to

A. 48.4

B. 48.9

C. 49.4

D. 50.9

E. 54.0

2014

Use the following information to answer Questions 10 and 11.

The seasonal indices for the first 11 months of the year, for sales in a sporting equipment store, are shown in the table below.

Month	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Seasonal index	1.23	0.96	1.12	1.08	0.89	0.98	0.86	0.76	0.76	0.95	1.12	

Question 10

The seasonal index for December is

A. 0.89
B. 0.97
C. 1.02
D. 1.23
E. 1.29

Question 11

In May, the store sold $213 956 worth of sporting equipment.

The deseasonalised value of these sales was closest to

A. $165 857
B. $190 420
C. $209 677
D. $218 322
E. $240 400

Question 12

The seasonal index for heaters in winter is 1.25.

To correct for seasonality, the actual heater sales in winter should be

A. reduced by 20%.
B. increased by 20%.
C. reduced by 25%.
D. increased by 25%.
E. reduced by 75%.

https://www.youtube.com/watch?v=7Ug2zA0M-1Y

Ex 5E