Time Series Summary

We often display cost, sales, rainfall etc. against time. Time is displayed as the independent variable along the x-axis and the other variable along the y-axis. This is called a time series. We normally join up the points in a time series.

A time series can have one or more of the following components:

- Trend (positive or negative secular trend)
- Seasonal Pattern
- Cyclic Pattern
- Random variation

Trend

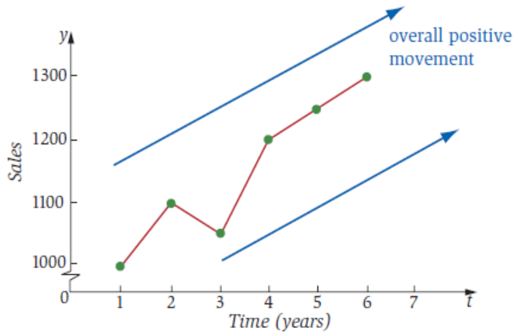
A trend exists if there is a long term increase (positive) or decrease (negative) in the dependent variable as time passes.

Positive Upward Trend

For example, the sales of a certain brand of bicycle steadily increase over the years. The time series plot below shows that the trend of the time series data over the period 1995 to 2000 is moving in an overall positive direction. We say the time series plot has a **positive secular trend**.

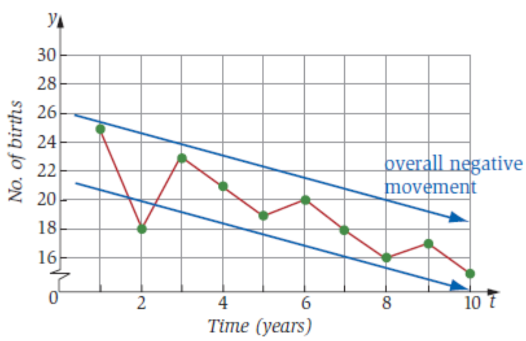
tip

For convenience, we set the scale of time as follows: for 1995, $t = 1$, for 1996, $t = 2$, and so on.



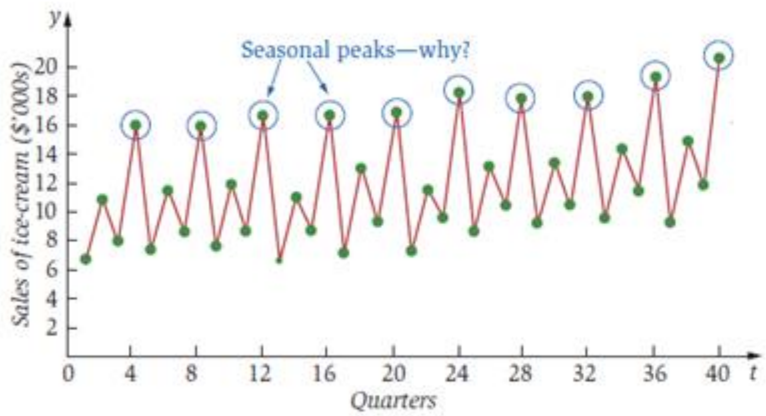
Negative Downward Trend

A time series plot can show an overall negative movement. This means that the time series has a negative secular trend, or downward trend. For example the number of births in a remote country hospital has decreased steadily over the years from 1996 to 2005.



Seasonal Trend

When the seasons of the year affect sales or production, peaks and troughs will appear at **regular intervals** during the year. For example, seasonal rainfall during summer, autumn, winter and spring in a year. The name seasonal is not specific to seasons of the year. It could be related to weekly sales in which sales on Saturday are consistently higher than the other week days. A key feature of seasonal trends is that the seasons occur at the same time each cycle.

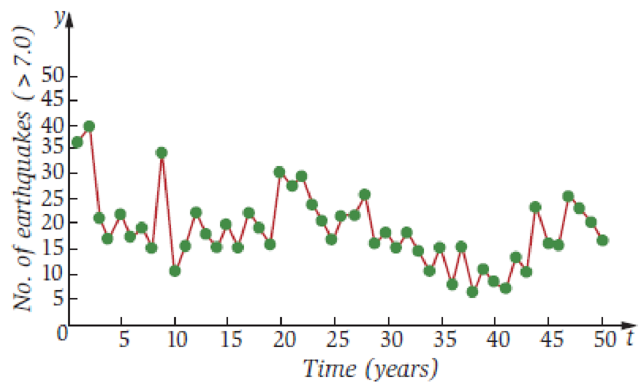


Notice that the graph peaks to times corresponding to t = 4, 8, 12 etc. which are the summer quarters of each year over 10 years.

Cyclic Trends

Like seasonal trends, cyclic trends show fluctuations upwards and downwards but not according to season. The peaks and troughs occur on an **irregular** basis.

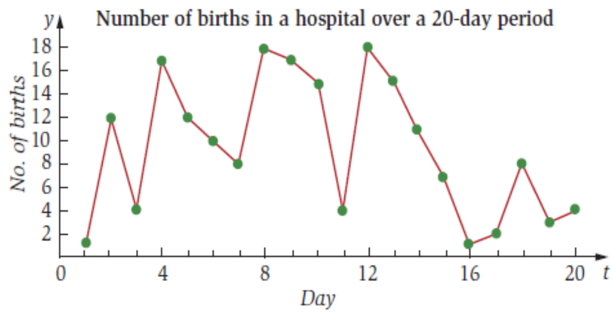
For example the number of large earthquakes recorded each year show significant peaks and troughs, but at unpredictable intervals.



Random Trends

Random variation or random pattern will not show predictable peaks and troughs nor will there be any significant peaks or troughs at unpredictable times. Instead there is a random movement about a relatively stable mean.

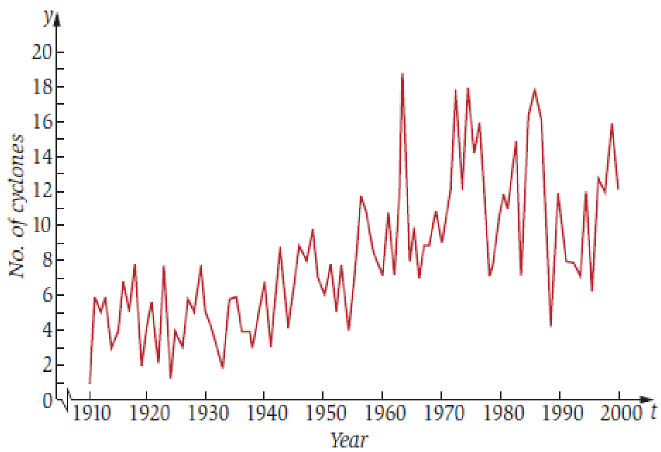
An example of a random pattern is the number of daily births in a particular hospital over a period of 20 days. A time series plot is shown on the right.



Sometimes it is difficult to decide whether a trend is cyclic or random. Choose random as a last resort if the trend is not seasonal or cyclic.

Example 1

For the time series plot shown, describe the components present.



Solution

This time series plot shows a positive secular trend with a cyclic pattern.

Note that although random variation is always a component, it is only mentioned when none of the other 3 components is present.

Fitting Trend Lines to Time Series Plots

There are 3 possible ways to fit a trend line to a time series plot:

- By eye
- Three median regression method
- Least squares regression method

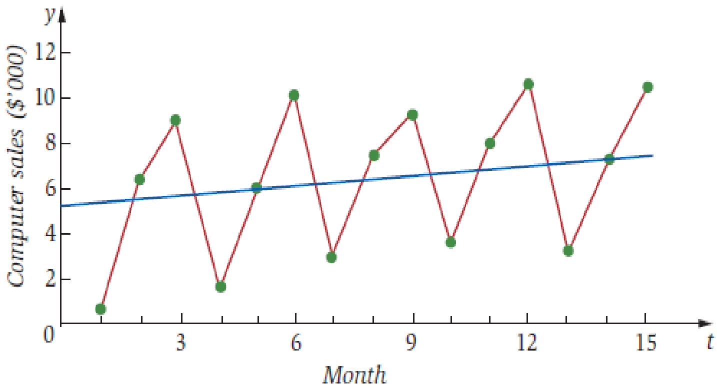
You can think of a time series plot as similar to a scatter plot with independent variable time along the x axis. Use these techniques on the original data when the trend is clearly linear. The methods cannot be applied effectively to cyclical or seasonal trends.

By Eye

Fitting a trend line by eye will only give approximate results when used to make predictions.

Solution

To fit a trend line by eye, try to place a line as close to all the points as possible. In the graph on the right, a reasonable line is shown. It passes through the data pairs (5, 6) and (14, 7).



The equation of the line can be found using the 2 points to be

y = 0.11t + 5.44

Note that making predictions using this equation would be very unreliable.

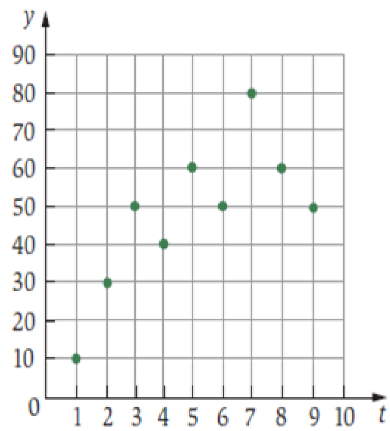
3-Median Regression Method

This method has been met before and CAS can be used to determine the equation of the line using the Median-Median option.

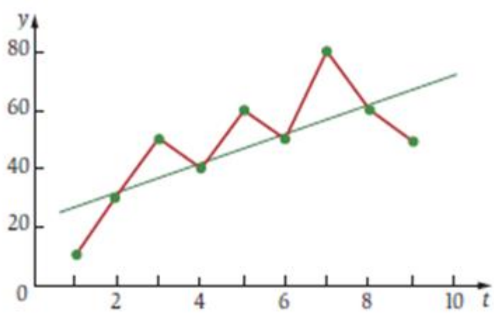
Example:

a Using the three median method, fit a trend line to the data shown.

b Find its equation and use the equation to find the value for y when t is 4. How accurate is this prediction?



a



b.

Equation of 3-median regression line is

$y = 5t + 21.7$

The true value of y at t = 4 is 40, compared to the predicted y = 41.7.

This is a fairly accurate prediction.

Least Squares Regression Method

This method has been met before and CAS can be used to determine the equation of the line using $y = mx + b$.

Smoothing Time Series

Time series data can be prone to large fluctuations from point to point. This means that at times a trend line cannot accurately predict the future if there is a large variation in how data moves. We can smooth out the fluctuations to show a clearer picture of the overall trend. We can use the following 3 techniques:

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 data points are averaged, then we move on to another group of data points in a systematic fashion and average them, and so on. Note that when finding the moving average we are finding the mean of the data points. There are two cases to consider.

Moving Average Smoothing with an Odd Number of Points.

The following example shows how we would use a 3-point moving average to smooth out the data points.

For example, the number of babies born in a remote hospital over the period 1996 to 2005 is given in the following table.

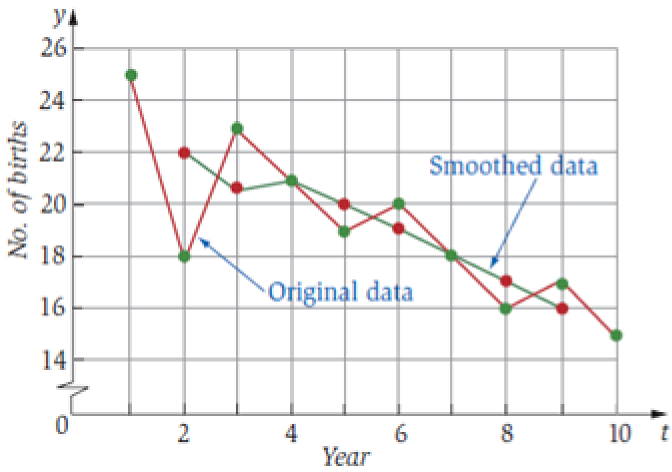
Year t	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Number of births y	25	18	23	21	19	20	18	16	17	15

We will smooth this data to determine the longterm trend in the number of births in this particular hospital. In this example, we use a three point moving average. That is, we use three successive data points to create a smoothed data value. Moving through the data by always using three points at a time means we end up with a new set of smoothed data.

Year t	No. of births y	Three point moving average
1	25	
2	18	$\frac{25+18+23}{3} = 22$
3	23	$\frac{18+23+21}{3} = 20.67$
4	21	$\frac{23+21+19}{3} = 21$
5	19	$\frac{21+19+20}{3} = 20$

Year t	No. of births y	Three point moving average
6	20	$\frac{19+20+18}{3} = 19$
7	18	$\frac{20+18+16}{3} = 18$
8	16	$\frac{18+16+17}{3} = 17$
9	17	$\frac{16+17+15}{3} = 16$
10	15	

By drawing a time series plot with the original data and the 3-point moving average data the general trend becomes more obvious.



The least squares equation for the **3-point moving average data** can be calculated using CAS and predictions made from it. This will give more accurate forecasts than if we used the original time series data.

Moving Average Smoothing with an Even Number of Points

Suppose a 4-point moving average is used in the previous example. We have to find the moving average in 2 steps to assist in locating the data point. The second step is called centering. Centering allows us to line up the moving average with a specific year. The number of babies born in a remote hospital over the period 1996 to 2005 is given by:

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Year t	1	2	3	4	5	6	7	8	9	10
No of births	25	18	23	21	19	20	18	16	17	15

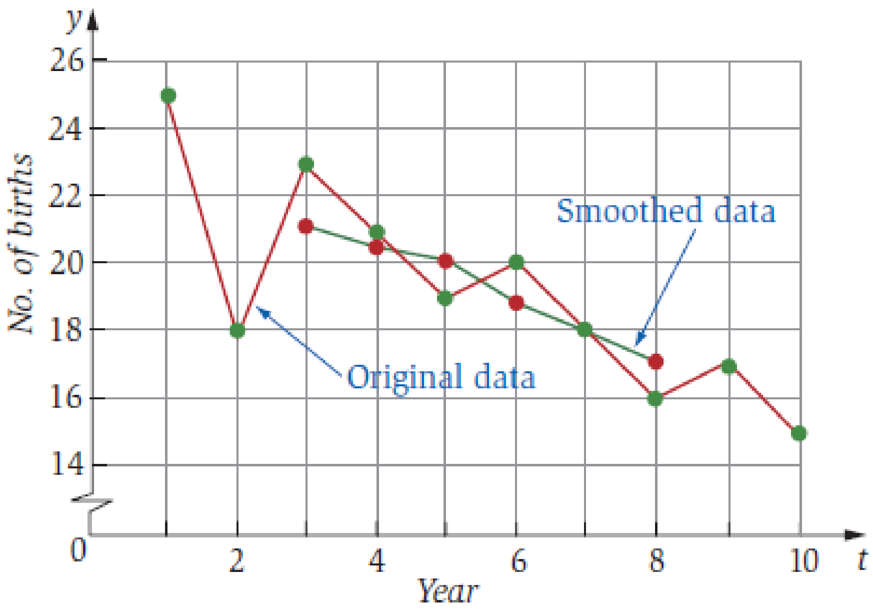
To calculate the 4 point moving average we form the table below:

Column1 Year	Column2 Year	No of Births	4-point Moving Average	Centred Moving Average
1996	1	25		
1997	2	18		
			21.75	
1998	3	23		21.000
			20.25	
1999	4	21		20.500
			20.75	
2000	5	19		20.125
			19.50	
2001	6	20		18.875
			18.25	
2002	7	18		18.000
			17.75	
2003	8	16		17.125
			16.50	
2004	9	17		
2005	10	15		

In Summary:

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Year t	1	2	3	4	5	6	7	8	9	10
Number of Births	25	18	23	21	19	20	18	16	17	15
Centred moving average			21.000	20.500	20.125	18.875	18.000	17.125		

Plotting the Number of Births and the Centred Moving Average on the same grid gives the graph below:



The least squares regression equation of the **centered moving average** points can be calculated using CAS to make predictions.

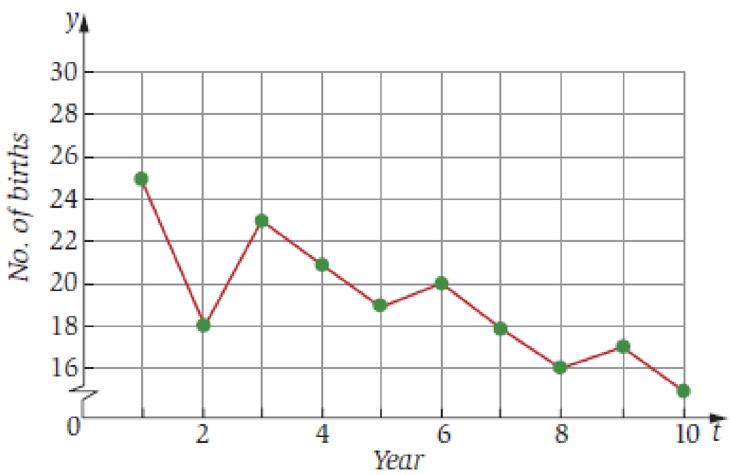
Median Smoothing (or Moving Medians)

Median smoothing uses a similar technique to moving averages; it uses the median values instead of the average values.

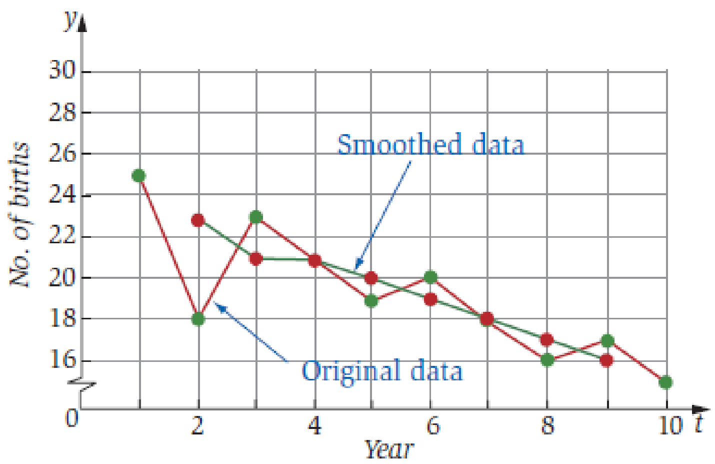
Median Smoothing with an Odd Number of Points

When we smooth with an odd number of points we can often do it by eye.

For example the time series for the data describing the number of births in a country hospital are shown in the graph below:



Using a **3-point median method** we can form the plot below:



Median Smoothing with an Even Number of Points

When we smooth with an even number of points an extra step called **centering** is needed to line up the moving median values with a specific year.

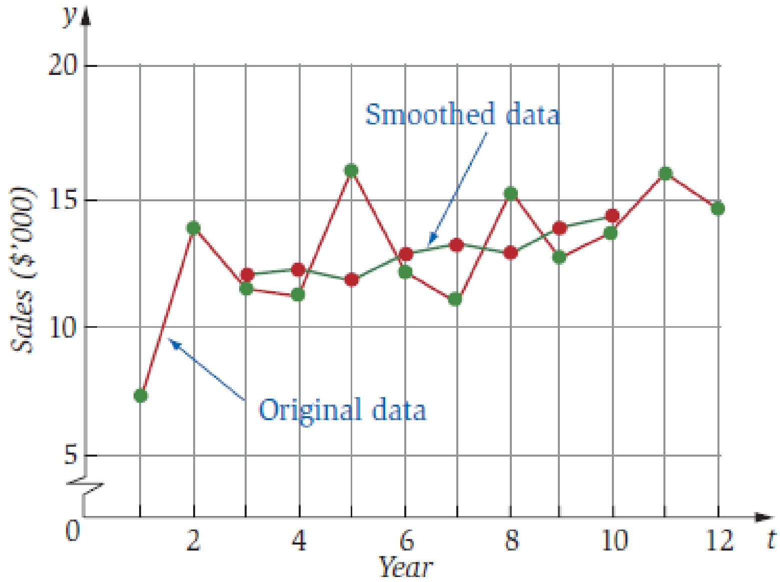
Example: The sales figures for a kiosk between 1993 and 2004 are given below.

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Sales ($'000)	7.3	13.8	11.4	11.2	16	12.2	10.9	15.3	12.7	13.7	16	14.5

Since we are using a four point median smoothing we must centre the smoothed data.

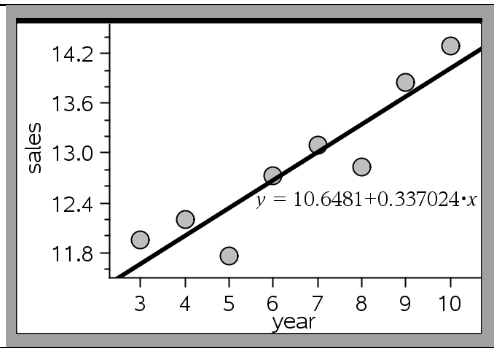
Year	t	Sales ($'000)	4-point moving median		
			Calculation	Smoothed value	Centred
1993	1	7.3			
1994	2	13.8			
			Median of (7.3, 13.8, 11.4, 11.2)	11.3	
1995	3	11.4			11.95
			Median of (13.8, 11.4, 11.2, 16)	12.6	
1996	4	11.2			12.2
			Median of (11.4, 11.2, 16, 12.2)	11.8	
1997	5	16			11.75
			Median of (11.2, 16, 12.2, 10.9)	11.7	
1998	6	12.2			12.73
			Median of (16, 12.2, 10.9, 15.3)	13.75	
1999	7	10.9			13.1
			Median of (12.2, 10.9, 15.3, 12.7)	12.45	
2000	8	15.3			12.83
			Median of (10.9, 15.3, 12.7, 13.7)	13.2	
2001	9	12.7			13.85
			Median of (15.3, 12.7, 13.7, 16)	14.5	
2002	10	13.7			14.3
			Median of (12.7, 13.7, 16, 14.5)	14.1	
2003	11	16			
2004	12	14.5			

The data can be plotted on a graph.



Using the **smoothed data** points CAS can be used to determine the equation of the least squares regression line. Note that we need to start from year 3 and finish at year 10.

	A year	B sales	C	D
1	3	11.95		
2	4	12.2		
3	5	11.75		
4	6	12.73		
5	7	13.1		
6	8	12.83		



From the graph the equation of best fit is: $y = 0.337024x + 10.6481$
$Sales = 0.337024 \times time + 10.6481$

We can predict the sales for year 2010 by substituting t = 18

Sales = 0.337024 x 18+ 10.6481 = 16.7145

This is approximately $16700

Seasonal Adjustments or Deseasonalisation

Seasonal factors are variations due to weather (seasons), the day of the week, the month of the year, the quarter of the year.
For example:

- Ice cream sales are greater in the summer than in the winter
- Sales of winter clothing are greater in the winter than the summer
- Sales of soft drinks vary with temperature
- Retail sales increase during Xmas period

Smoothing data of this type is more complex and smoothing the data using moving averages or moving medians may not be as effective.
To improve the smoothing process by taking out the seasonal effects we use a process called **deseasonalization** or **seasonal adjustment** so that a trend line can be fitted and long term trends can be predicted. This process involves finding the **seasonal indices** for each quarter, month or week. The steps to calculate the seasonal indices and the seasonally adjusted data is best shown in an example.

Example:
The quarterly sales figures (number of houses sold) were recorded by an estate agent for each of the years from 2003 to 2005.

	Quarterly sales figures			
Year	1	2	3	4
2003	5	7	9	3
2004	4	8	9	4
2005	5	9	10	5

Determine the Seasonal Index for each Quarter.

Step 1 Calculate the yearly average
Find the totals for each year and hence find the yearly averages.

Yearly averages 2003–2005

Year	Q_1	Q_2	Q_3	Q_4	Yearly average
2003	5	7	9	3	$\frac{5+7+9+3}{4} = 6$
2004	4	8	9	4	$\frac{4+8+9+4}{4} = 6.25$
2005	5	9	10	5	$\frac{5+9+10+5}{4} = 7.25$

Step 2 Calculate the quarterly proportions
Divide each sales figure by the corresponding yearly average.

Quarterly proportions 2003–2005

Year	Q_1	Q_2	Q_3	Q_4
2003	$\frac{5}{6} = 0.8333$	$\frac{7}{6} = 1.1667$	$\frac{9}{6} = 1.5000$	$\frac{3}{6} = 0.5000$
2004	$\frac{4}{6.25} = 0.6400$	$\frac{8}{6.25} = 1.2800$	$\frac{9}{6.25} = 1.4400$	$\frac{4}{6.25} = 0.6400$
2005	$\frac{5}{7.25} = 0.6897$	$\frac{9}{7.25} = 1.2414$	$\frac{10}{7.25} = 1.3793$	$\frac{5}{7.25} = 0.6897$

Step 3 Calculate the seasonal indices (SI)
Seasonal indices are found by averaging the quarterly proportions.

Year	Q_1	Q_2	Q_3	Q_4
2003	0.8333	1.1667	1.5000	0.5000
2004	0.6400	1.2800	1.4400	0.6400
2005	0.6897	1.2414	1.3793	0.6897
Total	2.1630	3.6881	4.3193	1.8297
SI	$\frac{2.1630}{3} = 0.7210$	$\frac{3.6881}{3} = 1.2294$	$\frac{4.3193}{3} = 1.4398$	$\frac{1.8297}{3} = 0.6099$

Seasonal indices

Quarter	Q_1	Q_2	Q_3	Q_4
Seasonal index	0.7210	1.2294	1.4398	0.6099

The seasonal indices should sum to 4. This is a useful check!
(Note: In many problems you are given the seasonal indices so you do not have to work them out from first principles.)

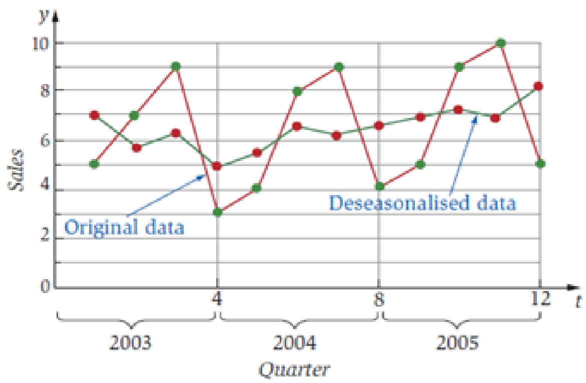
Deseasonalise the Original Time Series Data

Step 4 Deseasonalise the original time series data
Divide each sales figure by the corresponding seasonal index.

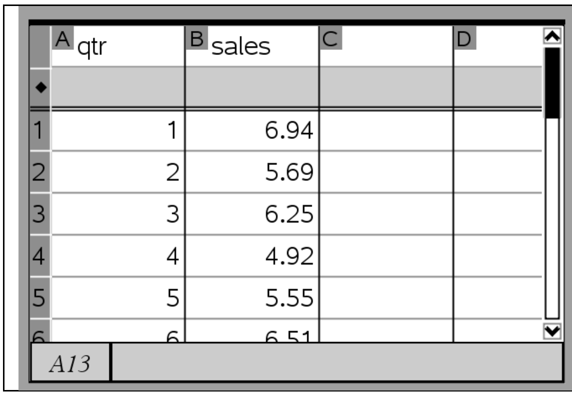
Deseasonalised figures 2003–2005

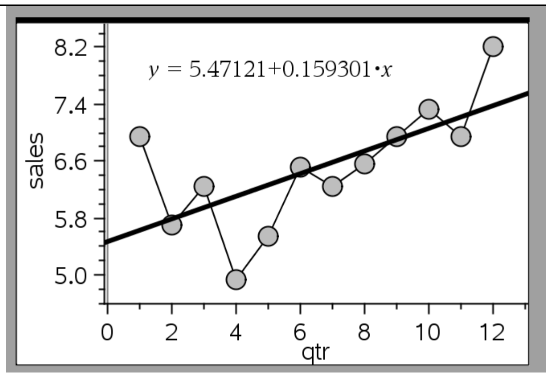
Year	Q_1	Q_2	Q_3	Q_4
2003	$\frac{5}{0.7210} = 6.94$	$\frac{7}{1.2294} = 5.69$	$\frac{9}{1.4398} = 6.25$	$\frac{3}{0.6099} = 4.92$
2004	$\frac{4}{0.7210} = 5.55$	$\frac{8}{1.2294} = 6.51$	$\frac{9}{1.4398} = 6.25$	$\frac{4}{0.6099} = 6.56$
2005	$\frac{5}{0.7210} = 6.94$	$\frac{9}{1.2294} = 7.32$	$\frac{10}{1.4398} = 6.95$	$\frac{5}{0.6099} = 8.20$

Using $t = 1$ to represent the first quarter of 2003, and so on, the original and deseasonalised time series data are shown in the following plot.



Using the deseasonalized sales data, we can create a least squares regression line using CAS and predict the deseasonalized sales for the first quarter of 2006.





Using the equation $Deseasonalised\ Sales\ =\ 0.159301\ \times qtr\ +\ 5.47121$ we can predict the **deseasonalised** sales in the first quarter of 2006.

Substitute qtr = 13 gives:
$\text{Deseasonalised Sales}\ =\ 0.159301\ \times\ 13 +\ 5.47121\ =\ 7.54212$

To calculate the actual sales you must remember to **seasonalise** the data! To do this, remember that:

$Actual\ sales\ =\ seasonal\ index\ \times deseasonalised\ prediction.$

In this case the seasonal index will be for Qtr 1.

Actual Sales = $0.7210 \times\ 7.54212\ =\ 5.43787$ which approximates 5.

More about Seasonal Indexes

Seasonal indices can be expressed as percentages as shown.

Seasonal indices

Quarter	Q_1	Q_2	Q_3	Q_4	Total
Seasonal index	0.7210	1.2294	1.4398	0.6099	4.0001

Seasonal indices

Quarter	Q_1	Q_2	Q_3	Q_4	Total
Seasonal index	72.10%	122.94%	143.98%	60.99%	400.01

The larger the seasonal index the higher the performance of the corresponding quarter compared to the average quarterly value. Thus in the real estate example the seasonal index of 143.98% for the third quarter indicates a performance of 43.98% above average.

Example:

A distributor of televisions predicts its deseasonalised quarterly sales to be $150 000 for each quarter. Predict its actual quarterly sales given the following seasonal indices: summer 84%; autumn 128%; winter 101%; spring 87%.

Solution

Prediction of actual value = seasonal index × deseasonalised prediction
= seasonal index × 150 000

Quarter	Seasonal index (%)	Seasonal index	Actual value SI × 150 000
Summer	84	0.84	126 000
Autumn	128	1.28	192 000
Winter	101	1.01	151 500
Spring	87	0.87	130 500

The final column gives the predicted (actual) quarterly sales.

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.

THINK

1. The sum of the seasonal indices is equal to the number of seasons.
2. The missing index is the sum of all the other seasons subtracted from the total.

WRITE

There are 7 seasons (Monday to Sunday), therefore the sum of indices is 7.

Friday index
= 7 – (sum of all the other indices)
= 7 – (0.5 + 0.2 + 0.5 + 0.6 + 2.2 + 1.1)
= 7 – 5.1
= 1.9