RSC | Advancing the Chemical Sciences

Time Series Analysis

Some data, such as sales data, may incorporate a seasonal trend. The underlying theory for approaches to forecasting in these situations is summarised in this document.

Consider a data series as shown in Table 5

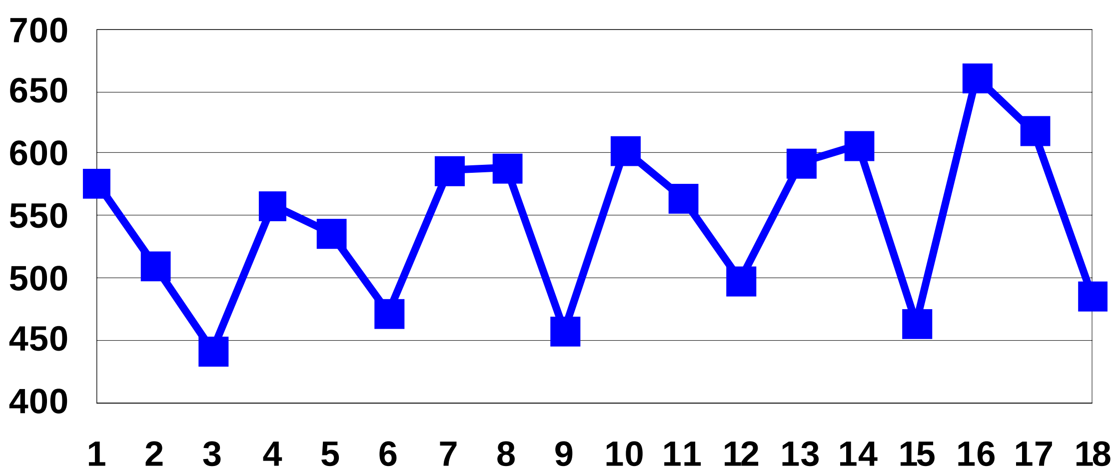
Table 5

Time period (t)	1	2	3	4	5	6	7
Sales (x)	577.2	510.2	442.8	559.0	536.5	471.0	585.9

Time period (t)	8	9	10	11	12	13	14
Sales (x)	587.8	457.1	602.5	564.7	498.9	593.5	607.2

Time period (t)	15	16	17	18
Sales (x)	463.8	661.6	617.7	485.7

Plotting the sales figures onto a graph shows a gradual increase but the variation within the data is significant. Indeed, careful observation shows that there is a distinct "seasonality" within the data. Every three time periods shows a similar peak or a trough. It would not be unreasonable to propose that to forecast future sales then the seasonality element should be included. A technique known as time series analysis may be used to forecast sales with seasonality.



The theory behind this approach is outlined below followed by a numerical calculation to show how it works in practice.

An additive model may be used; $y = x + s + r$

where ***y*** is the sales figure, ***x*** is the basic trend (secular) component, ***s*** is the seasonality (cyclical) component and ***r*** is the residual (random) component.

Observation of the data shows that there is a 3 season seasonality i.e. the peaks/troughs occur every 3 time periods

So, considering a sequential 3-season period: (note any number of periods can be used but an odd number is easier to centre and understand the theory)

For period n-1: $y_{n-1} = x_{n-1} + s_{n-1} + r_{n-1}$

For period n: $y_n = x_n + s_n + r_n$

For period n+1: $y_{n+1} = x_{n+1} + s_{n+1} + r_{n+1}$

Averaging these three sequential data points (a bar over the symbol indicates a mean) over the 3-season year and centering the average on period *n*

$$\bar{y}_n = \bar{x}_n + \bar{s}_n + \bar{r}_n$$

If seasonal variations are regular over time and the averaging has been done over a full season cycle then there will be as many data points above the average as there are below the average.

Hence $\bar{s}_n = 0$, in other words we have smoothed out the seasonality by taking an average. If the 3 sequential averages are taken at every data point along the data set then we will arrive at a set of data with the seasonality removed. This procedure is called centred moving averaging.

If residual component is random then an average of random values will cancel each other out so

$$\bar{r}_n = 0.$$

Hence as $\bar{y}_n$ is approximately x_n the centred moving average values may be used as a primary estimate of the x values hence the underlying trend equation.

The amount of seasonality (i.e. greater than or lower than the average amount) may be estimated by calculating the ratio of the actual values to the centred moving averages:

$$\frac{y}{\bar{y}} = \frac{x_n + s_n + r_n}{x_n} \qquad \text{So} \quad \frac{y}{\bar{y}} = 1 + \frac{s_n + r_n}{x_n}$$

These values will be close to a Seasonality Index given by the formula

$$I_n = \frac{x_n + s_n}{x_n} \qquad \text{So } I_n = 1 + \frac{s_n}{x_n}$$

Assuming seasonal variation remains the similar over time then:

$$I_n = I_{n+3} = I_{n+6} = I_{n+9} \quad \text{etc}$$

so the average of the estimates for a given season will provide the mean seasonal index for the season.

So, having obtained estimates for x_n and I_n successive seasonal variations can be estimated

$$s_n = (I_n \times x_n) - x_n \qquad \text{So } s_n = (I_n - 1)x_n$$

and the residuals from

$$r_n = y_n - x_n - s_n \qquad \text{So } r_n = y_n - (I_n \times x_n)$$

The difference between fitted data and actual data may be used to estimate the confidence limits of the extrapolation

The data from Table 5 have been re-drawn into Table 6.

Table 6

Time Period (t)	Sales (y) (=x+s+r)	Centred moving average (CMA)	y / CMA (=I_n)	Average seasonality index I_n	x estimated from graph	s (est)	r
1	577.2			1.09	504	46	27
2	510.2	510.1	1.000	1.05	509	25	-24
3	442.8	504.0	0.879	0.86	513	-71	1
4	559.0	512.8	1.090	1.09	518	47	-7
5	536.5	522.2	1.027	1.05	523	26	-13
6	471.0	531.1	0.887	0.86	528	-73	17
7	585.9	548.2	1.069	1.09	533	48	5
8	587.8	543.6	1.081	1.05	538	27	23
9	457.1	549.1	0.832	0.86	543	-76	-10
10	602.5	541.4	1.113	1.09	548	50	4
11	564.7	555.4	1.017	1.05	553	27	-16
12	498.9	552.4	0.903	0.86	559	-78	18

13	593.5	566.5	1.048	1.09	564	51	-21
14	607.2	554.8	1.094	1.05	569	28	10
15	463.8	577.5	0.803	0.86	575	-80	-31
16	661.6	581.0	1.139	1.09	580	53	29
17	617.7	588.3	1.050	1.05	585	29	3
18	485.7			0.86	591	-82	-23
			standard deviation of residuals =				18.6

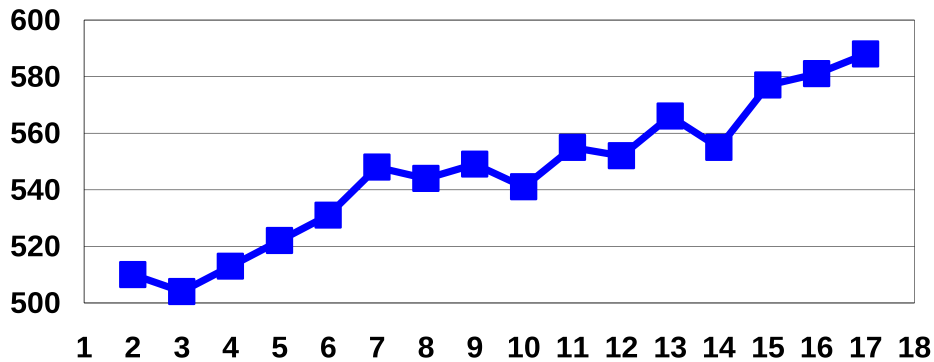
Procedure for Centred Moving Averaging

1. Construct a table of sequential sales data (y). Determine from the seasonality how many time periods represent a full cycle of seasonality. For this technique to work, the number of periods used to take centred moving averages must cover the full seasonality (i.e. peak to peak or trough to trough), no more and no less. This is because when the averages are calculated the high seasons must cancel out the low seasons. In this example, the periodicity is 3 time periods so the centred moving average values must be calculated from 3 sequential data points.

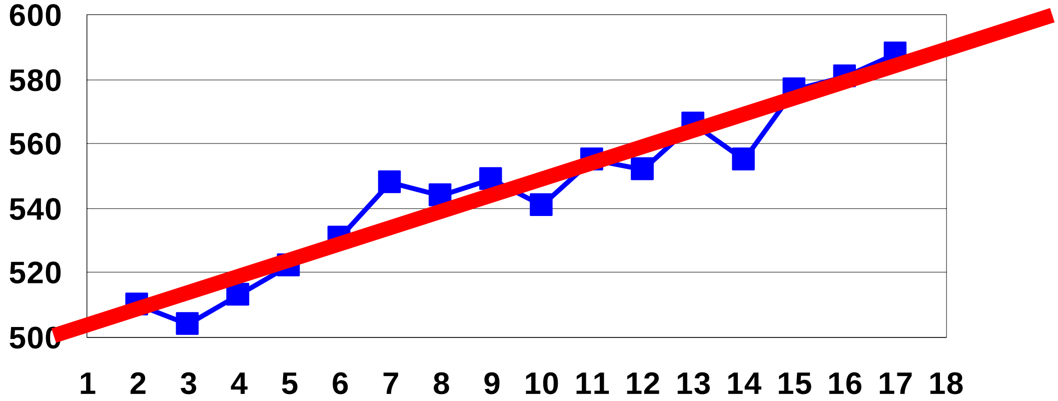
2. Calculate the average of 3 sequential ***y*** values and put the answer in the centre row in the ***CMA*** column of Table 6. For example, the first three points are 577.2, 510.2 and 442.8, the average is 510.1 and this value goes into the row t=2.

3. Divide the ***y*** values by the ***CMA*** values and put the answers into the ***y***/***CMA*** column, of Table 6 which represents the seasonality index. Take the average of all the values representing one particular period e.g. rows 4, 7, 10, 13 and 16 and put this average value (1.09) into the average seasonality column for every time period when this season appears i.e. rows 4, 7, 10,13 and 16. Do the same for the other 2 seasons (their averages are 1.05 and 0.86).

4. In order to determine the trend line (***x***) plot a graph of ***CMA*** against ***t*** as shown below.



This shows the upward trend without the complication of the seasonality. Decide what trend equation this trend is best represented by (linear, binomial, exponential, logistic) then determine the best equation to represent the trend. Assuming a linear trend is easier and the best straight line is drawn on the graph below.



The best straight line is represented by the linear equation: $x = 496.6 + 5.115t$

Now, either calculate the values of x from the equation or read off their values from the straight line graph and put these values into the (x estimated from the graph) column of table 6.

5. The values of ***s*** and ***r*** in Table 6 may be determined using the equations stated above.

6. In order to forecast the sales values for periods t = 19, 20 and 21, first determine the underlying trend value (***x***) either by calculation using the trend equation using t = 19, 20 and 21 or by extrapolation and reading off the values from the graph.

Time period (***t***)	Extrapolated Value of ***x*** from equation or graph	Average Seasonality Index (I_n) for the period	Forecasted value (= $x \times I_n$)	+/- standard deviation (***r***)
19	593.8	1.09	**647.2**	18.6
20	598.9	1.05	**628.8**	18.6
21	604.0	0.86	**519.4**	18.6

Multiply the x value by its respective seasonality index (I_n) to introduce the seasonality element and this represents the forecasted value +/- the standard deviation determined from the residuals ***r***.