Activity Solution: Smoothing for Seasonals

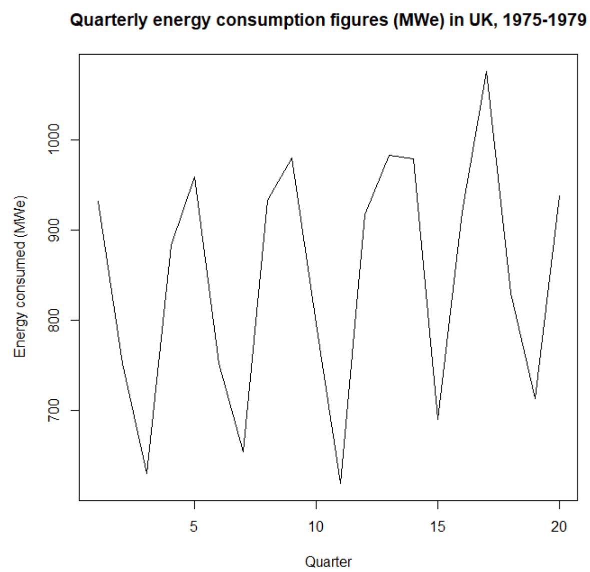
The following data are the quarterly energy consumption figures (in MWe) in the UK for the years 1975–1979.

Year	Quarter	t	$x(t)$	Sm(t)	Cm(t)	$x(t)$ –Cm(t)
1975	1	1	932			
1975	2	2	752			
1975	3	3	630		802.625	–172.625
1975	4	4	883		806.000	77.000
1976	1	5	959		809.000	150.000
1976	2	6	752		818.250	–66.250
1976	3	7	654		827.125	–173.125
1976	4	8	933		835.250	97.750
1977	1	9	980		836.375	143.625
1977	2	10	796		830.000	–34.000
1977	3	11	619		828.375	–209.375
1977	4	12	917		851.625	65.375
1978	1	13	983		883.375	99.625
1978	2	14	979		892.625	86.375
1978	3	15	690		904.625	–214.625
1978	4	16	920		897.625	22.375
1979	1	17	1076		⋆	⋆
1979	2	18	830		⋆	⋆
1979	3	19	713			
1979	4	20	938			

1. Sketch a rough plot of the time series, or use computer software if you have a laptop available. Comment on what you observe.
   Don't forget to label axes! In R we can create an object of "time series" type as follows:

```
> energyUK <- c(932,752,630,883,959,752,654,933,980,796,619,917,983,979,690,9
> energyUK <- as.ts(energyUK, start=c(1975,1), end=c(1979,4),
frequency=4)
> plot.ts(energyUK, main='Quarterly energy consumption figures
```

```
(MWe) in UK, 1975-1979', ylab='Energy consumed (MWe)', xlab='Quarter')
```



The series appears to have a slight upward trend and a seasonal effect of period 4.

2. *Clickers question:* Assuming the series contains an additive seasonal effect, how might we estimate the terms $S(1), \ldots, S(p)$? One approach involves *smoothing* the series $x(t)$ by averaging each p consecutive values to create a new series $\mathrm{Sm}(t)$. Find the first four values of the smoothed series. Each member of your group should compute (at least) one missing value. For what value of t would the first value of $\mathrm{Sm}(t)$ be most sensibly located?
   The first value would most sensibly be located at time point 2.5. This does not concur with the time of any observed value.

3. *Clickers question:* Find the numbers indicated by "$\star$" in the above table.
   To find $Cm(18)$ first, note that

$$\begin{aligned}\frac{938+713+830+1076}{4} &= 889.25,\\ \frac{713+830+1076+920}{4} &= 884.75\end{aligned}$$

and so averaging these

$$Cm\,(18) = \frac{884.75 + 889.25}{2} = 887.00.$$

Since

$$\frac{830 + 1076 + 920 + 690}{4} = 879.0$$

we see that

$$Cm\,(17) = \frac{879.0 + 884.75}{2} = 881.88.$$

Hence

$$\begin{aligned} x\,(17) - Cm(17) &= 194.12, \\ x\,(18) - Cm\,(18) &= -57.00. \end{aligned}$$

4. We can estimate the seasonal effects by looking at the mean differences between Cm(t) and the original series $x\,(t)$ for each quarter where the two series have values. The following table illustrates the calculation. Find the numbers indicated by "$\star$" in the table.

	Q1	Q2	Q3	Q4
	–	–	-172.625	77.000
	150.000	-66.250	-173.125	97.750
	$\star$	-34.000	$\star$	65.375
	99.625	86.375	-214.625	22.375
	$\star$	-57.000	–	–
Mean:	146.84	-17.72	$\star$	65.63

Using the data and calculating we find:

	Q1	Q2	Q3	Q4
	–	–	-172.625	77.00
	150.000	-66.250	-173.125	97.75
	143.625	-34.000	-209.375	65.375
	99.625	86.375	-214.625	22.375
	194.12	-57.00	–	–
Mean:	146.84	-17.72	-192.44	65.63

5. The above averages should give sensible estimates of the seasonal effects. But do they sum to zero? Find their sum and average. Adjust your seasonal estimates by subtracting this average from each one. Check your new seasonal estimates add to zero.
   Now since $146.84 - 17.72 - 192.44 + 65.63 = 2.31$, *subtracting 0.5775 from each term we find the adjusted seasonal indices as*

$$\begin{aligned} S(1) &= 146.26 \\ S(2) &= -18.30 \\ S(3) &= -193.02 \\ S(4) &= 65.05 \end{aligned}$$

6. Why would the method you applied in part 4 be preferable here to the method first applied to the "death rate" data that does not use smoothing?
   As the series has an apparent trend, the simpler method used on the death rate data would ignore the fact that values in each quarter are tending to be higher at the end than the start.

7. *Clickers question:* We could now deduct our estimates of the seasonal effects from the data and fit a linear model to the de-seasonalized series by least squares. (It may have been preferable to estimate the trend and then de-trend the series *before* estimating the seasonals.) When the smoothed data are regressed against t, the fitted line is

$$T(t) = 776.18 + 6.98t.$$

   Using this, forecast the energy consumption for the first two quarters of year 1980.

t	$T(t)$	$S(t)$	$\hat{x}(t)$
21	922.76	146.26	1069.02
22	929.74	−18.30	911.44

8. Re-cap what you have done during this activity. Why do you think you were asked to do this activity? What have you learned?
   The activity describes models for time series data that have a trend and/or a seasonal effect. An approach to estimating seasonal effects has been performed which involves smoothing the series over complete

periods, centering and subtracting the new series from the original. In this way average differences can be found between the original series and the smoothed version, the latter being deseasonalized. The estimates of the seasonal effects (sometimes called "seasonal indices", or "seasonals") can be adjusted to sum to zero, as would be expected for an additive seasonal model. A trend can be fitted and a model of the form

$$X(t) = \beta_0 + \beta_1 t + S(t) + \varepsilon$$

fitted. This can be used for forecasting, and would be preferable to using a regression approach alone when a seasonal effect is present.
The learning outcomes encountered in this activity include:

(a) *Apply a filter (that is, a smoother) to a time series, centring if necessary.*

(b) *Use a filter to estimate the seasonal indices in a time series that has an additive seasonal component.*