Risto B. Rushford
GSCM 451-Spr 2019
HW05

Data Analysis Concepts

Briefly comment on each of the following concepts, usually just several sentences.

1. List three different data structures from which a time series visualization can be created with R.

First, a time series visualization can be plotted without the dates as a run chart;

Second, it can be plotted by manually adding a dates column;

Finally, it can be plotted by directly using a time series (ts) object.

2. A forecast for a single time value includes the forecasted value and at least one _____. Why?

At least one corresponding index value, usually corresponding to a unit of time.

3. What is the distinction between additive and multiplicative seasonality? What is one easy and useful way to help assess if either is present in the data?

Additive seasonality is when the data have a consistent seasonality added to the trend. Multiplicative seasonality is when the seasonal variation increases over time, with larger differences between the peaks and troughs on a graph.

An easy way to assess if additive seasonality is present is to see if the difference between peaks and troughs in each cycle are similar. A way to tell if the graph exhibits multiplicative seasonality is to assess if the difference between the peak and the trough is growing in each subsequent cycle. Looking at it mathematically, the seasonal effects are a changing percentage of the trend line as opposed to added to it.

4. What is an “n-period running mean”?

An “n-period running mean” is the average value for the data, taking into account the data from n number of periods going back from the current time.

5. When deseasonalizing data, what is the purpose of computing the n-period running mean?

The n-period running mean is used to calculate the data’s underlying trend. The trend is best calculated after the data has been deseasonalized.

6. What is a “centered running mean”? Why and when is it needed?

The centered running mean (or centered moving average) calculates the mean of a centerpoint using two moving averages over a period of time equal to the number of points being considered (if quarterly, then using four points, if monthly, using 12).

7. For additive seasonality, given the data and running means, how are the deseasonalized data values computed?

For additive seasonality, the deseasonalized data values are computed by calculating the difference between the original data point for the time considered and subtracting the centered moving average.

8. For additive seasonality, what constraint must the seasonal indices satisfy as a group?

For an additive model, the seasonal indices must sum to zero. If they don’t, then the index must be adjusted in accordance with the number of points used per period.

Worked Problems

1. Trend and Seasonality

Consider the quarterly time series data that begins at the last quarter of 2012 of Sales at:

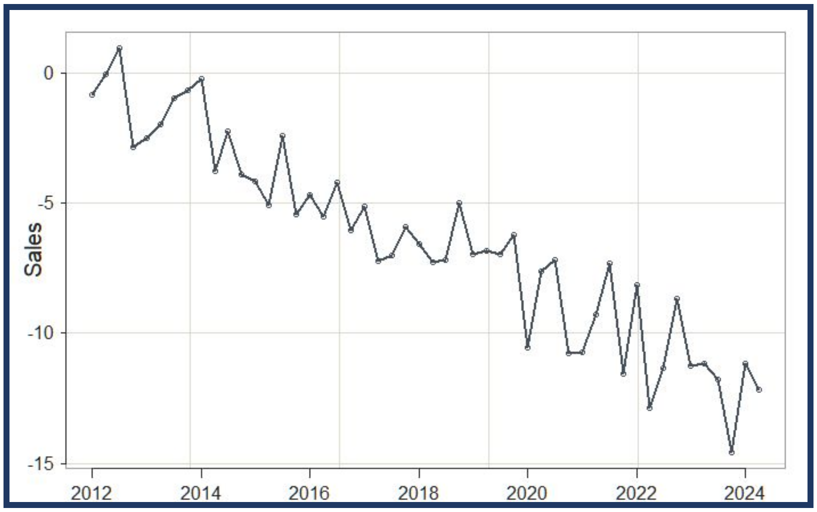
http://web.pdx.edu/~gerbing/451/Data/HW6_1.xlsx

a. Create an R ts object of the data and display its values.

```
d <- rd("http://web.pdx.edu/~gerbing/451/Data/HW6_1.xlsx")
d
     Y
     <dbl>
1    -0.8642
2    -0.0587
3    0.9501
4    -2.8683
5    -2.5074
6    -1.9898
7    -0.9886
8    -0.6619
9    -0.2329
10   -3.7591
...
```

b. Plot the time series:

```
Sales <- ts(d$Y, frequency=4, start=c(2012,1))
Plot(Sales)
```

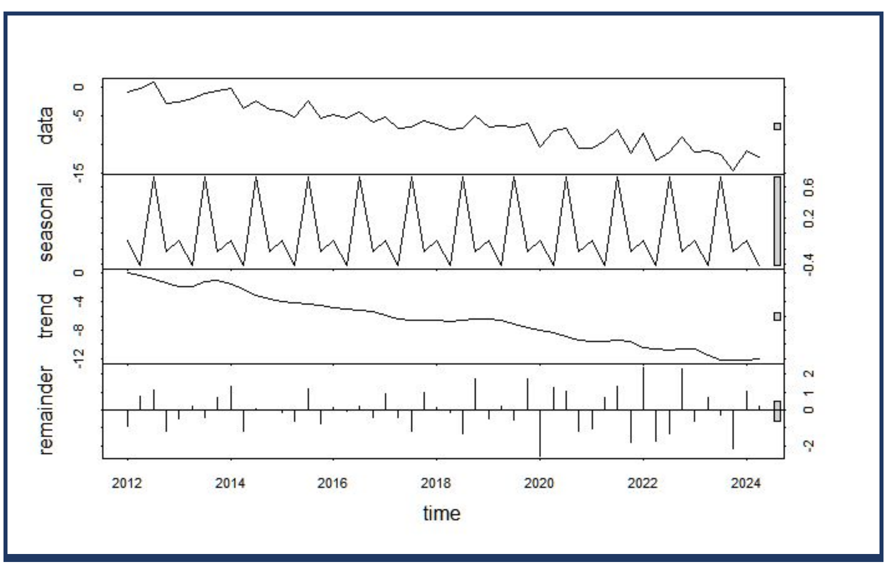


c. Characterize the time series qualitatively, that is, describe its general behavior verbally.

This time series displays seasonal behavior with an overall downward trend.

d. Deconstruct any trend and seasonality with a visualization (can plot in a single visualization).

```
decomp_Sales <- stl(Sales, s.window="periodic")
plot(decomp_Sales)
```



e. Specify and interpret the numerical seasonal components.

Detrended, the seasonal components have a mean of 0.2, with a maximum of 0.6 and a minimum of -0.4 (in $10k USD).

f. Express the second data value in terms of the trend, seasonal, and error component values:

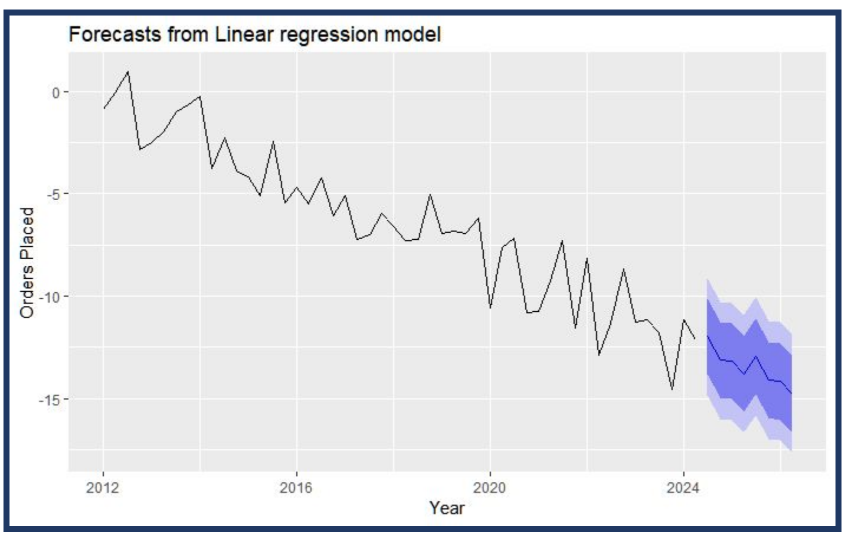
```
Components
           seasonal        trend    remainder
2012 Q2 -0.41058305  -0.41300973  0.764892783
```
<- The second data value

g. Generate the forecast of sales for the next two years numerically.

```
fit_Sales <- tslm(Sales ~ trend + season)
f <- forecast(fit_Sales, h=8)
f

      Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
2024 Q3     -11.96784 -13.81044 -10.12523 -14.82117  -9.11450
2024 Q4     -13.12779 -14.97039 -11.28519 -15.98113 -10.27446
2025 Q1     -13.19354 -15.03933 -11.34776 -16.05181 -10.33527
2025 Q2     -13.79683 -15.64262 -11.95105 -16.65510 -10.93857
2025 Q3     -12.94053 -14.80033 -11.08073 -15.82050 -10.06056
2025 Q4     -14.10049 -15.96029 -12.24069 -16.98046 -11.22052
2026 Q1     -14.16624 -16.03041 -12.30206 -17.05298 -11.27949
2026 Q2     -14.76953 -16.63371 -12.90535 -17.65628 -11.88278
```

h. Generate the visualization of the data with the forecast.



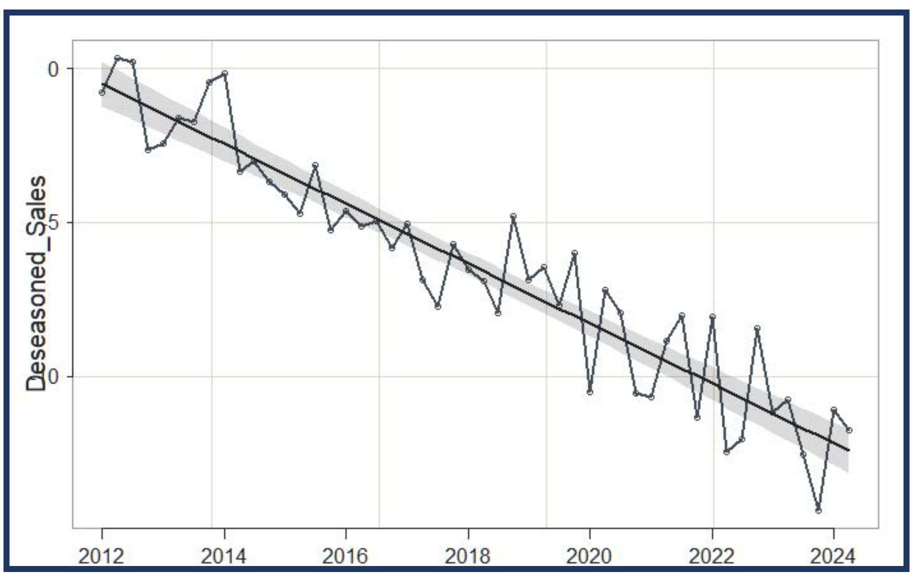
i. Interpret the forecast.

There is a steady, overall downward trend of sales from 2012 through the first quarter 2024. The model thus forecasts a continued decline. The dark purple region in the forecasted part of the graph represents the 80% confidence interval, and the lighter purple regions (encompassing the darker purple region as well) represents the 95% confidence interval.

j. Compute the deseasonalized data.

```
seasonality_Sales <- decomp_Sales$time.series[,1]
Deseasoned_Sales <- Sales - seasonality_Sales
Deseasoned_Sales
                 Qtr1        Qtr2        Qtr3        Qtr4
     2012  -0.7820311   0.3518831   0.2283775  -2.6393294
     2013  -2.4252311  -1.5792169  -1.7103225  -0.4329294
     2014  -0.1507311  -3.3485169  -2.9874225  -3.6592294
     2015  -4.0906311  -4.6850169  -3.1457225  -5.2227294
     2016  -4.6033311  -5.1037169  -4.9448225  -5.8250294
     2017  -5.0353311  -6.8372169  -7.7386225  -5.7032294
     2018  -6.5082311  -6.8857169  -7.9262225  -4.7863294
     2019  -6.8720311  -6.4283169  -7.6718225  -5.9810294
     2020 -10.4866311  -7.2039169  -7.9168225 -10.5453294
     2021 -10.6679311  -8.8582169  -8.0329225 -11.3412294
     2022  -8.0643311 -12.4489169 -12.0435225  -8.4321294
     2023 -11.1808311 -10.7369169 -12.5147225 -14.3472294
     2024 -11.0654311 -11.7423169
```

k. Plot the deseasonalized data.



I. Write the data frame with both the original data and the deseasonalized data to an Excel file. Open the file and display a screen pic or copy of the data in the Excel file.

```
d$Sales_Deseasoned <- as.vector(Deseasoned_Sales)
d$Qtr <- as.vector(time(Sales))
Write("Deseasonalized_Data", format="Excel")
```

Y	Sales_Deseasoned	Qtr
-0.8642	-0.7820311277	2012
-0.0587	0.3518830524	2012.25
0.9501	0.2283774842	2012.5
-2.8683	-2.639329429	2012.75
-2.5074	-2.425231128	2013
-1.9898	-1.579216948	2013.25
-0.9886	-1.710322516	2013.5

. . .

3. Check for Non-Existent Trend and Seasonality

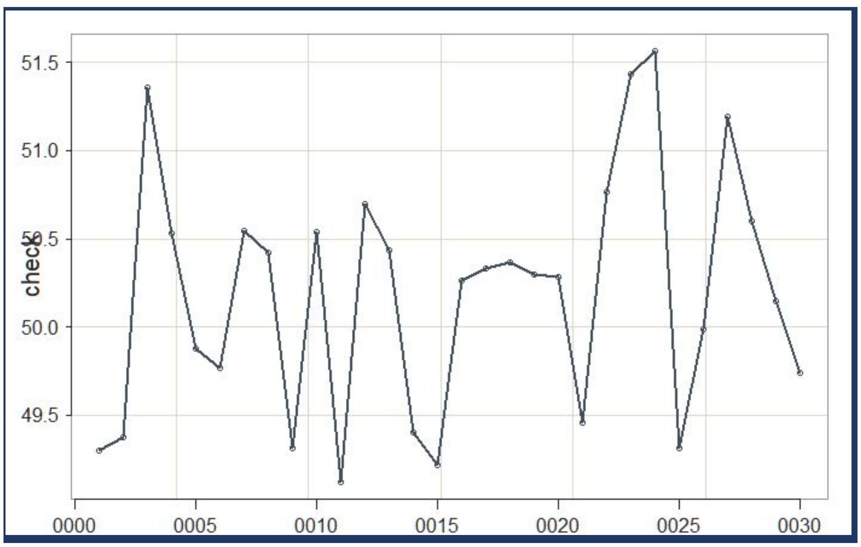
HW #2 featured four time series data sets of a stable process, each with increasing amounts of error. Return to the third time series, on the *third* (not the first) worksheet of the following data file:

http://web.pdx.edu/~gerbing/451/Data/HW3_1.xlsx

Because the process is stable, the fluctuations are irregular, not the result of seasonality. Moreover, there is no trend. To understand more about stl(), use it to decompose the time series into trend and seasonal components. In other words, you know this particular time series is stable, but in a real-world data analysis, you would not likely know that. How can you tell that there are no real trend and seasonal components in the data by using stl()?

a. Plot (once again) the time series and verbally describe.

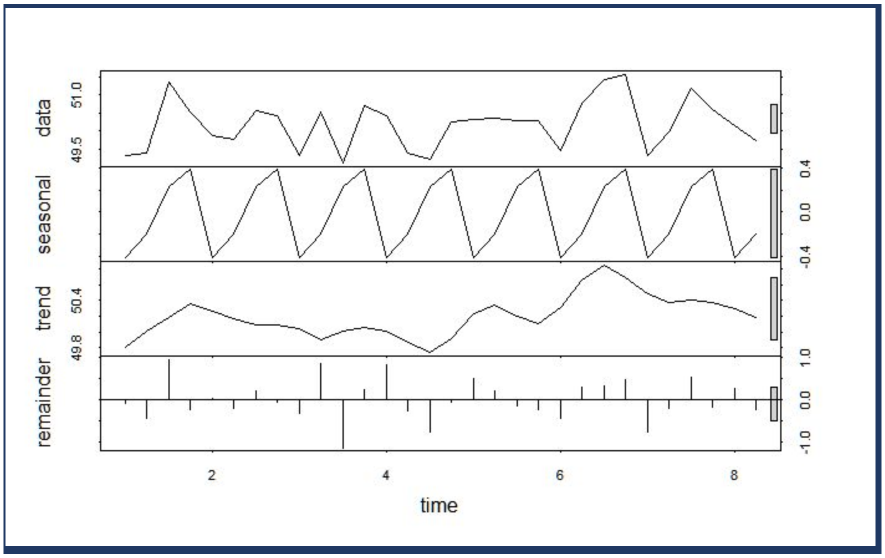
```
check <- ts(d$Y3)
Plot(check)
```



b. Do the stl() decomposition. Does it report trend and seasonality?

Assuming that I set “frequency=n” when I assign the ts object, stl will provide an output that assumes seasonality. However, it does not display any consistent trend whatsoever.

c. Plot the corresponding graph of the decomposition (here, frequency=4):



d. To informally evaluate the fit of the trend + seasonality model, compare the range of the errors to the range of the data values. Does the model appear to fit (i.e., adequately describe) the data?

The range of the errors are three or four orders of magnitude less than the data values, while the trend hovers right around the mean of the overall dataset.

e. We have other, more intuitive evidence that the up and down peaks in this time series do not represent seasonality. Count the time periods between each maximum. Show the resulting set of values. Are the intervals between maximum values in the time series regular or irregular?

The intervals between the maximum and minimum values are irregular (ranging from 1 to 3 steps between), indicating some other source of variation than a seasonal influence.

4. Forecast from Linearity

Return to the forecasting problem from HW #4.

Currency conversions from the dollar are important to the global supply chain. Consider the following data that shows the value of the Euro and the Pound to the USD from the end of October, 2018 into April of 2019.

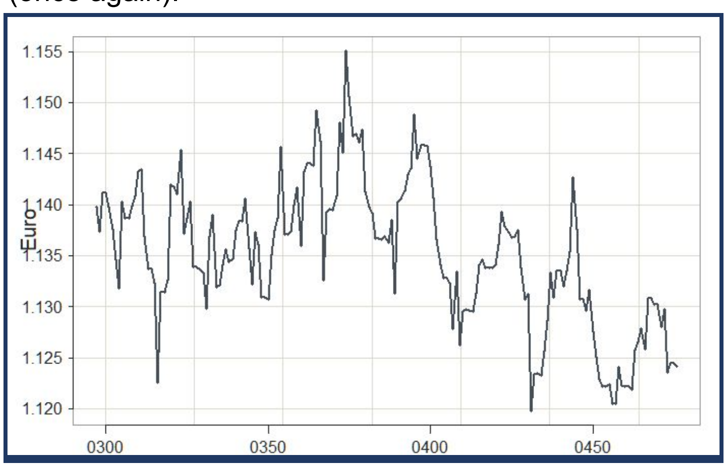
Data: http://web.pdx.edu/~gerbing/451/Data/HW4_2.xlsx

Here use the forecast() function to forecast from the linear trend.

a. Create the R ts object of the data. These data are daily, so the frequency is 1. The data begin on Oct 24 of 2018, which is the 297th day of the year. So use start=c(297,1).

```
Euro <- ts(d$Euro, start=c(297,1))
     Time Series:
     Start = 297
     End = 476
     Frequency = 1
```

b. Plot the data (once again).

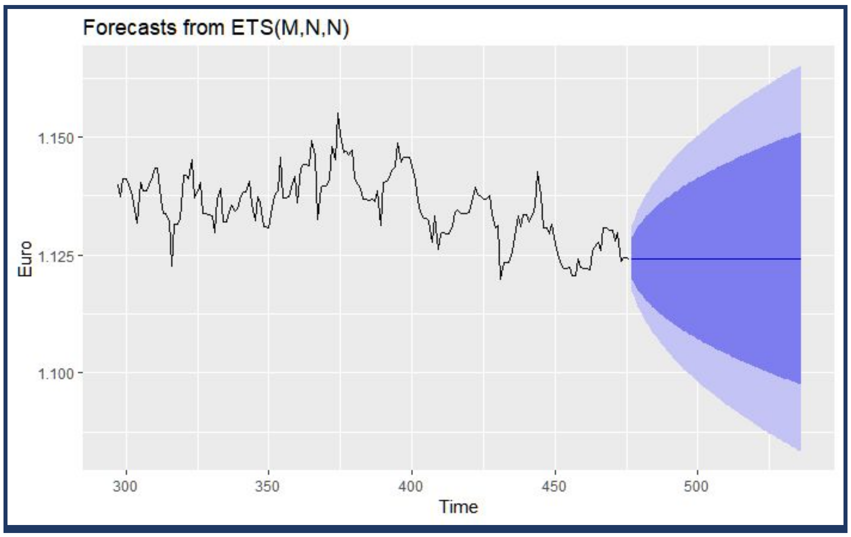


c. Verbally describe the pattern of the data.

The data has no consistent trend and indeterminate seasonality.

d. In HW #4 we did a linear forecast with the Regression() function. Here use forecast() to obtain a plot of the data and the linear forecasted values for two months from the last data value. Show the visualization.

```
f <- forecast(Euro, h=60)
autoplot(f, ylab="Euro")
```



e. In general, the forecast is not wrong, but the first several values of the forecast should likely be adjusted, if even by intuition. How would you adjust those values? Why?

This may be due to my tired state as I'm working this problem, but I don't understand what you're asking for here.

5. Logic of Deseasonalization

Return to the Sales data in Worked Problem #1. Just focus on the last 20 values of the time series.

The highlight colors correspond to the part of the table where the calculations were performed:

a. Using Excel, calculate the 4-pt moving average for each data value for which the average can be computed.

b. Calculate the corresponding centered moving average for each relevant data value.

c. Calculate the specific seasonal index for each data value.

d. Calculate the average seasonal index for each of the four seasons.

e. Adjust the initial averages so that they sum to 0. What are the four seasonal indices?

Year	Qtr	Sales	4-point Moving Average	Centered Moving Average	Specific Seasonal Index		0.25	Average Seasonal Index	**Seasonal Index**
2019	Summer	-6.95	-7.84				Winter	0.08	-0.17
2019	Fall	-6.21	-7.90	-7.87	1.66		Spring	0.03	-0.23
2020	Winter	-10.57	-9.04	-8.47	-2.10		Summer	1.07	0.82
2020	Spring	-7.61	-9.08	-9.06	1.45		Fall	-0.15	-0.41
2020	Summer	-7.20	-9.50	-9.29	2.10			1.03	0.00
2020	Fall	-10.77	-9.53	-9.51	-1.26				
2021	Winter	-10.75	-9.73	-9.63	-1.12				
2021	Spring	-9.27	-9.07	-9.40	0.13				
2021	Summer	-7.31	-9.97	-9.52	2.21				
2021	Fall	-11.57	-10.97	-10.47	-1.10				
2022	Winter	-8.15	-10.25	-10.61	2.46				
2022	Spring	-12.86	-11.03	-10.64	-2.22				
2022	Summer	-11.32	-10.60	-10.81	-0.51				
2022	Fall	-8.66	-10.72	-10.66	2.00				
2023	Winter	-11.26	-12.19	-11.46	0.19				
2023	Spring	-11.15	-12.17	-12.18	1.03				
2023	Summer	-11.79	-12.42	-12.29	0.50				
2023	Fall	-14.58	-12.63	-12.52	-2.05				
2024	Winter	-11.15	-11.65	-12.14	0.99				
2024	Spring	-12.15	-12.15	-11.90	-0.25				