DATA 624 -- HW 1

**Libraries**

|  |
| --- |
| library(forecast)  library(ggplot2)  library(seasonal)  library(fma)  library(mlbench)  library(corrplot)  library(caret)  library(e1071)  library(mice) |

**HW 6.2**

|  |
| --- |
| **Question**  The plastics data set consists of the monthly sales (in thousands) of product A for a plastics manufacturer for five years.  a) Plot the time series of sales of product A. Can you identify seasonal fluctuations and/or a trend-cycle?  b) Use a classical multiplicative decomposition to calculate the trend-cycle and seasonal indices.  c) Do the results support the graphical interpretation from part a?  d) Compute and plot the seasonally adjusted data.  e) Change one observation to be an outlier (e.g., add 500 to one observation), and recompute the seasonally adjusted data. What is the effect of the outlier?  f) Does it make any difference if the outlier is near the end rather than in the middle of the time series? |
| **Code**  ```{r}  p = plastics  print(p)  ```  # Part A: Plot the time series of sales of product A. Can you identify seasonal fluctuations and/or a trend-cycle?  ```{r}  p <- plastics  autoplot(p, series="Data") + xlab("Year") + ylab("Sales") + ggtitle("Product A: Monthly Sales '000'")  ```  # Part B: Use a classical multiplicative decomposition to calculate the trend-cycle and seasonal indices.  ```{r}  # Decompose plastics data  m = decompose(p, type="multiplicative")  # Calculating trend-cycle  print(m$trend)  # Calculating seasonal indices  print(m$seasonal)  # Calculating remainder component  print(m$random)  ```  # Part C: Do the results support the graphical interpretation from part a?  ```{r}  # Plot seasonal indices  autoplot(m$seasonal)  # plot trend-cycle  autoplot(m$trend)  # plot trend-cycle  autoplot(m$random)  ```  # Part D: Compute and plot the seasonally adjusted data.  ```{r}  # Seasonal adjustment of multiplicative decomposition  p\_adj <- p/m$seasonal  autoplot(p\_adj)  ```  # Part E: Change one observation to be an outlier (e.g., add 500 to one observation), and recompute the seasonally adjusted data. What is the effect of the outlier?  ```{r}  p[15] <- p[15] + 500  m\_1 <- decompose(p, type = "multiplicative")  p\_adj\_1 <- p/m\_1$seasonal  autoplot(p\_adj\_1)  autoplot(m\_1$trend)  autoplot(m\_1$random)  ```  # Part F: Does it make any difference if the outlier is near the end rather than in the middle of the time series?.  ```{r}  p[10] <- p[10] + 500  p[50] <- p[50] + 500  m\_2 <- decompose(p, type = "multiplicative")  p\_adj\_2 <- p/m\_2$seasonal  autoplot(p\_adj\_2)  autoplot(m\_2$trend)  autoplot(m\_2$random)  ``` |
| Response   1. The plot above shows that there is a seasonal nature to the plastics data set. It appears that the seasonality is annual will sales peaking at mid-year.   There trend cycle appears to be positive with strong sales growth every year.   1. See output above.   Note that the estimate of the trend-cycle is unavailable for the first 6 and last 6 observations and as a result there is also no estimate of the remainder component for the same period.  This is a common problem with the classical decomposition.   1. The plot of seasonal decomposition shows a clear pattern of annual seasonality while the trend-cycle plot shows a strong positive trend of growth in sales over the 5-year period.   This confirms the conclusion made under graph number 1.   1. See output above. 2. The addition of an outlier to the data impacts decomposition because the classical decomposition methods (multiplicative in this case) are unable to capture these seasonal changes over time.   The trend-cycle also tends to over-smooth this rapid rise in the data.  This is an inherent weakness of this method.   1. The position of the outlier does not seem to make any difference in the decomposition since the classical methods are unable to capture these seasonal changes over time.   It’s also evident that the trend-cycle tend to over-smooth the two rapid rises in the data because of the two introduced outliers. |