**FORECASTING OF SHAMPOO SALES DATA [JANUARY-MARCH]**

**INTRODUCTION**

***Problem Description:*** Here in this problem we are interested in

1. Fitting a suitable predictive model for the dataset using Holt’s exponential smoothing technique.
2. And futher in obtaining the insample and out of sample forecast for next 5 data points.
3. We are also interested in Obtaining the accuracy measures and comment about the findings.

***Objective:*** The main objective of this problem is to fit a predictive model using Holt’s exponential smoothing for the time series dataset and we want to find the insample forecast as well as out of sample forecast for next 5 data points. Further we are also interseted in obtaining the accuracy measure and draw a conclusion out of it.

***Holt’s Exponential smoothing:*** Exponential smoothing may readily be generalized to deal with time series containing trend and seasonal variation. The version for handling a trend with non-seasonal data is usually called Holt’s (two-parameter) exponential smoothing.

***In sample forecast:*** In-sample forecast is the process of formally evaluating the predictive capabilities of the models developed using observed data to see how effective the algorithms are in reproducing data. It is kind of similar to a training set in a machine learning algorithm.

***Out of Sample forecast:*** An out of sample forecast instead uses all available data in the sample to estimate a models. For example, here estimation would be performed over 1, Jan- 12,March, and the forecast(s) would commence in March, 13th- March, 17th.

*#Setting and getting the current working directory.*  
**setwd**("E:/M.Sc/SEM III/TIME\_SERIES\_ANALYSIS(MST371)/Practical Labs")  
**getwd**()

## [1] "E:/M.Sc/SEM III/TIME\_SERIES\_ANALYSIS(MST371)/Practical Labs"

***Data Description:***

This dataset describes the monthly number of sales of shampoo over a 3 year period.

The units are a sales count and there are 36 observations.

Also it is observed that the data has been recorded from january to Feburary for first 12 days in each month.

*#Loading the package required to load the dataset.*  
**library**(readxl)  
  
*#Loading the 'shampoo sales' dataset.*  
shampoo <- **read\_excel**("E:/M.Sc/SEM III/TIME\_SERIES\_ANALYSIS(MST371)/shampoo.xlsx")  
  
*#Obtaining the first few records of the dataset.*  
**head**(shampoo)

## # A tibble: 6 x 2  
## Month Sales  
## <dttm> <dbl>  
## 1 2021-01-01 00:00:00 266   
## 2 2021-01-02 00:00:00 146.  
## 3 2021-01-03 00:00:00 183.  
## 4 2021-01-04 00:00:00 119.  
## 5 2021-01-05 00:00:00 180.  
## 6 2021-01-06 00:00:00 168.

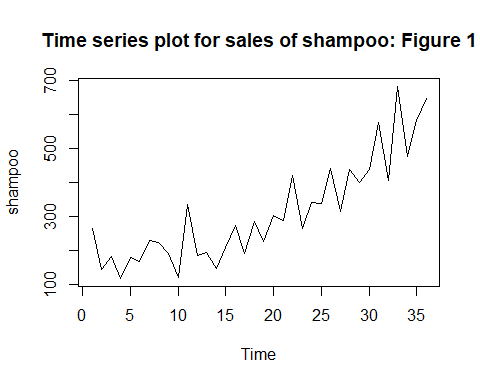
**ANALYSIS**

*#Extracing the data for the sales variable which we are interested in.*  
shampoo <- shampoo**$**Sales  
  
*#Now converting the it into a time series data.*  
shampoo1=**ts**(shampoo)  
  
*#Here. we are checking if the dataset has been converted into a timeseries plot.*  
**class**(shampoo1)

## [1] "ts"

Hence, now the dataset we are interested in is a timeseries data.

*#Now we will first obtain the time series plot of the data to understand the nature of the time series data.*  
**ts.plot**(shampoo1, gpars = **list**(main="Time series plot for sales of shampoo: Figure 1", xlab="Time",ylab="shampoo",lty=**c**(1**:**20)))



***Interpretation:*** From the above time series plot we observe that there exists a trend component in the dataset since there is observed a increase pattern for a longer period of time. Also we observe that there is some kind of irregularity in the dataset hence we can say that there also exists a error component in the dataset.

Since, from the above time series plot we observed that our time series data has only trend component therefore we fit the redictive model for the dataset using Holt’s exponential smoothing technique.

1. Fitting a suitable predictive model for the dataset using Holt’s exponential smoothing technique.

*#Forecasting using Holt's exponential smoothing.*  
shampoo\_forecast<-**HoltWinters**(shampoo1, beta = TRUE, gamma = FALSE)  
shampoo\_forecast

## Holt-Winters exponential smoothing with trend and without seasonal component.  
##   
## Call:  
## HoltWinters(x = shampoo1, beta = TRUE, gamma = FALSE)  
##   
## Smoothing parameters:  
## alpha: 0.3563877  
## beta : TRUE  
## gamma: FALSE  
##   
## Coefficients:  
## [,1]  
## a 612.67168  
## b 23.92143

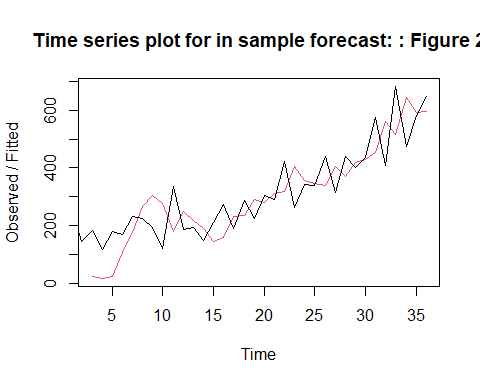
***Interpretation:*** Thus, we observe from the above model that the alpha value = 0.3563877 which is farther from 1 which indicates that the weightage is given to all the observations that means forecast is not just on the basis of recent past observation it is based on all the observations.

1. Obtaining the in sample and out of sample forecast for next 5 data points.

*#Performing in-sample forecast.*  
shampoo\_forecast**$**fitted

## Time Series:  
## Start = 3   
## End = 36   
## Frequency = 1   
## xhat level trend  
## 3 25.80000 145.90000 -120.1000000  
## 4 17.81958 81.85979 -64.0402083  
## 5 26.11213 53.98596 -27.8738318  
## 6 108.13963 81.06280 27.0768356  
## 7 178.23986 129.65133 48.5885311  
## 8 265.00475 197.32804 67.6767092  
## 9 303.81067 250.56935 53.2413146  
## 10 277.92630 264.24782 13.6784744  
## 11 181.10583 222.67683 -41.5709983  
## 12 250.29599 236.48641 13.8095796  
## 13 218.20569 227.34605 -9.1403603  
## 14 192.02594 209.68599 -17.6600536  
## 15 144.05444 176.87022 -32.8157768  
## 16 158.31432 167.59227 -9.2779484  
## 17 230.99534 199.29381 31.7015389  
## 18 234.47429 216.88405 17.5902434  
## 19 289.50357 253.19381 36.3097616  
## 20 280.54954 266.87168 13.6778668  
## 21 310.65721 288.76444 21.8927666  
## 22 317.75475 303.25960 14.4951511  
## 23 406.26825 354.76392 51.5043265  
## 24 356.72364 355.74378 0.9798606  
## 25 347.42268 351.58323 -4.1605489  
## 26 337.75760 344.67041 -6.9128191  
## 27 404.00577 374.33809 29.6676757  
## 28 370.87381 372.60595 -1.7321392  
## 29 417.91418 395.26007 22.6541155  
## 30 428.72612 411.99309 16.7330247  
## 31 451.64167 431.81738 19.8242907  
## 32 559.74914 495.78326 63.9658803  
## 33 515.26684 505.52505 9.7417910  
## 34 643.85194 574.68850 69.1634439  
## 35 592.87569 583.78210 9.0935983  
## 36 593.71842 588.75026 4.9681630

*#Visualizing the in-sample forecast.*  
**plot**(shampoo\_forecast, main = "Time series plot for in sample forecast: : Figure 2")



Thus, the in sample forecasted values are obtained in the above table and are plotted in figure 2.

*#Loading the library 'forecast'.*  
**library**(forecast)

## Warning: package 'forecast' was built under R version 4.0.5

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

*#Forecasting the last observation to compare with actual out-sample forecast result if the model obtained is good or not.*  
shampoo\_forecast1<-**HoltWinters**(shampoo1[1**:**31], beta=TRUE, gamma = FALSE)  
shampoo\_forecast1

## Holt-Winters exponential smoothing with trend and without seasonal component.  
##   
## Call:  
## HoltWinters(x = shampoo1[1:31], beta = TRUE, gamma = FALSE)  
##   
## Smoothing parameters:  
## alpha: 0.3950397  
## beta : TRUE  
## gamma: FALSE  
##   
## Coefficients:  
## [,1]  
## a 500.72748  
## b 67.78064

forecast1<-**forecast**(shampoo\_forecast1,h=5)  
forecast1

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 32 568.5081 455.0339 681.9824 394.9642 742.052  
## 33 636.2888 491.6714 780.9061 415.1155 857.462  
## 34 704.0694 506.5872 901.5515 402.0465 1006.092  
## 35 771.8500 505.1098 1038.5902 363.9061 1179.794  
## 36 839.6307 491.2251 1188.0363 306.7903 1372.471

Thus, from the above for table it is observed that the forecasted values are very different from the actual values in the dataset.

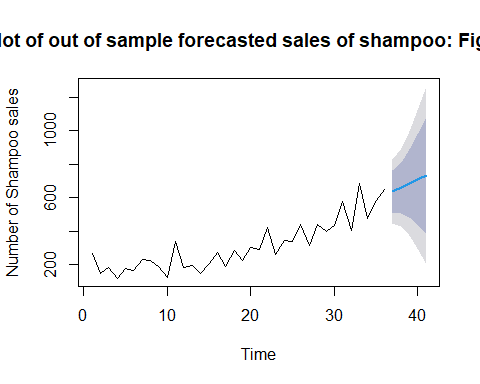
*#Performing out-sample 5 step ahead forecast for March, 13- March,17.*  
forecast2<-**forecast**(shampoo\_forecast, h=5)  
forecast2

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 37 636.5931 512.7343 760.4519 447.1675 826.0187  
## 38 660.5145 508.4127 812.6164 427.8948 893.1342  
## 39 684.4360 482.7644 886.1075 376.0059 992.8660  
## 40 708.3574 440.3139 976.4009 298.4202 1118.2945  
## 41 732.2788 385.0618 1079.4958 201.2562 1263.3014

Thus, the out of sample forecast for next 5 data points, i.e. shampoo sales values from March, 13- March,17 are obtained.

*#Plotting the forecasted value.*  
**plot**(forecast2, main = "Plot of out of sample forecasted sales of shampoo: Figure 3", xlab ="Time", ylab="Number of Shampoo sales")

**Plot of out of sample forecasted values of sales of shampoo**





Thus, the forecasted values are plotted in the graph.

1. Obtaining the accuracy measures and comment about the findings.

***Root mean square value***

*#Loading the package 'Metrics'.*  
**library**(Metrics)

## Warning: package 'Metrics' was built under R version 4.0.5

##   
## Attaching package: 'Metrics'

## The following object is masked from 'package:forecast':  
##   
## accuracy

*#Obtaining the accuracy measure for the forecasting model obtained by removing the last value.*  
ma<-**rmse**(shampoo1,shampoo\_forecast**$**fitted)  
ma

## [1] 210.079

***Interpretation:*** It is observed from the above calculations that the values obtained using forecasting model are way too far from the actual values also the root mean square for the above model is 210.079 which is evident to the fact that the above model is not that good.

**CONCLUSION**

Thus, from the above analysis we interpret that the forecasting model obtained for the shampoo sales data is not good that is the model is less accurate. Thus, the forecasted value for shampoo sales from March, 13- March,17 is also not correct.