*Name: Hagar Haytham*

*Sec: 2*

*B.N: 28*

***Data Mining, Big Data and Analytics***

**1.After inspecting the time series:**

**a) What is the time range (start and end) of this time series?**

Jan-2003 to Dec-2014

**b) How many values are there in this time series?**

144 Value

**c) What is the time interval separating between each two consecutive**

**values in the time series? (monthly, weekly, daily, yearly, etc.)**

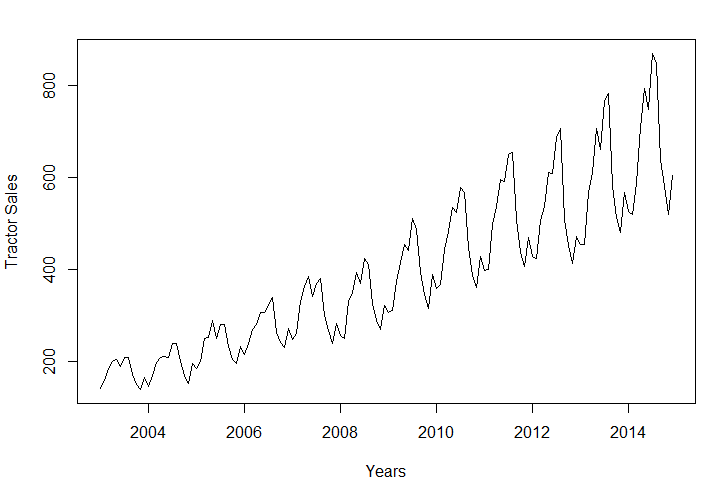
Monthly

**2. What does the parameter frequency mean? Why did we set it to 12?**

Frequency is the number of observations per unit of time. As the unit here is year as we start from 2003, each year contains 12 months which means we have 12 values per unit of time.

**3. After visualizing the time series:**

**a) Add a neat plot of the generated time series.**

****

**b) Do you think there is a trend in the time series? If yes, then what is the**

**degree of the trend (i.e. is it linear, quadratic, etc.)?**

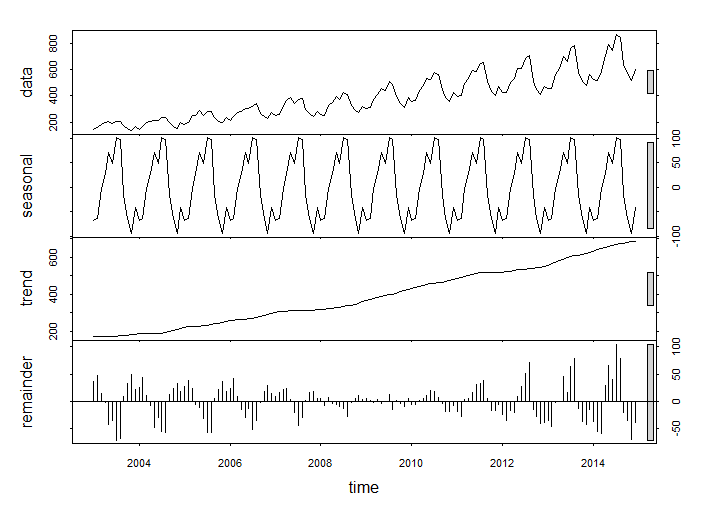
Yes, Linear.

**c) Do you think there is a seasonality in the time series?**

Yes

**4. What does the function stl do? Add a neat plot of the plot generated.**

Stl decompose a time series into seasonal, trend and irregular components (remainder)

****

**5. Back to the original time series:**

**a) What are the two conditions imposed on the mean and the variance**

**of a time series to be stationary?**

The mean and variance have to not change over time for a time series to be stationary.

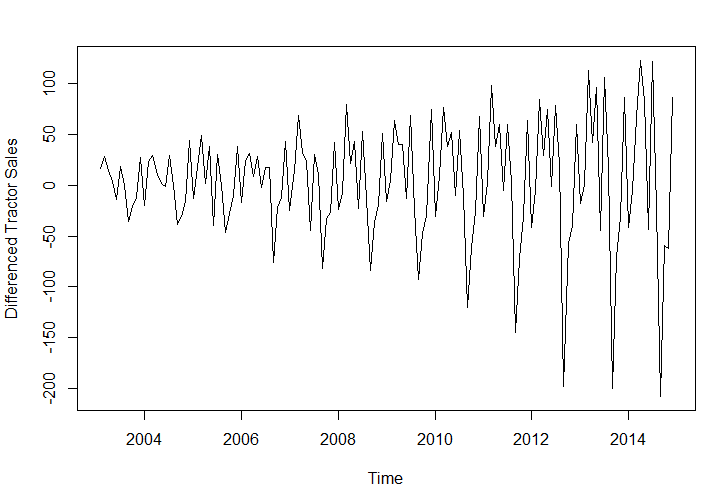
**b) Is this time series stationary? Mention the reasons behind your**

**answer.**

No, As the time series contains the trend and seasonal effect. If we made detrending the mean and variance won’t change over time, also the auto correlation structure should be stable over time (i.e. remove seasonal effects) and the time series would be stationary.

**6. After differencing the time series:**

**a) Add a neat plot of the time series after differencing.**

****

**b) Does the time series become stationary? Are the two conditions of the**

**mean and variance satisfied?**

No, the time series is not stationary yet.

The mean condition is satisfied while the variance condition is not satisfied.

**c) If no, which of the two conditions is still not satisfied for a stationary**

**time series?**

The variance still changes over time.

**d) How does differencing help (not guarantee) to make a time series**

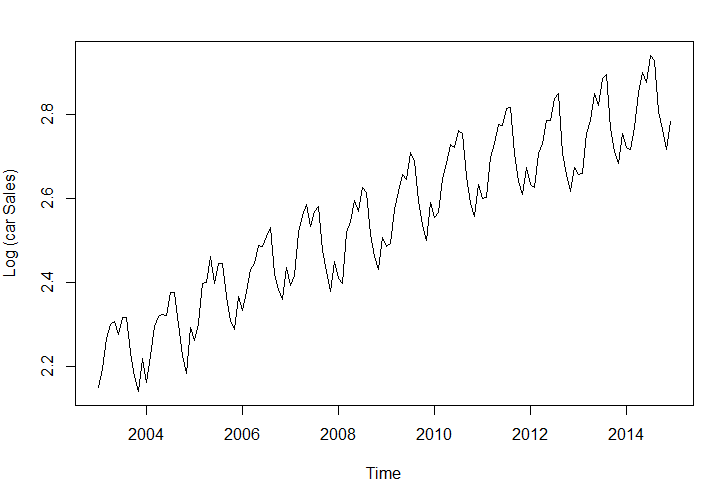
**stationary?**

Differencing helps in detrending (especially in stabilizing the mean over time).

**7. After applying logarithm to the time series, compare visually (5) and (7)**

**and you will know the answer to the following questions:**

**a) Add a neat plot of the time series after applying logarithm.**

****

**b) Does the time series become stationary? Are the two conditions of the**

**mean and variance satisfied?**

No, it is not stationary.

No, the variance condition is satisfied while the mean condition is not satisfied.

**c) If no, which of the two conditions is still not satisfied for a stationary**

**time series?**

The mean condition as it changes with time.

**d) How does applying logarithm help (not guarantee) to make a time**

**series stationary? [This is a new piece of information never told in the**

**lecture or the tutorial].**

It helps in stabilizing the variance (make it not change over time) so helps in detrending

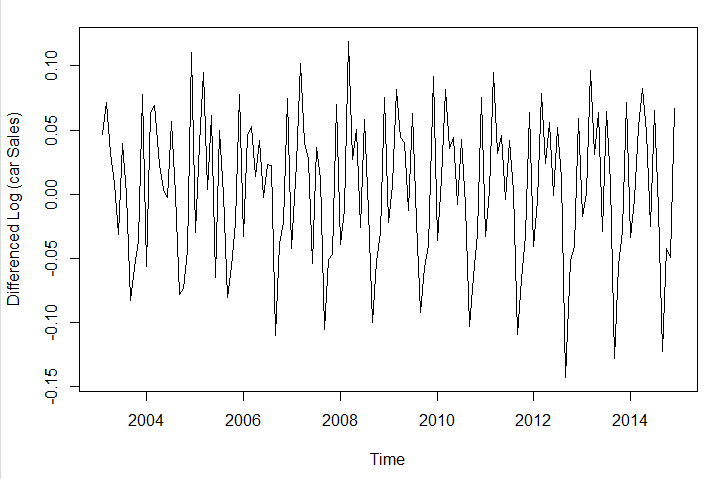
**8. After applying both differencing and logarithm to the time series:**

**a) Does the time series become stationary? Are the two conditions of the**

**mean and variance satisfied?**

No, the time series is not stationary, as to be stationary we have to detrend(mean and variance don’t change over time) and remove seasonal effect(autocorrelation doesn’t change over time).

**b) Add a neat plot of the final time series.**

****

**9. After fitting an ARIMA Model with the logarithm of the time series:**

**a) What are the two requirements of ARIMA (or ARMA) models on the**

**time series data?**

The two requirements are detrended and seasonally adjusted time series.

**b) Does the time series passed to the ARIMA model successfully hold the**

**two requirements? Why?**

No, it only holds a part from the requirement of detrending. As detrending contains both adjusting mean and variance. The log only adjusts the variance.

Also, the Time Series still contains seasonal effects.

**c) Inspect the summary of the model.**

**The output of the ARIMA model is: (p, d, q) (P, D, Q) [S]**

**What does the (p, d, q) mean?**

The p means how many terms before the current term are taken into consideration for the autoregressive (AR) model

The d is the number of differences to detrend the time series (d=1 to remove a linear trend)

The q is the number of moving average terms for the moving average (MA) model

**What do you think the (P, D, Q) [S] relate to? (no details are needed).**

They relate to the seasonal part of the time series

The S part is for the units of time in which seasonal effects occur (12 months)

**d) Do you think that ARIMA model achieved the two requirements of (9-a)**

**internally? If yes, how did it happen briefly? You don’t need to give any**

**mathematical proofs or so. You just need to observe the ARIMA model**

**output (9-c) and you will get it.**

Yes, Cause ARIMA model applies differencing to detrend the time series(make mean stable over time)and also search for non-seasonal models (since seasonal parameter is true by default)

**e) What do you think will be more suitable for the case of forecasting the**

**tractor sales, an autoregressive (AR) model or a moving average (MA)**

**model? Why?**

MA model is more suitable as the problem (time series) shows short term dependencies between data.

**10. After changing trace = True:**

**a) How is the best model selected? What is the information criterion used**

**in selecting the best model? [Mention only the name] [Check the help]**

The model is selected using Information Criteria (IC) which is

AICc (it is AIC with a correction for small sample sizes.)

**b) What other information criteria are there that can be used as well?**

**[Mention only the names] [Check the help]**

AIC (Akaike’s Information Criterion)

or BIC (Bayesian Information Criterion)

**c) Do we seek to get the minimum value or the maximum value of this**

**criterion?**

We seek to get the minimum value.

**11. What is the meaning of n.head = 36?**

The number of steps ahead for which prediction is required is 36 months (3 years)

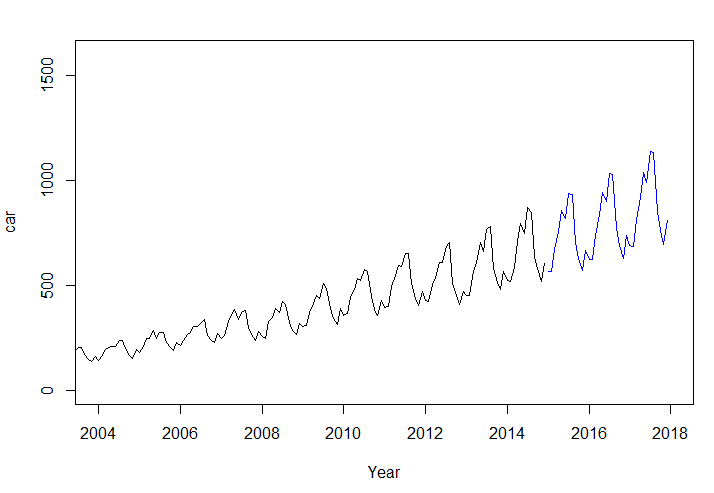
**12. After forecasting and plotting the future values,**

**a) According to your observation, does this forecast work well?**

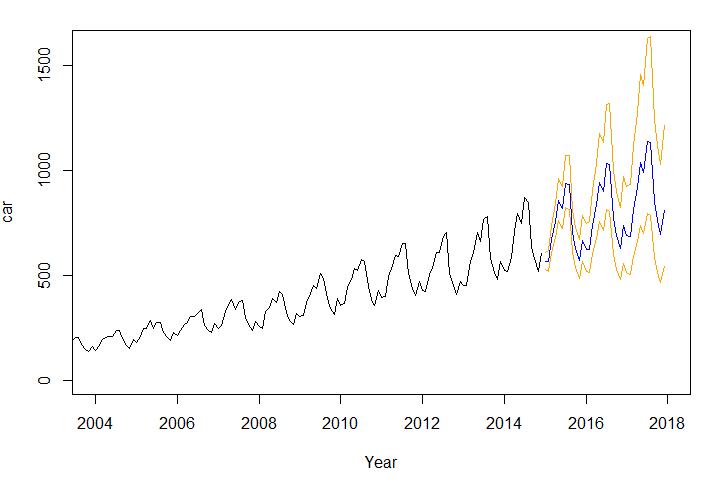
Yes, it works well as it similar to what happened in the previous years and following the same trend and seasonality.

**b) Add a neat plot of the generated time series.**

The prediction only

****

The prediction and its +- 2 std dev. (expected error)

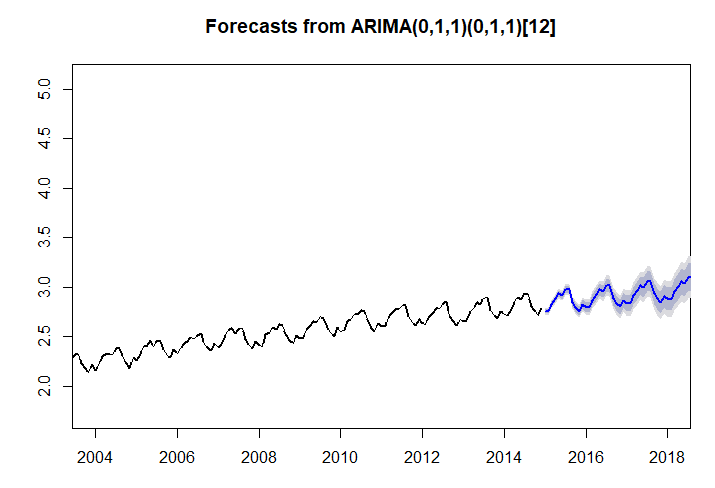
****

**13. After using TSPred library:**

**a) Does this library generate a similar plot to (12)? Why?**

No, cause the data is passed to log so the variance doesn’t change over time here

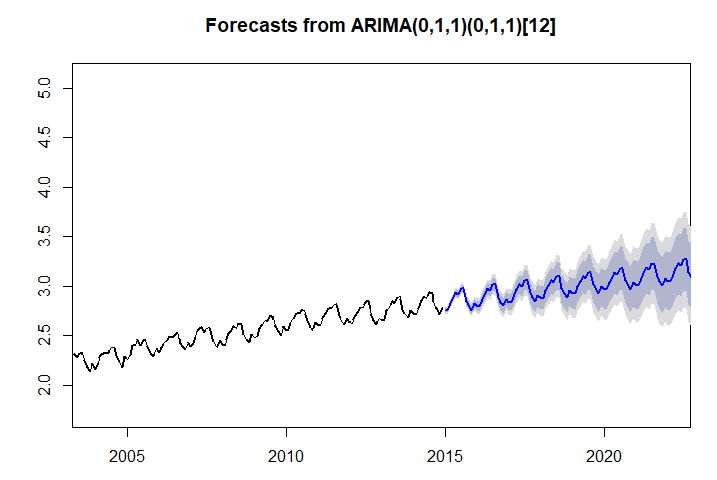
**b) Add a neat plot of the generated time series.**

****

**14. What happened when we tried to forecast the tractor sales for an**

**extended or longer time range? What do you notice?**

The range for the expected error increases

****