INTERMEDIATE ANALYTICS

ALY6015, FALL 2019

MODULE 5 ASSIGNMENT

TIME SERIES

SUBMITTED BY: SHIVANI ADSAR

NUID: 001399374

CRN: 71933

SUBMITTED TO: PROF. LI, TENGLONG

DATE: 10/12/2019

Introduction

We have worked on the “Births” and “Skirts” datasets to perform time series analysis.

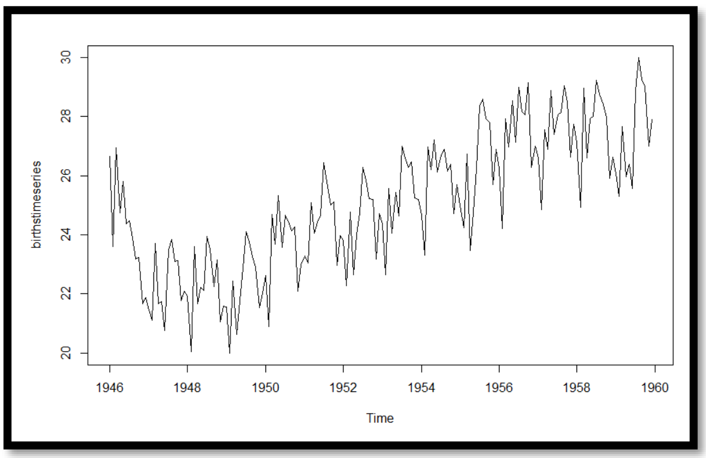
Analysis

Decomposition of Seasonal Time Series

- We have loaded the “Births” dataset in a variable.
- We have used the ts() function to analyze the time series data.

```
> birthsts <- ts(births, frequency=12, start=c(1946,1))
> birthsts
        Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
1946 26.663 23.598 26.931 24.740 25.806 24.364 24.477 23.901 23.175 23.227 21.672 21.870
1947 21.439 21.089 23.709 21.669 21.752 20.761 23.479 23.824 23.105 23.110 21.759 22.073
1948 21.937 20.035 23.590 21.672 22.222 22.123 23.950 23.504 22.238 23.142 21.059 21.573
1949 21.548 20.000 22.424 20.615 21.761 22.874 24.104 23.748 23.262 22.907 21.519 22.025
1950 22.604 20.894 24.677 23.673 25.320 23.583 24.671 24.454 24.122 24.252 22.084 22.991
1951 23.287 23.049 25.076 24.037 24.430 24.667 26.451 25.618 25.014 25.110 22.964 23.981
1952 23.798 22.270 24.775 22.646 23.988 24.737 26.276 25.816 25.210 25.199 23.162 24.707
1953 24.364 22.644 25.565 24.062 25.431 24.635 27.009 26.606 26.268 26.462 25.246 25.180
1954 24.657 23.304 26.982 26.199 27.210 26.122 26.706 26.878 26.152 26.379 24.712 25.688
1955 24.990 24.239 26.721 23.475 24.767 26.219 28.361 28.599 27.914 27.784 25.693 26.881
1956 26.217 24.218 27.914 26.975 28.527 27.139 28.982 28.169 28.056 29.136 26.291 26.987
1957 26.589 24.848 27.543 26.896 28.878 27.390 28.065 28.141 29.048 28.484 26.634 27.735
1958 27.132 24.924 28.963 26.589 27.931 28.009 29.229 28.759 28.405 27.945 25.912 26.619
1959 26.076 25.286 27.660 25.951 26.398 25.565 28.865 30.000 29.261 29.012 26.992 27.897
```

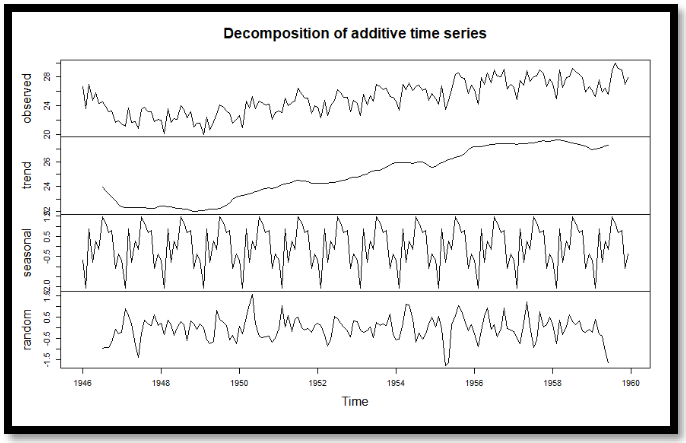
- We have plotted the time series using the plot.ts() function.



Observations: It can be seen that the number of births every month in New York is seasonal. Also, it can be noted that it reaches the peak every summer and winter. The variations are random and constant with time.

- Decomposition of time series: This involves separation of data into components. The method includes estimating the seasonal and trend component in the data.
- We have used the decompose() function for estimating the components of time series.

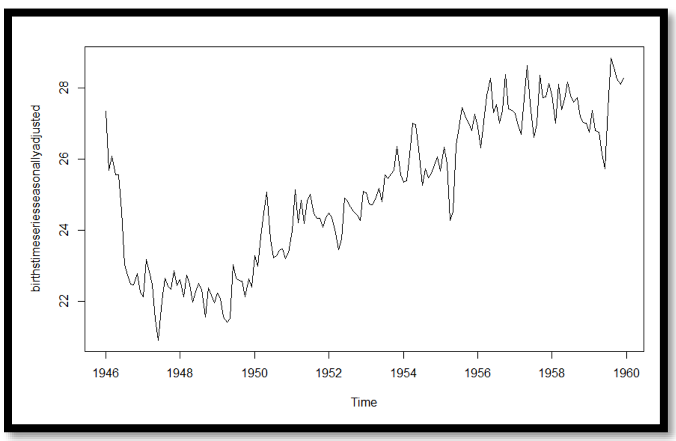
```
> birthstsc <- decompose(birthsts)
> birthstsc$seasonal
```



Observations: The plot displays the observed, trending, seasonal and random estimates. It is noted that the estimated trend plot has a minimum decrease from 24 (in 1947) to 22 (in 1948) and then a steady increase to 27 (in 1959).

- In order to have seasonally adjusted time series data, we can estimate the seasonal component by using the decompose() function and then subtract the seasonal component from the original data

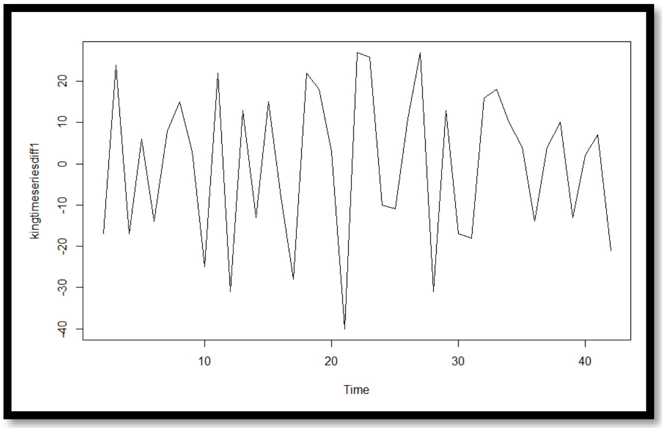
```
> birthstsc <- decompose(birthstimeseries)
> birthstimeseriesseasonallyadjusted <- birthsts - birthstsc$seasonal
> plot(birthstimeseriesseasonallyadjusted)
```



Observations: The seasonal component has been removed from the data. So, the seasonally adjusted time series has the irregular and the trend component.

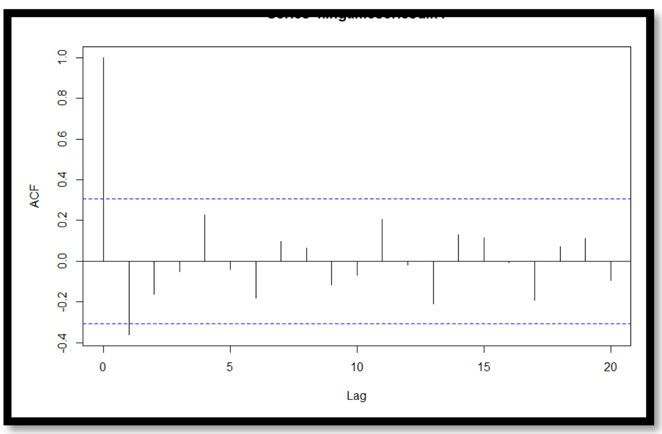
<u>Autocorrelation and ARIMA</u>

- Autoregressive Integrated Moving Average models allow for the non-zero autocorrelations in the irregular component and also include a statistical model.
- The time series for the age of death of kings is not stationary in mean. So, we will calculate the time series of differences and plot the data.



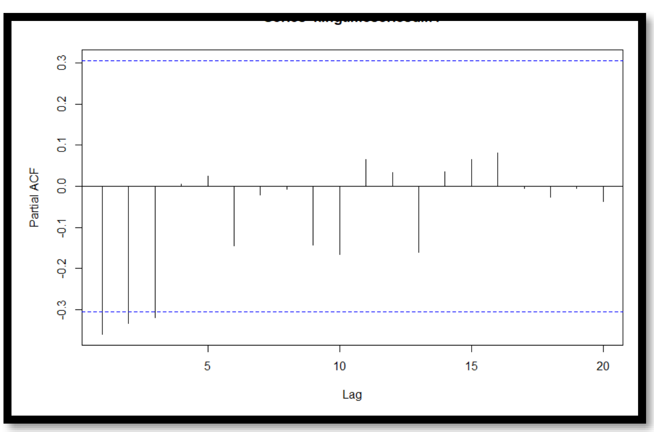
Observations: Now, the time series of differences in the age of deaths seems to be stationary and hence we can perform the ARIMA. We can use the data to examine the correlations between the data.

- Since the time series is stationary, we can now select the ARIMA model for finding the values of p and q by scrutinizing the correlations.
- In order to get the actual values of autocorrelations, we have set the plot=FALSE in acf() and pacf() functions.



Observations: It is seen that the autocorrelation at lag 1 is greater than the significance level whereas the correlations between lags 1 to 20 are not greater than the significance levels

- Plotting partial correlations:



Observations: It can be seen that the partial autocorrelations at lags 1,2 and 3 are negative and decrease in magnitude. The correlations move to 0 after lag 3.

- Using the principle of parsimony, the model with lesser parameters is more efficient. ARIMA model is chosen to be the best model for time series of ages of deaths of kings.

- An ARIMA model is the moving average model having order 1 and is beneficial for displaying the time series with short term dependencies amongst its successive observations

References

- Using R for Time Series Analysis¶. (n.d.). Retrieved from https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html#decomposing-time-series.