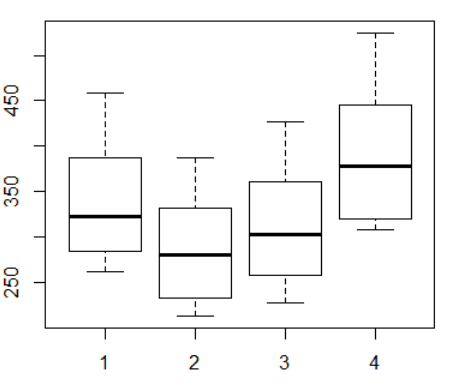
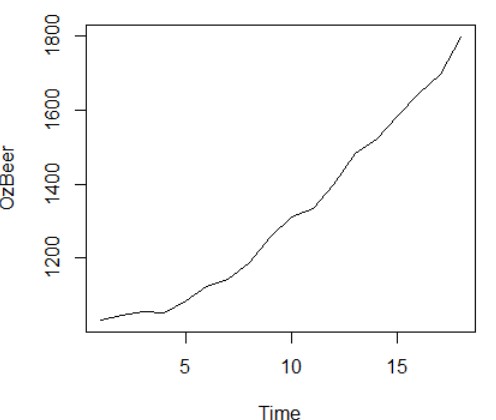
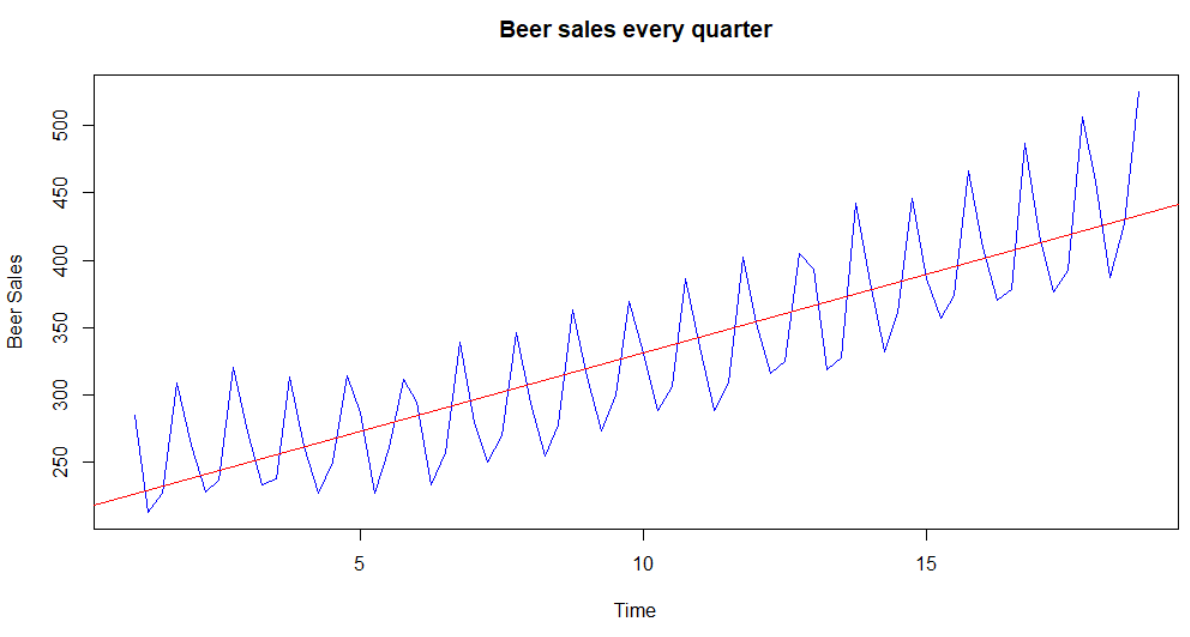
**Problem Statement**

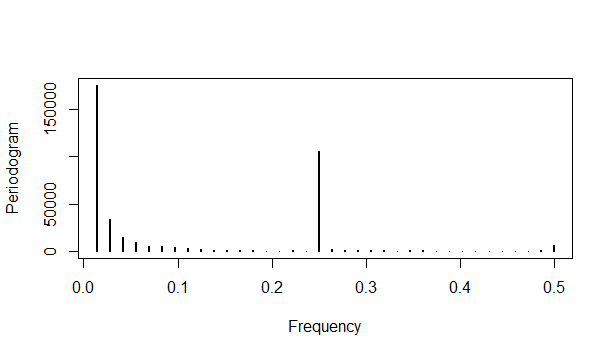
The dataset contains Beer Sales for the last 72 Quarters. Using two different Time Series Models: Exponential & ARIMA, predicting the next 8 Quarter Sales for the Beer.

1. **Reading & Plotting the Data**

The beer data consist of quarterly Sales Data for 72 Quarters. The plot of the data has both Trend and Seasonality factor to it



1. **Detecting Seasonality**

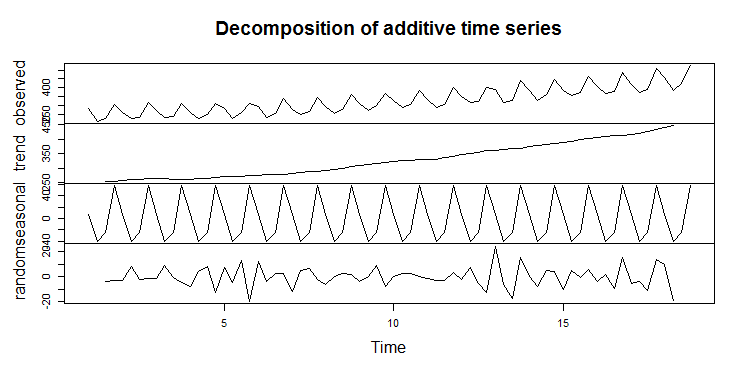
Creating a **Periodogram** plot shows the following two things:

1. The signal is the strongest in the beginning and then it is strongest at a frequency of 0.25.
2. The relationship between periodicity and frequency is given by Periodicity = 1/Frequency Periodicity = 1/.25 = 4
3. The highest signal in the periodogram occurred at a frequency of 0.01388889 which correspond to 72 quarters which is the entire data set!
4. The next highest signal is detected at a frequency of 0.25, which corresponds to 4 quarters. This gives a periodicity of 4 quarters
5. The Australian beer production clearly follows an annual seasonality and records data quarterly i.e. 4 times in a year

# Inference: The time series has a linear trend with a constant (additive) seasonal pattern as

**Shown in visual plot above and proven mathematically.**

1. **Decomposing Time Series Data**



**Holt-Winters Time Series Model**

The Holt- Winters Exponential Method is applied to the data with seasonal component being “Additive” since the seasonality is constant with time

**The Model has an AIC value of 653.67**

> beer.model1 <- hw(beer.data, seasonal = "additive")

> beer.model1$model

Holt-Winters' additive method

Call:

hw(y = beer.data, seasonal = "additive")

Smoothing parameters:

alpha = 0.0395

beta = 0.0395

gamma = 0.1854

Initial states:

l = 258.2137

b = 0.0239

s=55.077 -23.9983 -37.4121 6.3335

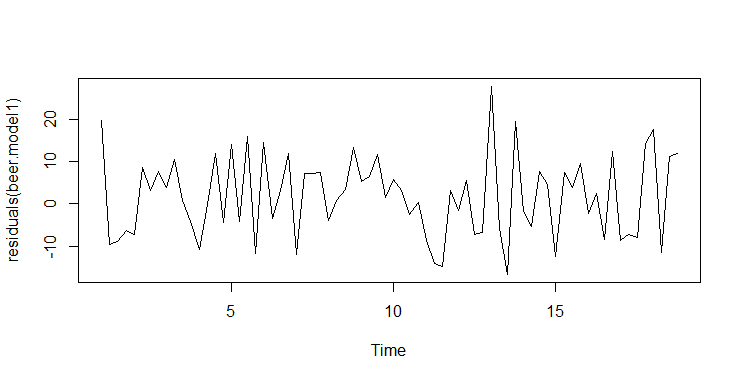
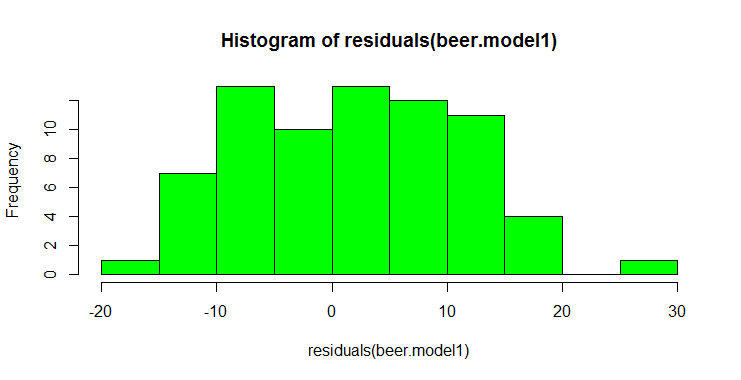
sigma: 9.738

AIC AICc BIC

653.6687 656.5719 674.1587

**Check for Model Residuals**

* Residuals are plotted to see that all the information from the model has been harnessed and all the remains is **White Noise N(0, sigma^2)**
* From the plots below, it is evident that thevariance is constant and the mean is approx. 0, signifying that all that remains from the residuals is White Noise and that the model has been fit properly to the data



* **Ljung Box Test:** The test is used to check if data is randomly distributed.

**Ho <- Random Distribution of the data**

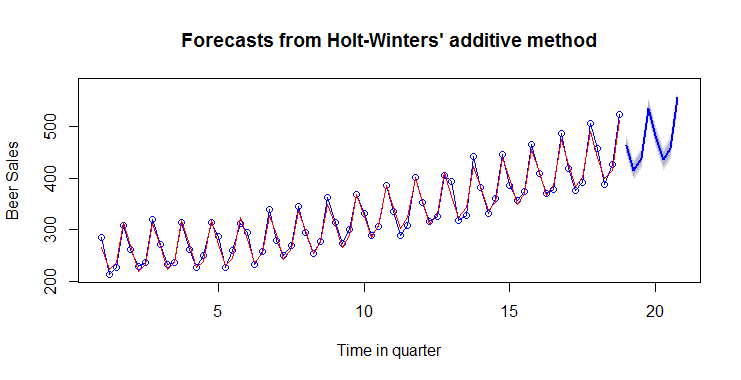
**Ha <- Data not distributed randomly**

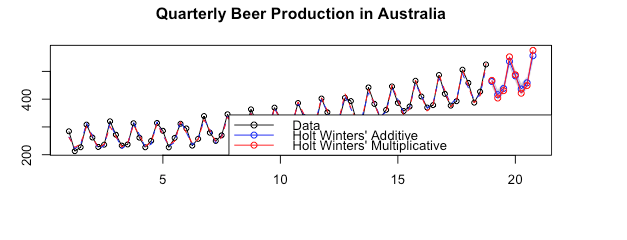
> Box.test(residuals(beer.model1), lag = 4, type = "Ljung")

Box-Ljung test

data: residuals(beer.model1)

X-squared = 3.3368, df = 4, p-value = 0.5031

* The result of the test with p>0.5 signifies that the residuals are randomly distributed and hence, the model is properly fitted.
* Multiplicate and aditive model result sin similar kind of forecast.

****

**ARIMA Model**

1. **Check for Data Stationarity**

As see above from the time series plot, data is not stationary and has both Trend and Seasonality component. To make the data stationary, we Difference the timeseries data.

1. **KPSS Test**

The test is used to test the Hypothesis for data stationarity

**Ho <- Data is not Stationary**

**Ha <- Data is not Stationary**

1. The results below with p value of 0.01 signify that the Hypothesis of Data being Stationary is rejected.

> kpss.test(beer.data)

KPSS Test for Level Stationarity

data: beer.data

KPSS Level = 3.0456, Truncation lag parameter = 1, p-value = 0.01

1. **Find Order of differencing to make data Stationary**

We use the below commands to check for no. of times differencing is required to remove the seasonal and the trend component, thereby making the data stationary

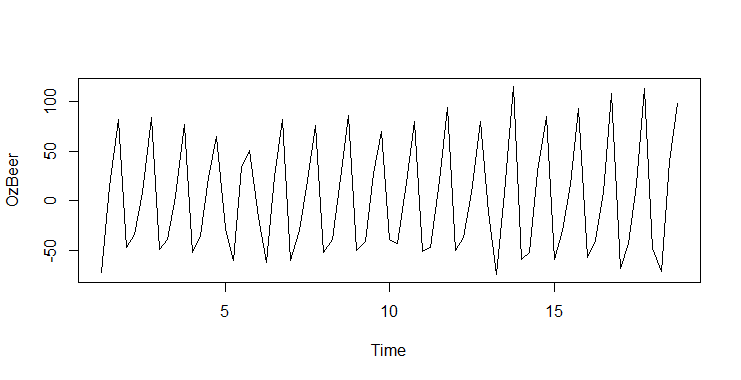
> nsdiffs(beer.data) #differencing required to remove seasonality

[1] 1

> ndiffs(beer.data) #differencing required to remove trend

[1] 1

**Step – 1 Removing Trend Component**

> d.beer <- diff(beer.data)

> plot(d.beer)

> kpss.test(d.beer)

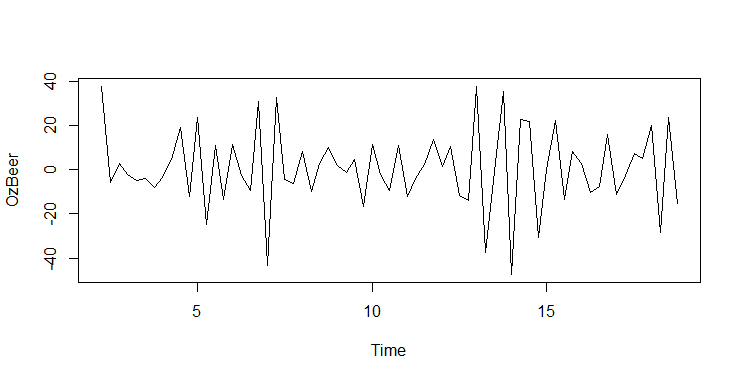
KPSS Test for Level Stationarity

data: d.beer

KPSS Level = 0.036853, Truncation lag

parameter = 1, p-value = 0.1

* As can be seen from the results and the graph, the data has been detrended, but seasonality component remain in the data

**Step – 2 Removing Seasonal Components**

> dd.beer <- diff(d.beer, lag = 4)

> plot(dd.beer)

> kpss.test(dd.beer)

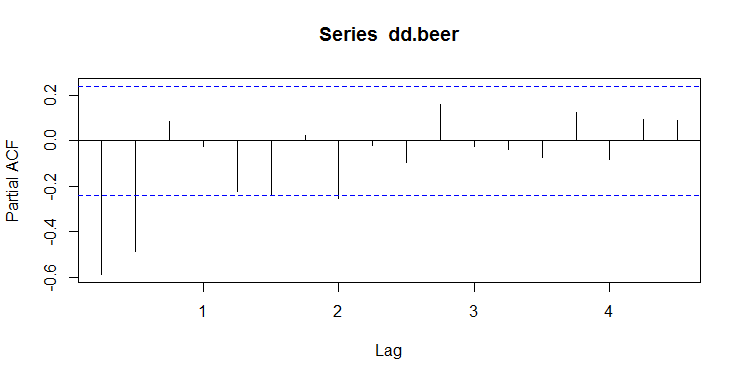
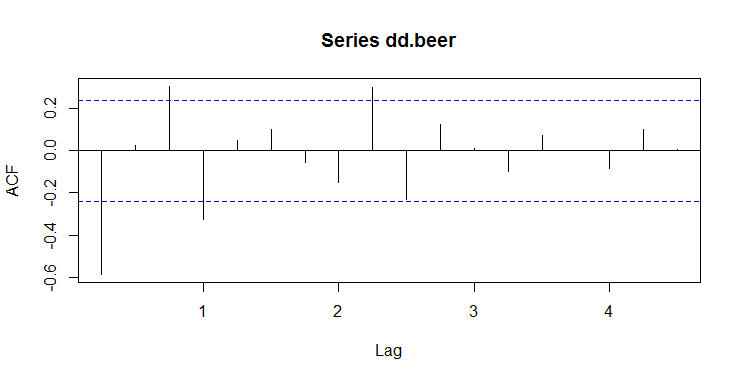
KPSS Test for Level Stationarity

data: dd.beer

KPSS Level = 0.039765, Truncation lag

parameter = 1, p-value = 0.1

* As can be seen from the results and the graph, the data has been detrended and seasonality component removed. The data is now stationary
* **Plotting ACF and PACF Plots to determine the Seasonal parameters of the ARIMA Model**



* **Non-Seasonal Components:** PACF Cuts of Sharply after lag 2. Suggesting **ARIMA(2,1,0)**
* **Seasonal Components:** ACF plot shows negative at lag 4 (period of the series). This suggest

**ARIMA(0,1,1)**

1. **Build ARIMA Model**

* The model has an AIC value of 515.29**.**

> beer.model.arima <- arima(beer.data, order = c(2,1,0),

seasonal = list(order = c(0,1,1), period = 4))

> beer.model.arima

Call:

arima(x = beer.data, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 1),

period = 4))

Coefficients:

ar1 ar2 sma1

-0.9111 -0.5086 -0.5868

s.e. 0.1146 0.1121 0.1208

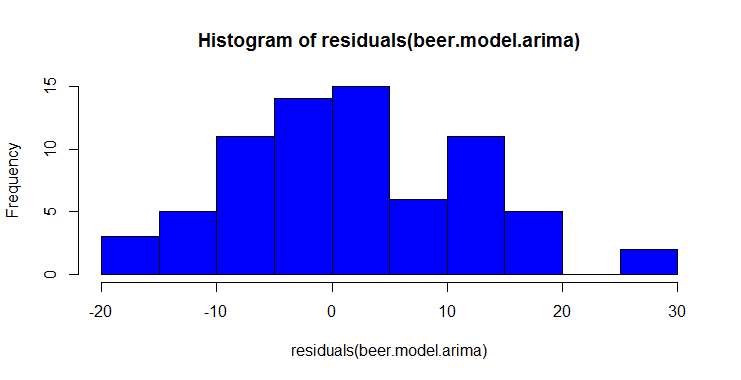
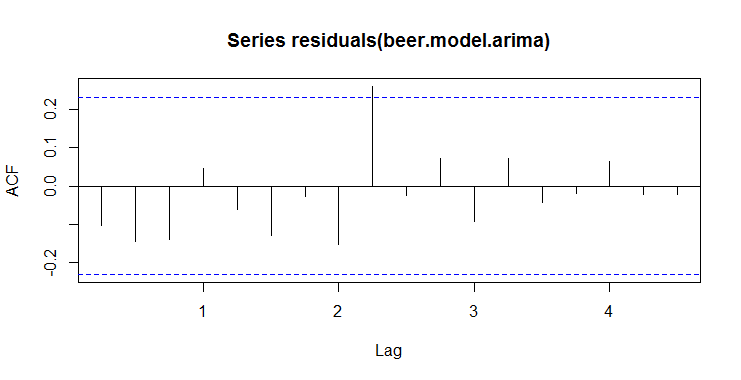
sigma^2 estimated as 111.6: log likelihood = -254.65, aic = 515.29

1. **Model Validation**

* The ACF plot below shows that the residuals are independently distributed. The Histogram on the right confirms that the residuals only contain **White Noise with N(0, Sigma^2)**
* The Ljung Box test also has p-value of 0.5 signifying that the Null Hypothesis is not rejected which states that the Residuals are independently distributed
* Based on the plots and the graphs, we can say that the model is a good fit and will be able to predict accurate results

> acf(residuals(beer.model.arima))

> hist(residuals(beer.model.arima))



> Box.test(residuals(beer.model.arima), lag = 4, type = "Ljung")

Box-Ljung test

data: residuals(beer.model.arima)

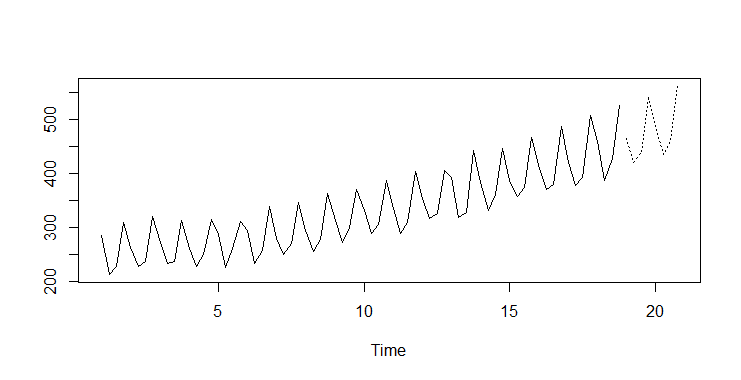
X-squared = 4.042, df = 4, p-value = 0.4004

**NOTE: Running the auto.arima function in R gave the parameters as (2,1,1) (0,1,1) and an AIC value of 515.14. There is only slight improvement in the AIC value with the automatic parameters selection. We have chosen to go ahead with the model we build and neglect the auto.arima model.**

1. **Prediction**

> beer.forecast <- predict(beer.model.arima, n.ahead = 8)

> ts.plot(beer.data, beer.forecast$pred, lty = c(1,3))



**Final Summary**

On comparing the results of the two models used to predict the next 8 quarters of Beer sales, Seasonal ARIMA Model has done far better in the accuracy of predicting the results with the AIC value of 515 as compared to the AIC value 653 in the holtz-Winters method model.

Hence, we conclude that ARIMA model has proven to be the more powerful and accurate model in the timeseries prediction.