**Business Forecasting**

Last name: Bhatnagar First name: Meru Section number (40 or 60) 40

Forecaster toolbox

**Step 1** Create a directory on your computer where the home work will be implemented. That means all input files should be placed there and R will create the output files in this directory.

**Step 2** Open R-studio. In the command prompt select “Session” then “Set working directory” and “Choose directory”. Using the command prompt with file manager select the directory you created in Step 1.



**Step 3** Install fpp library 9if it is not installed yet, and connect this library by command library(fpp).



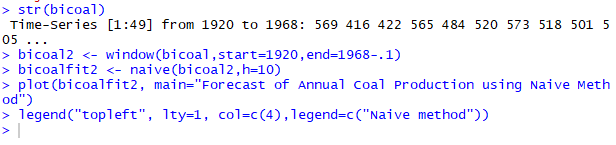
**Problem set 1** For each of the following series, make a graph of the data with forecasts using the most appropriate of the four benchmark methods:

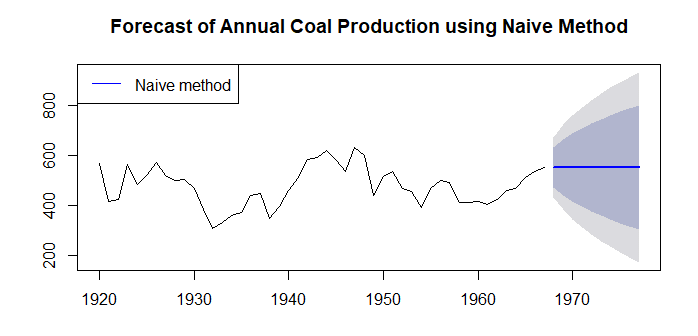
mean, naive, seasonal naive or drift.

Provide plots for time series. Implement the autocorrelation analysis including correlogram and the Ljung-Box test with lag 8. Answer the question if autocorrelation, and seasonality exists.

1. Annual bituminous coal production (1920–1968). Data set bicoal. Use naïve. (1 mark)

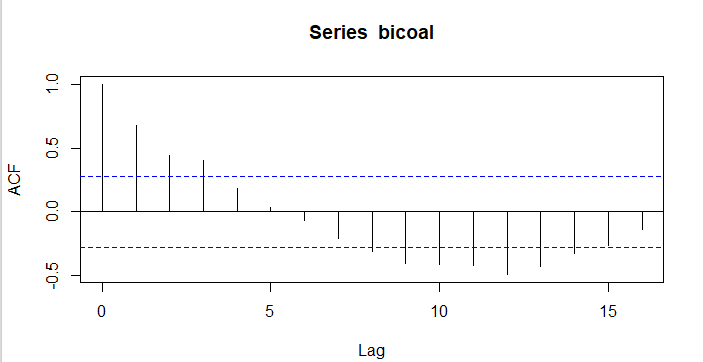
**Naïve Method**





**Auto- Correlation:**





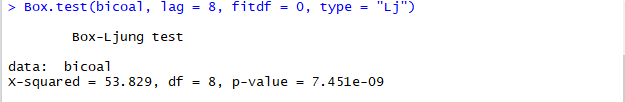
Since the spikes is above the dotted blue (sqrt(2)/T) outside the range of +.28 hence it can be confirmed that the time series dataset is auto correlated.

**Seasonality:**



Since there is no uniform trend available for this dataset , hence seasonality does not exist.

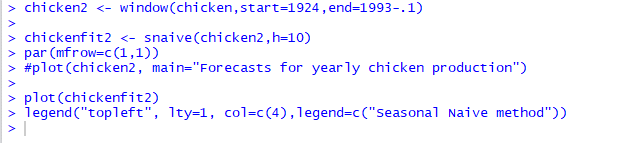
**Box-Ljung Test**

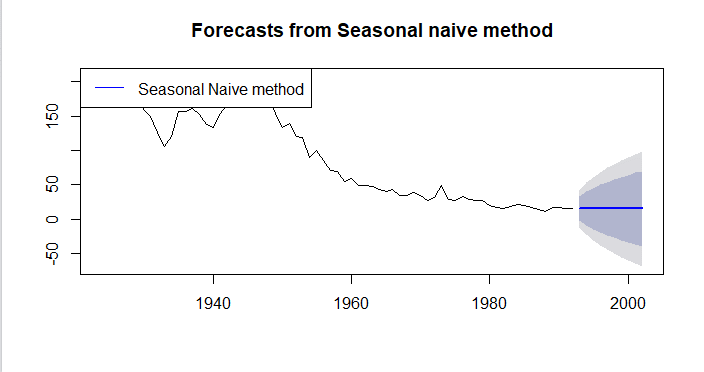


Since the p-value of time series is less than .05, hence we can reject the null hypothesis and conclude that the model is significant for forecasting.

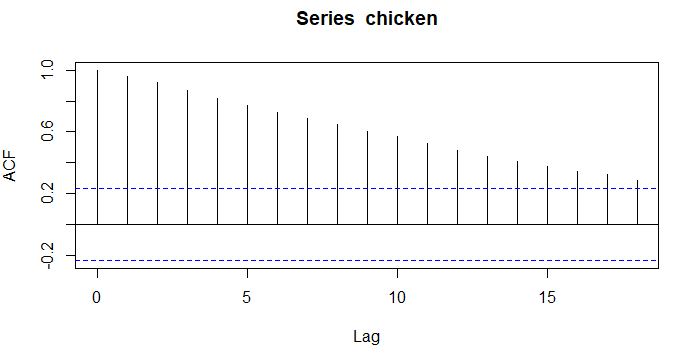
1. Price of chicken (1924–1993). Data set chicken. Use seasonal naïve. (1 mark)

**Seasonal Naïve Method**





**Auto Correlation:**



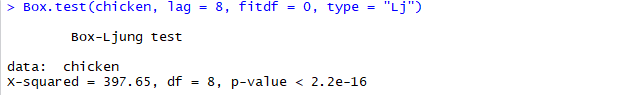
Since the spikes is above the dotted blue (sqrt(2)/T) outside the range of +.28 hence it can be confirmed that the time series dataset is auto correlated.

**Seasonality:**



Since there is no uniform trend available for this dataset, hence seasonality does not exist.

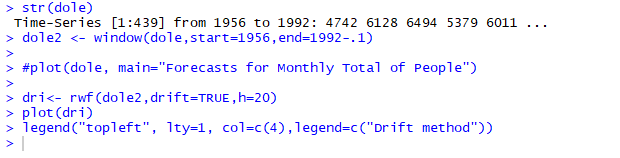
**Box-Ljung Test:**

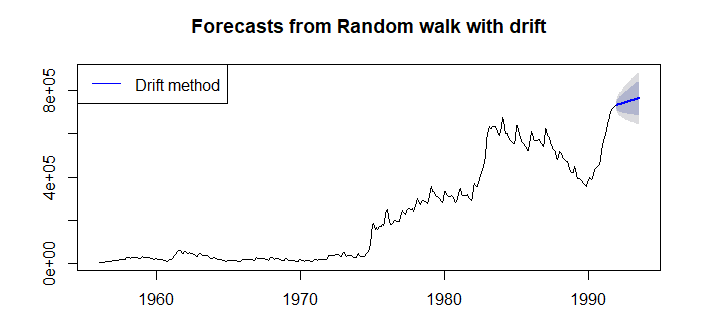


Since the p-value of timeseries is less than .05, hence we can reject the null hypothesis and conclude that the model is significant for forecasting.

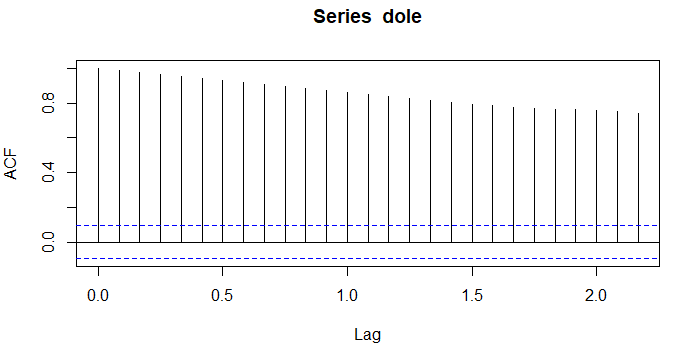
1. Monthly total of people on unemployed benefits in Australia (January 1956–July 1992). Data set dole. Use drift (1 mark)

**Drift Method**





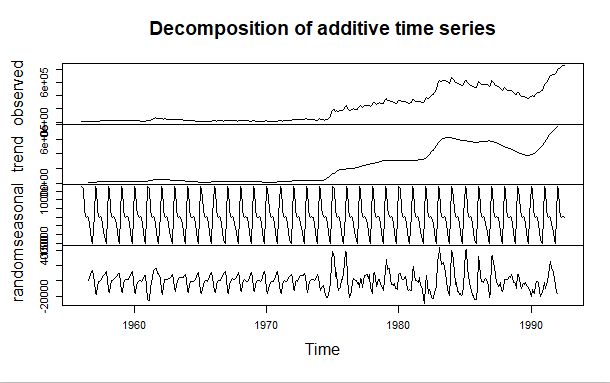




Since the spikes is above the dotted blue (sqrt(2)/T) outside the range of +.28 hence it can be confirmed that the time series dataset is auto correlated.

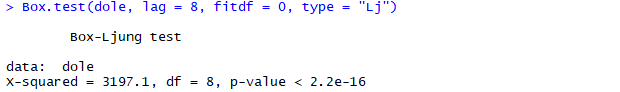
**Seasonality:**





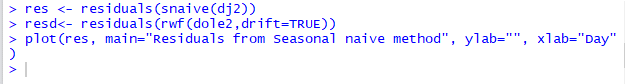
Seasonality exists for the given dataset with increasing trend.

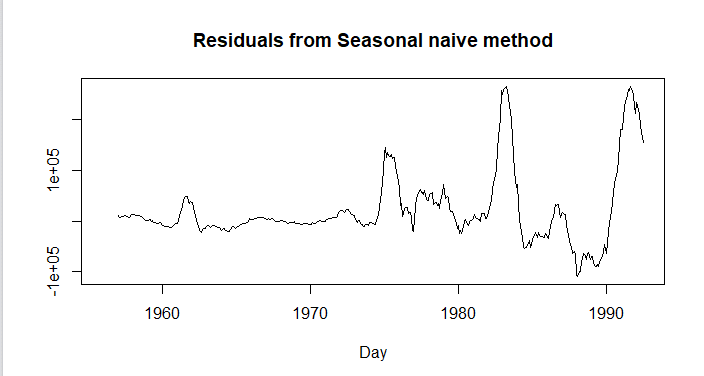
**Box-Ljung Test**



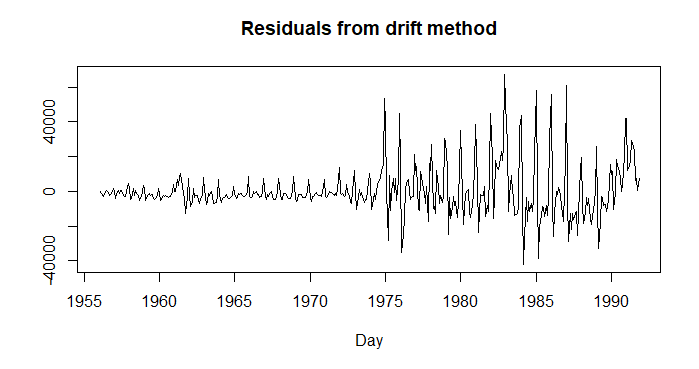
since the p-value of time series is less than .05, hence we can reject the null hypothesis and conclude that the model is significant for forecasting.

(d) Plot residuals for this forecast and compare them with the residuals of seasonal naive forecast. (1 mark, no partial marks)



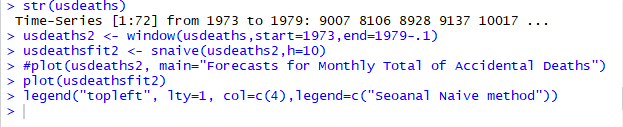


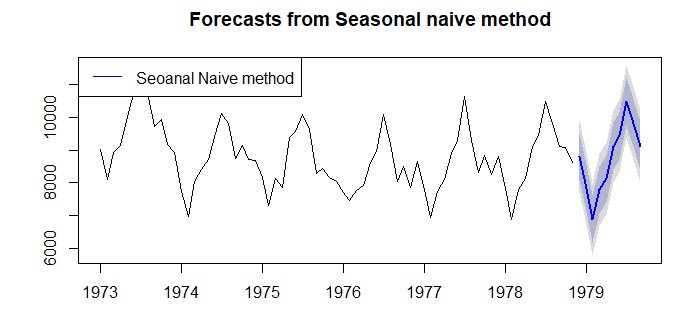




1. Monthly total of accidental deaths in the United States (January 1973–December 1978). Data set usdeaths. Use seasonal naïve. (1 mark)

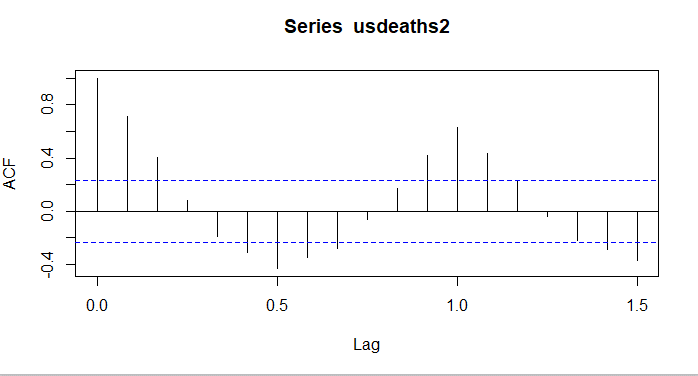
**Seasonal Naïve:**





**Auto Correlation:**

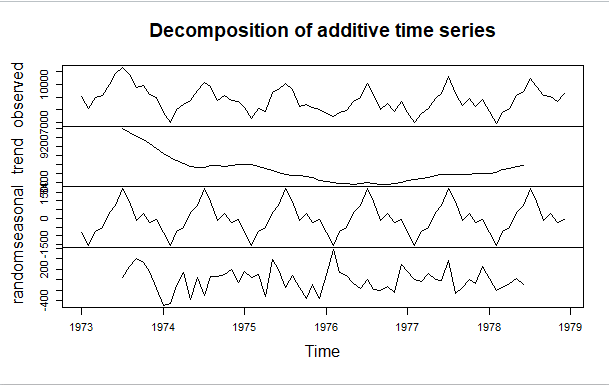




Since the spikes is above the dotted blue (sqrt(2)/T) outside the range of +.28 hence it can be confirmed that the time series dataset is auto correlated.

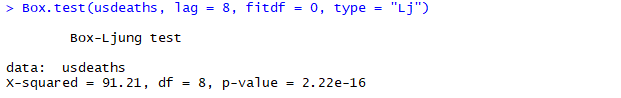
**Seasonality:**





The seasonality exists for given dataset with decreasing trend which is unevenly distributed.

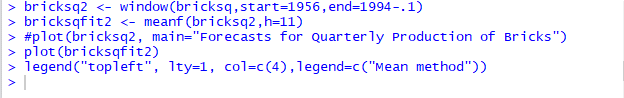
**Box-Ljung Test:**

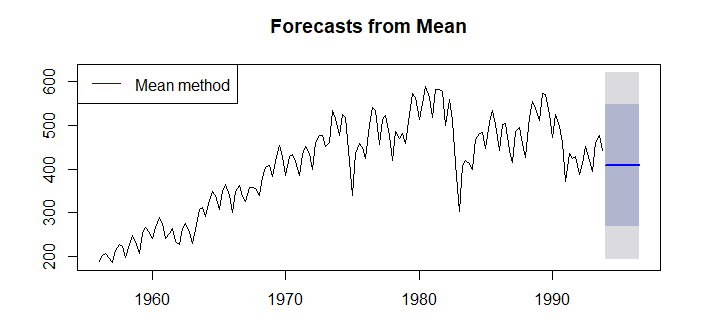


Since the p-value of time series is less than .05, hence we can reject the null hypothesis and conclude that the model is significant for forecasting.

1. Quarterly production of bricks (in millions of units) at Portland, Australia (March 1956–September 1994). Data set bricksq. Use mean method. (1 mark)

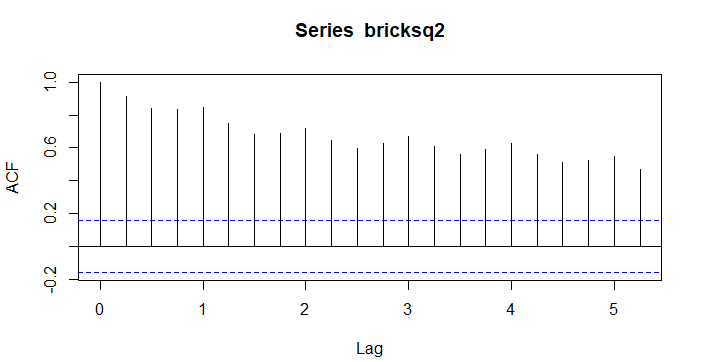
**Mean Method**





**Auto Correlation:**

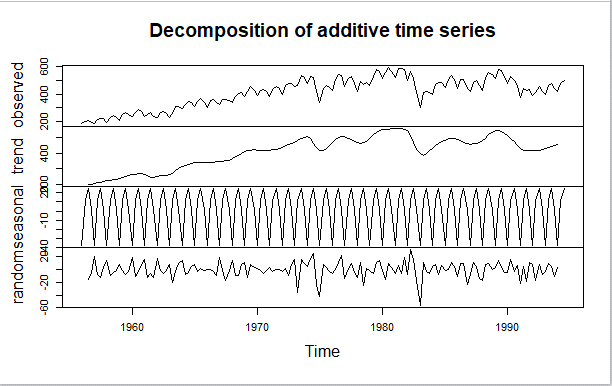




Since the spikes is above the dotted blue (sqrt(2)/T) outside the range of +.28 hence it can be confirmed that the time series dataset is auto correlated.

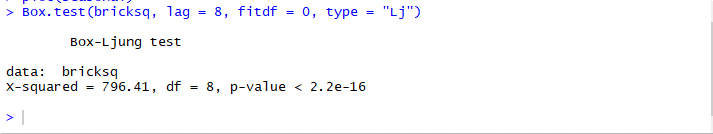
**Seasonality:**





The seasonality exists for given dataset with increasing trend which is unevenly distributed.

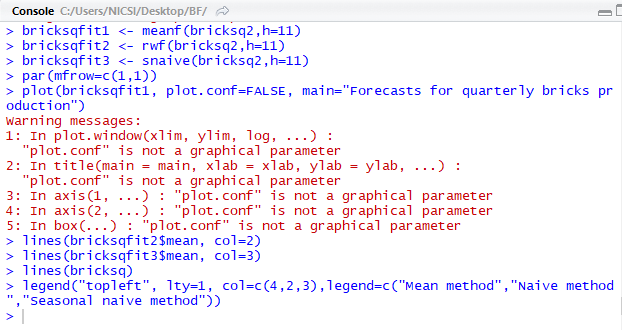
**Ljung-Box Test:**

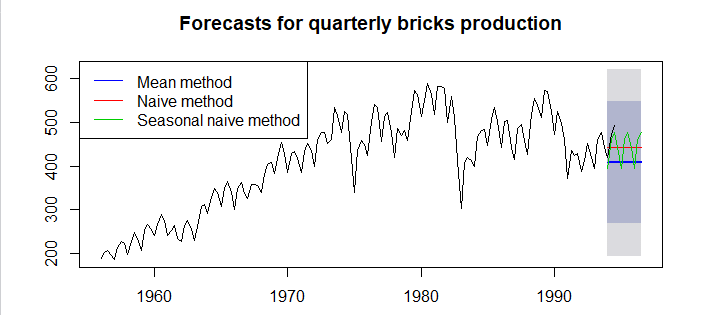


Since the p-value of time series is less than .05, hence we can reject the null hypothesis and conclude that the model is significant for forecasting.

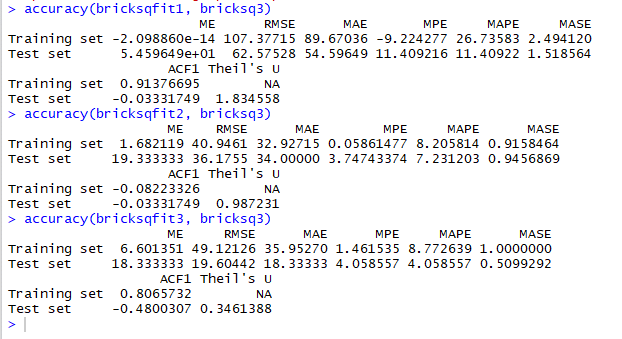
1. What is the more appropriate method of forecast for this data? Implement it and justify your answer. (1 mark, no partial marks)

**Optimized Method Determination:**





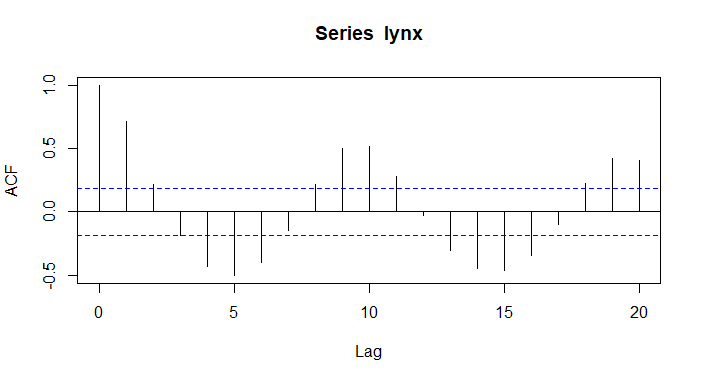




Since the MASE and MAPE is less bricksq3 hence we can conclude that seasonal naïve method is the most appropriate method to forecast the data.

(h) Annual Canadian lynx trappings (1821–1934). Data set lynx. Choose the optimal method. Implement for it the same autocorrelation analysis as above. (1 mark)

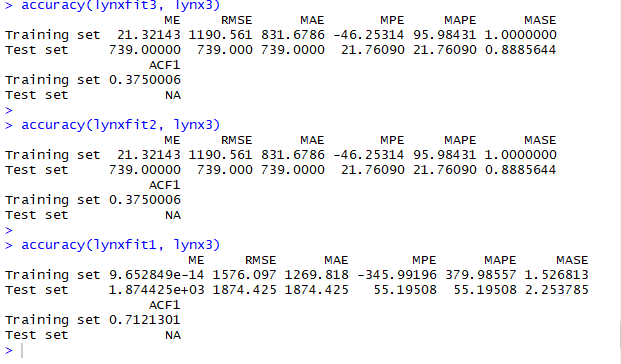
**Auto Correlation**



We can see that there is a cyclic trend which keeps on having a positive and negative trend.

**Optimised Method determination:**





Since the MASE and MAPE is less lynxfit3 hence we can conclude that seasonal naïve method is the most appropriate method to forecast the data.

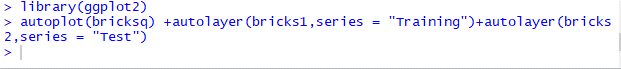
**Problem set 2** For the data set bricksq:

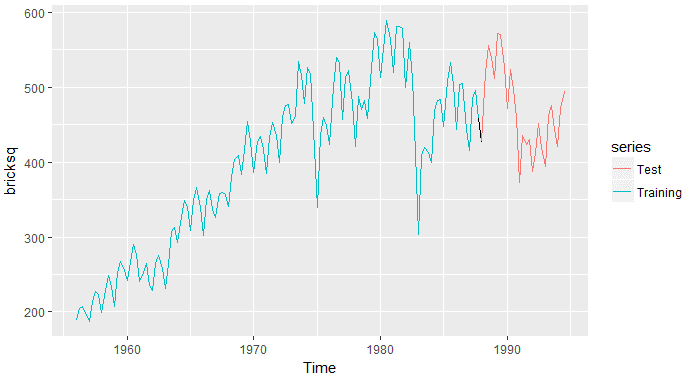
(a) Split the data into two parts using

> bricks1 <- window(bricksq, end=1987.99)

> bricks2 <- window(bricksq, start=1988)

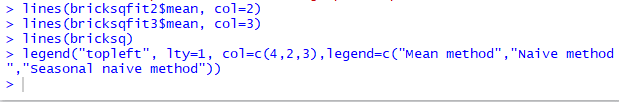
(b) Check that your data have been split appropriately by producing the following plot. (0.5 marks)

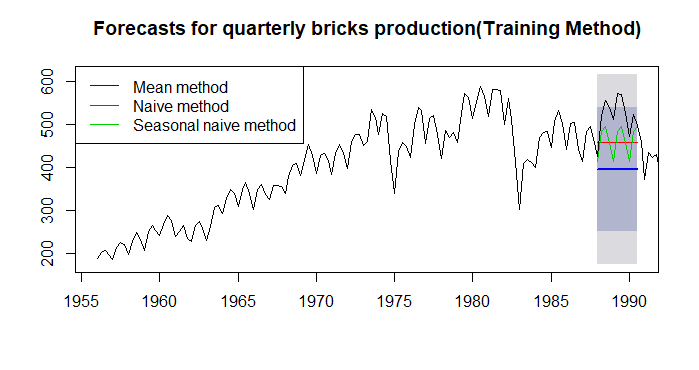


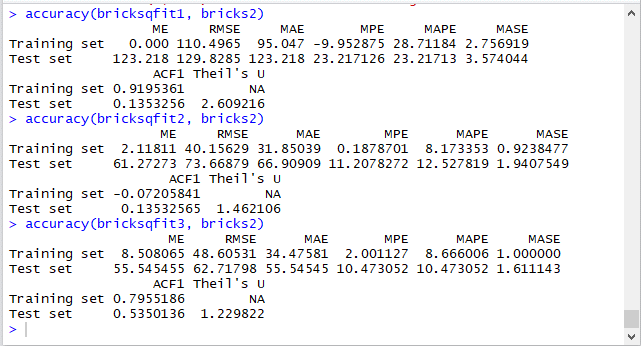


(c) Calculate forecasts using each of the mean, naïve, and seasonal naive benchmark methods applied to bricks1. Compare the accuracy of your forecasts against the actual values stored in bricks2. Which method does best? Why? (1 mark)

**Optimized Model Determination:**

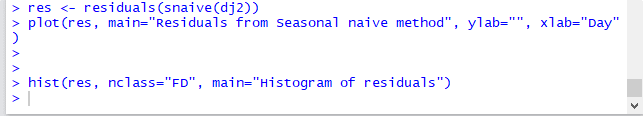


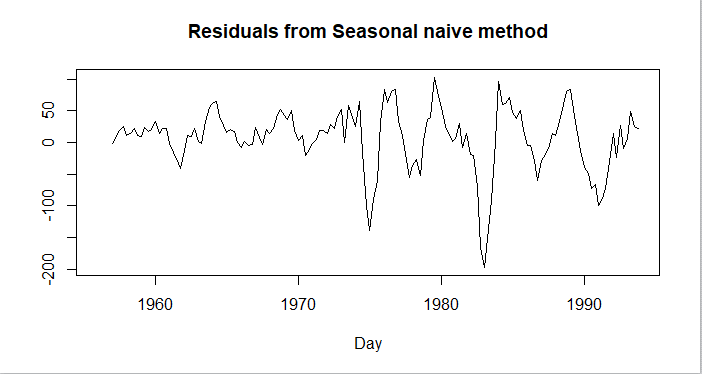


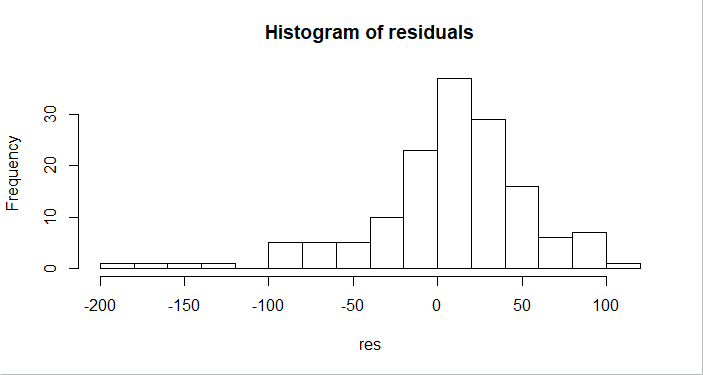


By comparing all the models with the actual values stored in brick2 we can conclude that seasonal naïve is the best method for forecasting this series.

(d) For the best method, compute the residuals and plot them. (0.5 marks)







Since the residual is not normally distributed hence we can say that seasonal naïve model is a good model of forecasting for this dataset.