Business Forecasting - Assignment 01

Ashita Shetty

2023-10-23

```
library(datasets)
library(fpp2)
```

```
## Warning: package 'fpp2' was built under R version 4.1.3
```

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

```
## -- Attaching packages -------------------------------------- fpp2 2.5 --
```

```
## v ggplot2   3.4.3      v fma        2.5
## v forecast  8.21       v expsmooth 2.3
```

```
## Warning: package 'forecast' was built under R version 4.1.3
```

```
## Warning: package 'fma' was built under R version 4.1.3
```

```
## Warning: package 'expsmooth' was built under R version 4.1.3
```
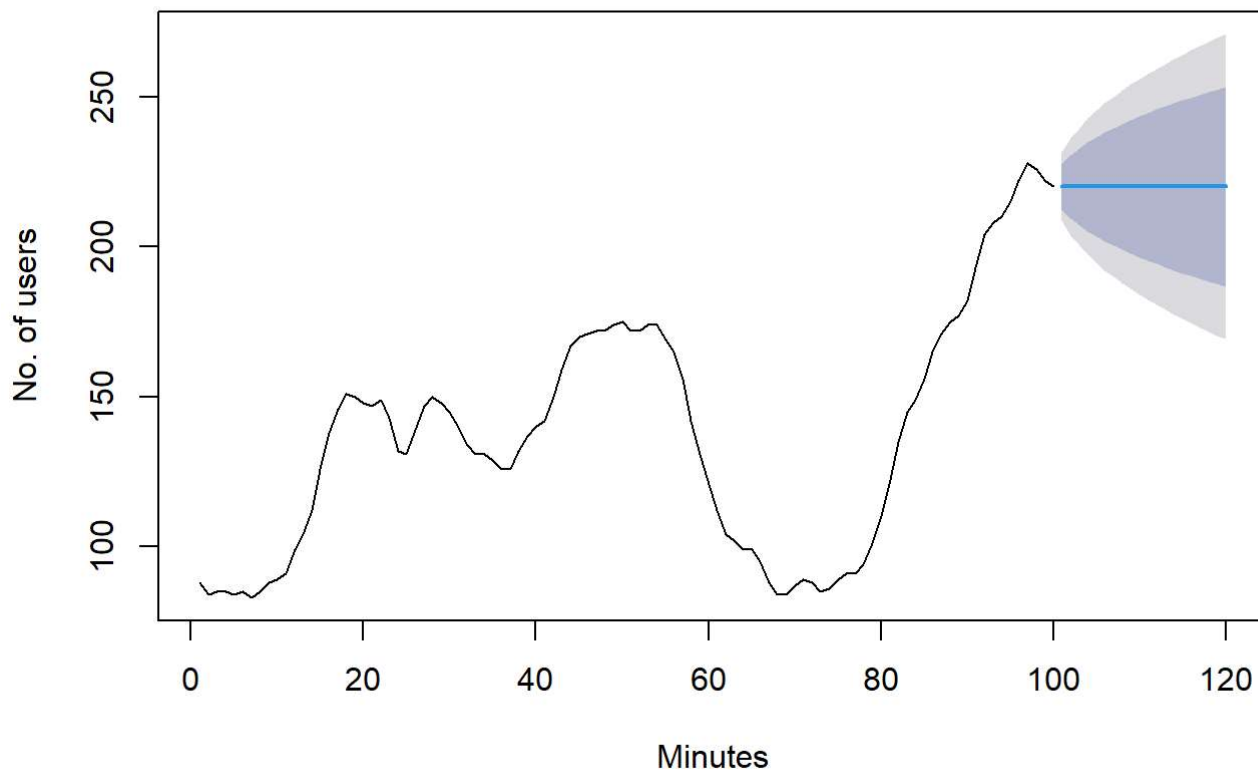
```
##
```

Exercise 1

WWWusage

(i) Plot the time series along with a forecast for next 20 minutes using the naive method

```
n_www <- naive(WWWusage, h = 20)
plot(n_www, main = "Internet usage time series and forecast", ylab = "No. of users", xlab="Minut
es")
```
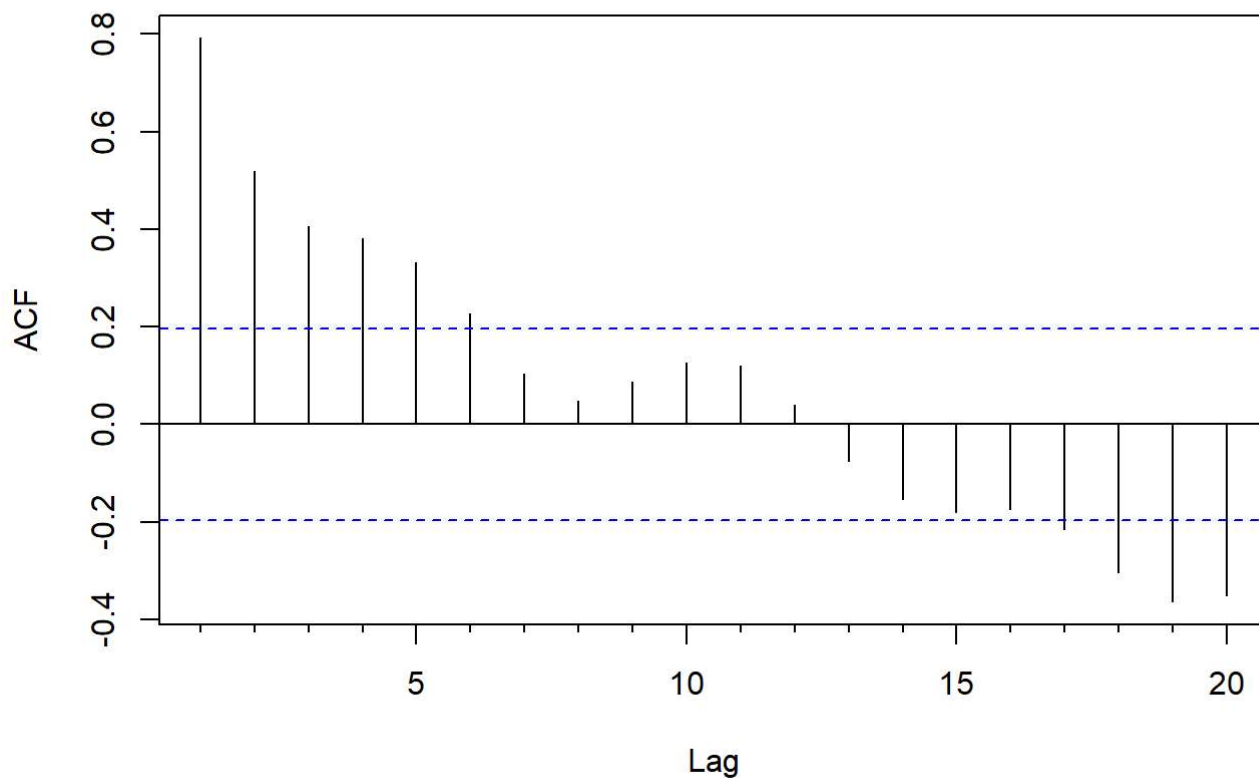
Internet usage time series and forecast

(ii) Plot a correlogram of residuals

```
Acf(n_www$residuals, main = "Correlogram of Residuals")
```

Correlogram of Residuals

(iii) Perform a Ljung-Box test with lag 8 on residuals

```
ljung <- Box.test(n_www$residuals, type = "Ljung-Box", lag = 8)
print(ljung)
```

```
##
##   Box-Ljung test
##
## data:  n_www$residuals
## X-squared = 142.99, df = 8, p-value < 2.2e-16
```

```
accuracy(n_www)
```

```
##                        ME     RMSE      MAE       MPE     MAPE MASE      ACF1
## Training set 1.333333 5.799687 4.525253 0.8286134 3.401706    1 0.7917635
```

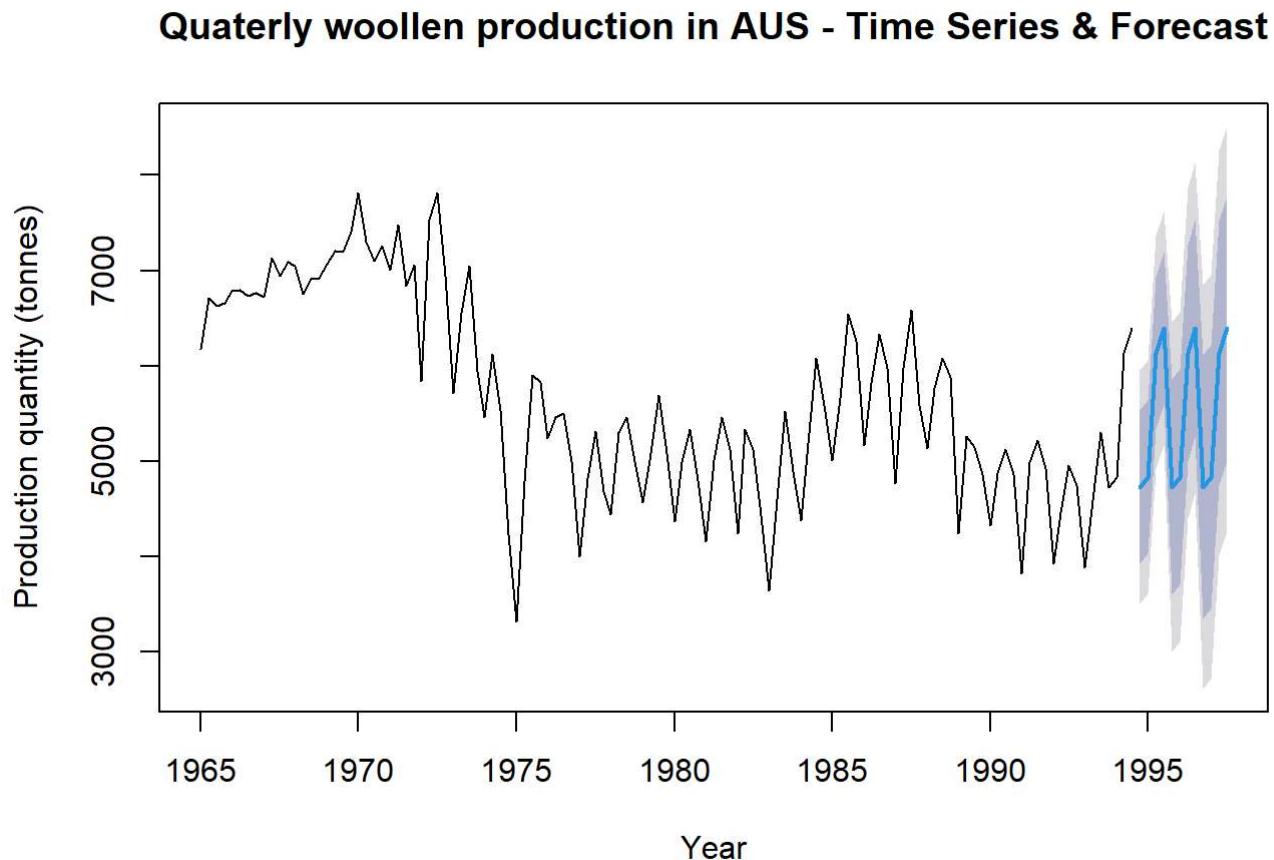### (iv) Use parts (ii) and (iii) to comment on whether there is autocorrelation or seasonality

- In the Acf plot, it is visible that the lines cross the confidence interval levels and it crosses on both sides of the Acf plot.This suggests that there is seasonality in the data.
- In the Ljung box test, the p < 0.05 suggests that the autocorrelations in the residuals are statistically significant and therefore, the hypothesis that there is autocorrelation in the residuals can be rejected.

Exercise 2

woolyrnq

(i) Plot the time series along with a forecast for the next 3 years using the seasonal naïve method
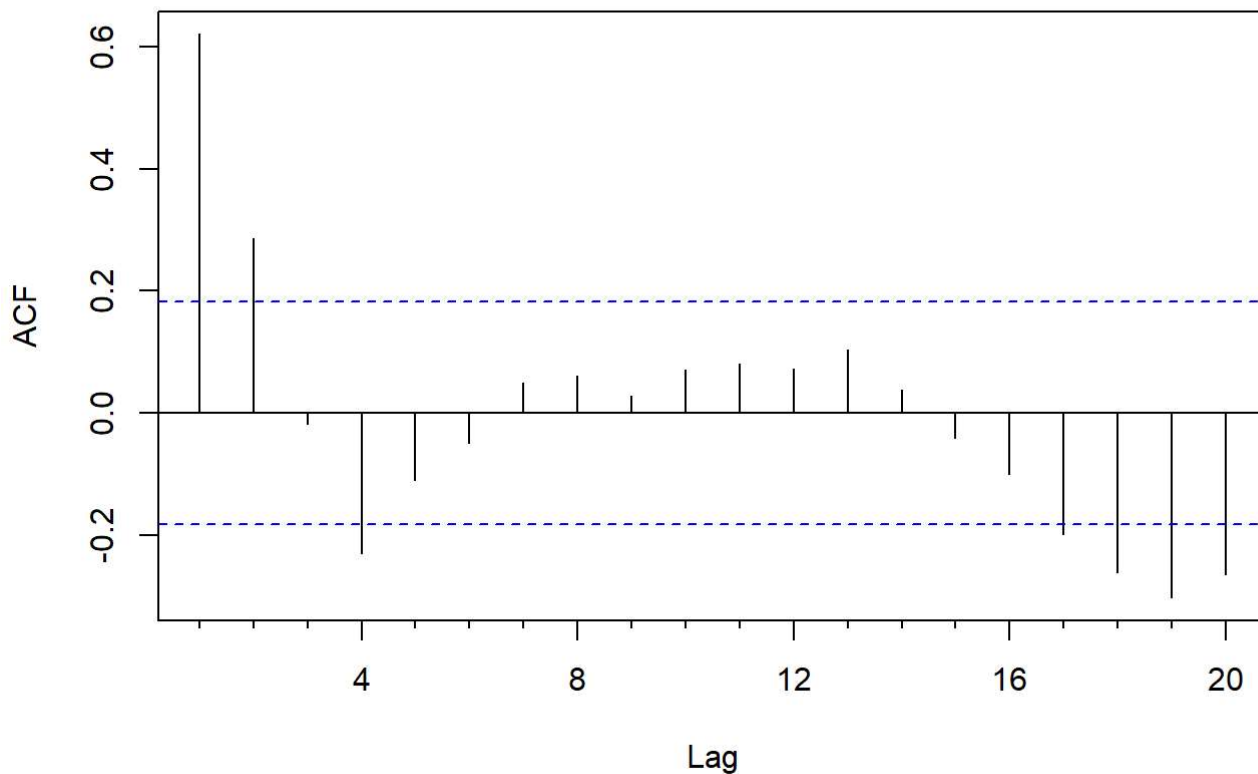
```
snaive_woolyrnq <- snaive(woolyrnq, h = 12)
plot(snaive_woolyrnq, main = "Quaterly woollen production in AUS - Time Series & Forecast", ylab
= "Production quantity (tonnes)", xlab="Year")
```

Quaterly woollen production in AUS - Time Series & Forecast

(ii) Plot a correlogram of the residuals

```
Acf(snaive_woolyrnq$residuals, main = "Correlogram of Residuals")
```

Correlogram of Residuals

(iii) Perform a Ljung-Box test with lag 8 on the residuals

```
ljung <- Box.test(snaive_woolyrnq$residuals, type = "Ljung-Box", lag = 8)
print(ljung)
```

```
##
##  Box-Ljung test
##
## data:  snaive_woolyrnq$residuals
## X-squared = 64.299, df = 8, p-value = 6.641e-11
```

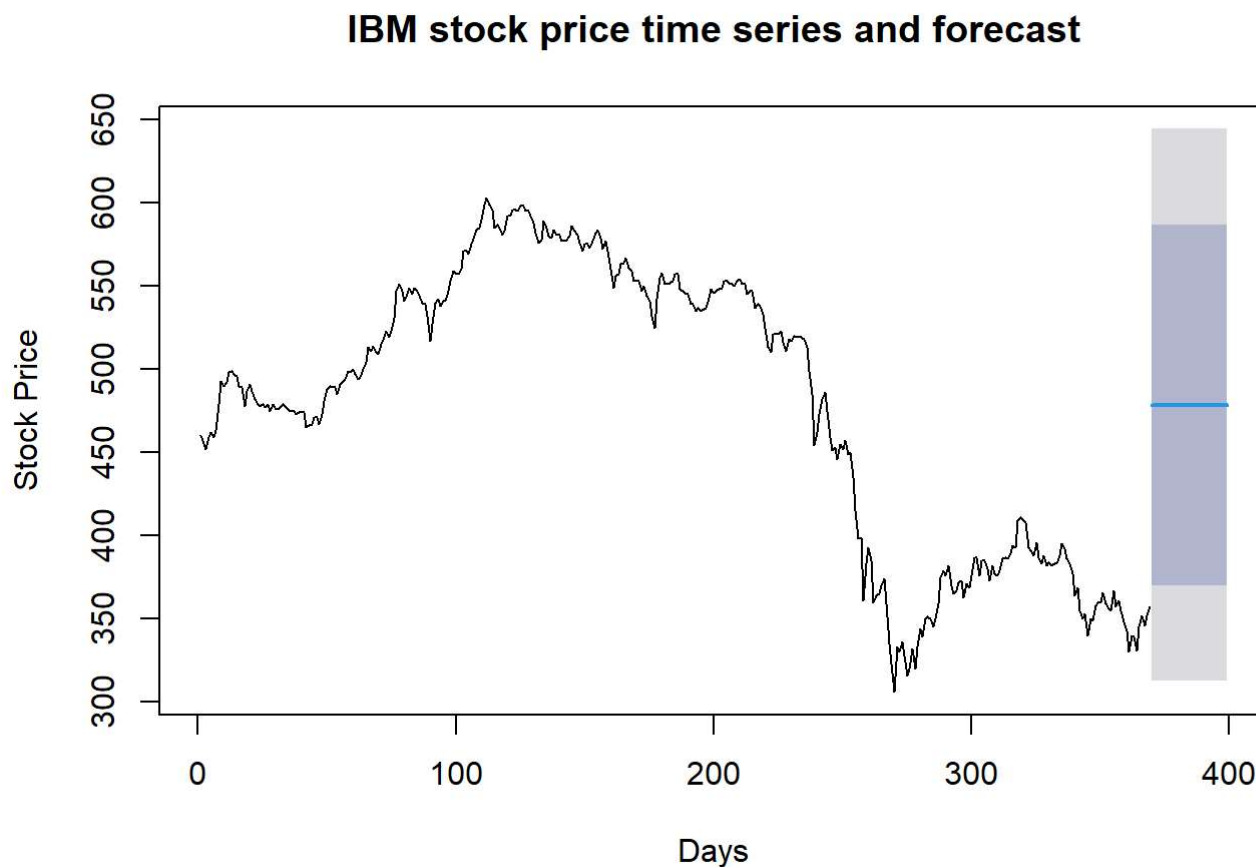### (iv) Use parts (ii) and (iii) to comment on whether there is autocorrelation or seasonality

- In the Acf plot, it is visible that the lines cross the confidence interval levels and it crosses on both sides of the Acf plot.This suggests that there is seasonality in the data.
- In the Ljung box test, the $p < 0.05$ suggests that the autocorrelations in the residuals are statistically significant and therefore, the hypothesis that there is no autocorrelation in the residuals can be rejected.

Exercise 3

ibmclose

(i) Plot the time series along with a forecast for the next 30 days using the mean method.

```
mean_ibmclose <- meanf(ibmclose, h = 30)
plot(mean_ibmclose, main = "IBM stock price time series and forecast", ylab = "Stock Price", xla
b="Days")
```

IBM stock price time series and forecast

(ii) Plot a correlogram of the residuals.

```
Acf(mean_ibmclose$residuals, main = "Correlogram of Residuals")
```

Correlogram of Residuals
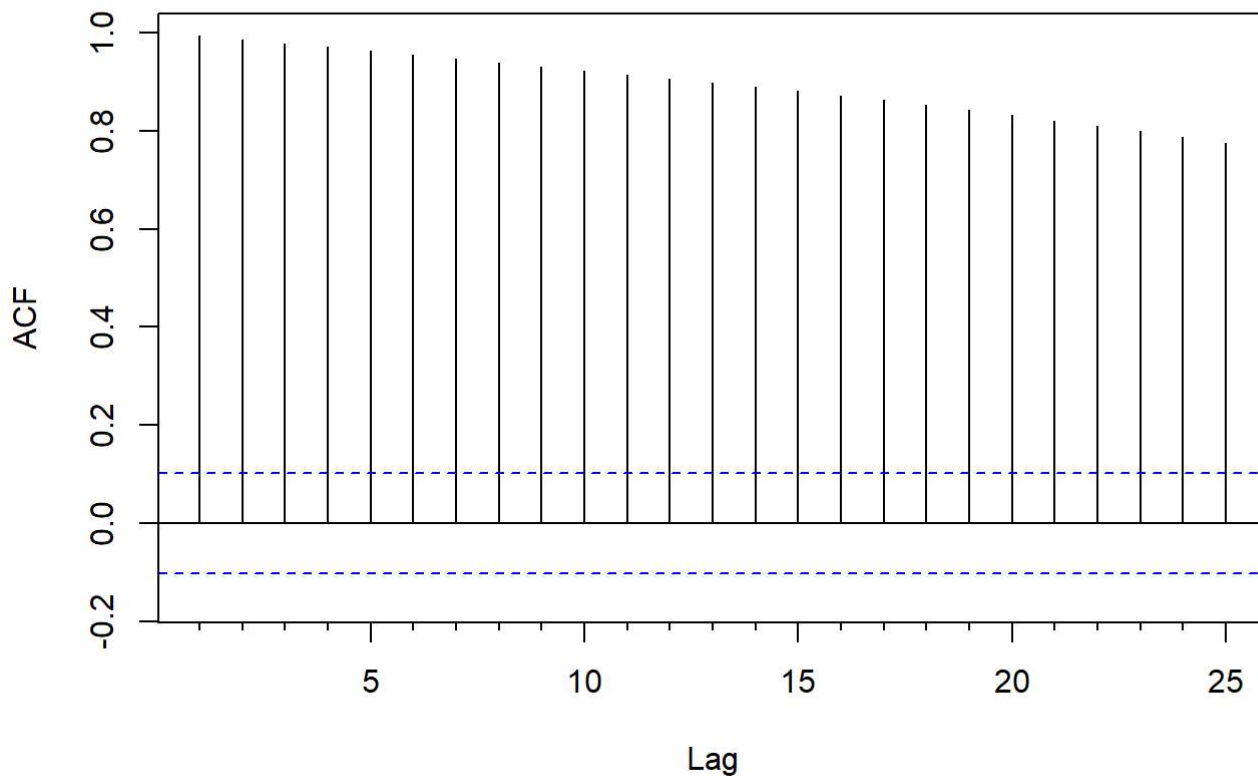
(iii) Perform a Ljung-Box test with lag 8 on the residuals.

```
ljung <- Box.test(mean_ibmclose$residuals, type = "Ljung-Box", lag = 8)
print(ljung)
```

```
##
##   Box-Ljung test
##
## data:  mean_ibmclose$residuals
## X-squared = 2809.9, df = 8, p-value < 2.2e-16
```

(iv) Use parts (ii) and (iii) to comment on whether there is autocorrelation or seasonality

- Based on the ACF plot, it can be inferred that the data follows a particular trend and since, it crosses the confidence level - there is autocorrelation.
- In the Ljung box test, the $p < 0.05$ suggests that the autocorrelations in the residuals are statistically significant and therefore, the hypothesis that there is no autocorrelation in the residuals can be rejected.

(v) What might be a more appropriate method of forecast for this time series?

- *Naive or Seasonal Naive* would be a more appropriate method of forecast for this time series.
- In the code provided below, it can be observed that the MAPE for Mean and Seasonal Naive/Naive have a significant difference, making the latter a better choice of forecasting method.
- The MAPE and other error metric values are the same for Naive and Seasonal Naive because the data provided is for each day and therefore, the 'Season' here is the 'Day'.

(vi) Implement your chosen method from part (v) and investigate whether it is indeed a better choice than the mean method

```
n_ibmclose = naive(ibmclose, h=30)
sn_ibmclose = snaive(ibmclose, h=30)
```
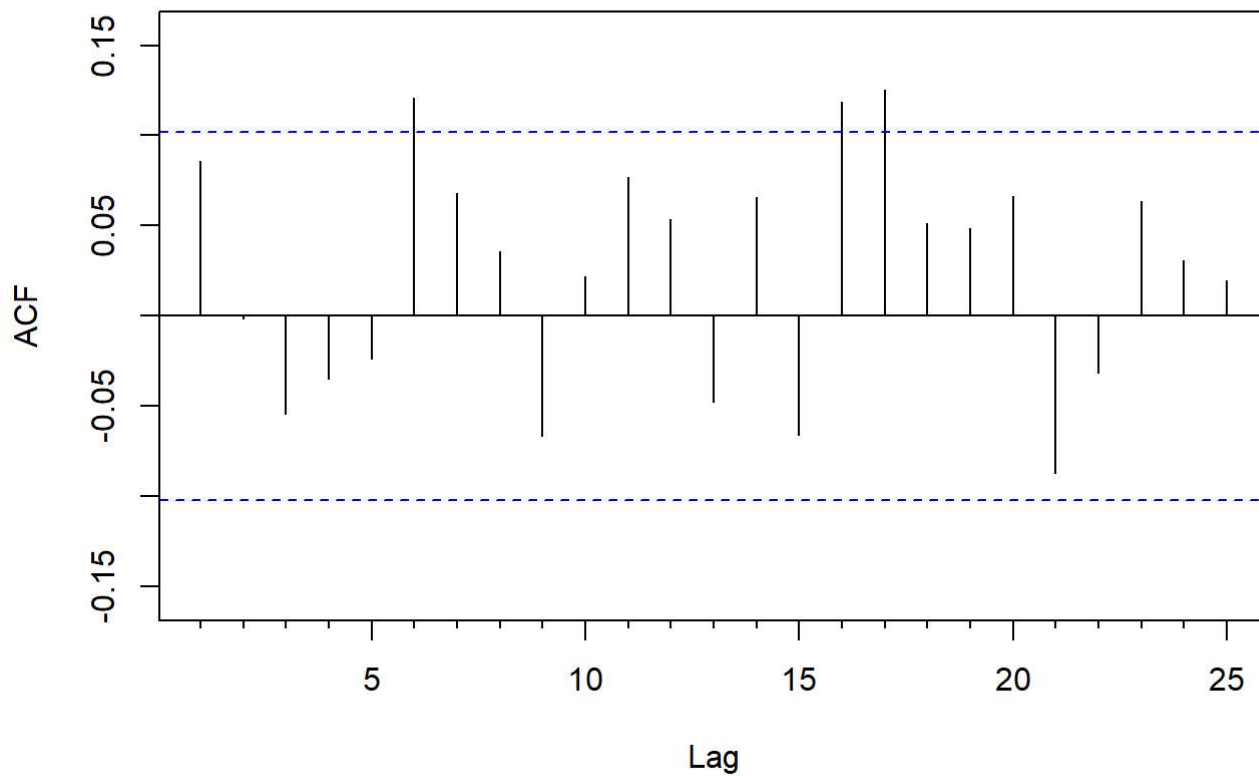
```
Acf(n_ibmclose$residuals, main = "Correlogram of Residuals - Naive Method")
```

Correlogram of Residuals - Naive Method
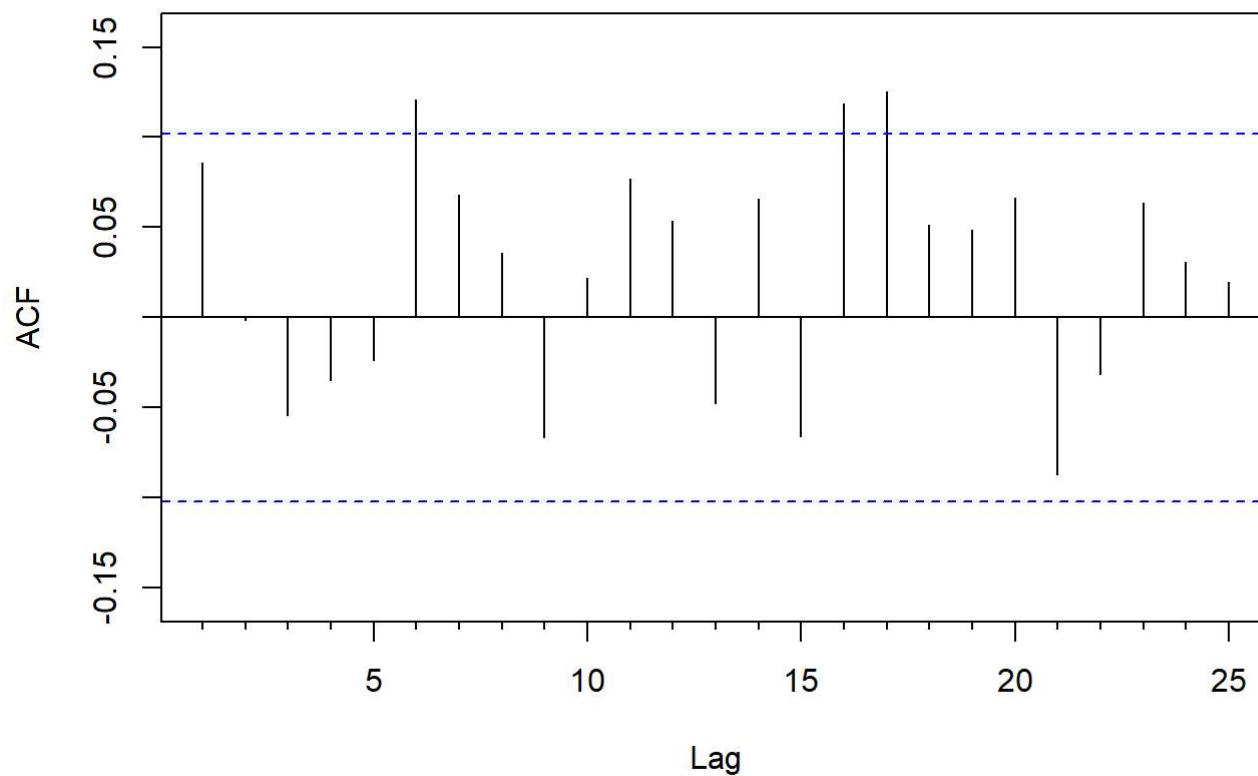

```
Acf(sn_ibmclose$residuals, main = "Correlogram of Residuals - Seasonal Naive Method")
```

Correlogram of Residuals - Seasonal Naive Method


```
print("Mean Method")
```

```
## [1] "Mean Method"
```

```
accuracy(mean_ibmclose)
```

```
##                        ME      RMSE     MAE       MPE      MAPE      MASE       ACF1
## Training set 9.42456e-15 84.10504 72.8893 -3.498933 16.54369 13.99231 0.9933992
```

```
print("Naive Method")
```

```
## [1] "Naive Method"
```

```
accuracy(n_ibmclose)
```

```
##                      ME      RMSE      MAE        MPE      MAPE MASE       ACF1
## Training set -0.2798913 7.253747 5.209239 -0.08469131 1.197176    1 0.08557559
```

```
print("Seasonal Naive Method")
```

```
## [1] "Seasonal Naive Method"
```

```
accuracy(sn_ibmclose)
```

```
##                       ME      RMSE      MAE         MPE      MAPE MASE        ACF1
## Training set -0.2798913 7.253747 5.209239 -0.08469131 1.197176    1 0.08557559
```

Exercise 4

sunspotarea

Consider the time series sunspotarea, which shows the annual average sunspot area between 1875 and 2015. Investigate whether the seasonal naıve method is a good way to forecast this series.

```
sn_sunspotarea = snaive(sunspotarea, h=10)
plot(sn_sunspotarea, main = "Seasonal Naive Method for Sun Spot Area", ylab = "Annual average su
nspot area", xlab = "Year")
```


Seasonal Naive Method for Sun Spot Area
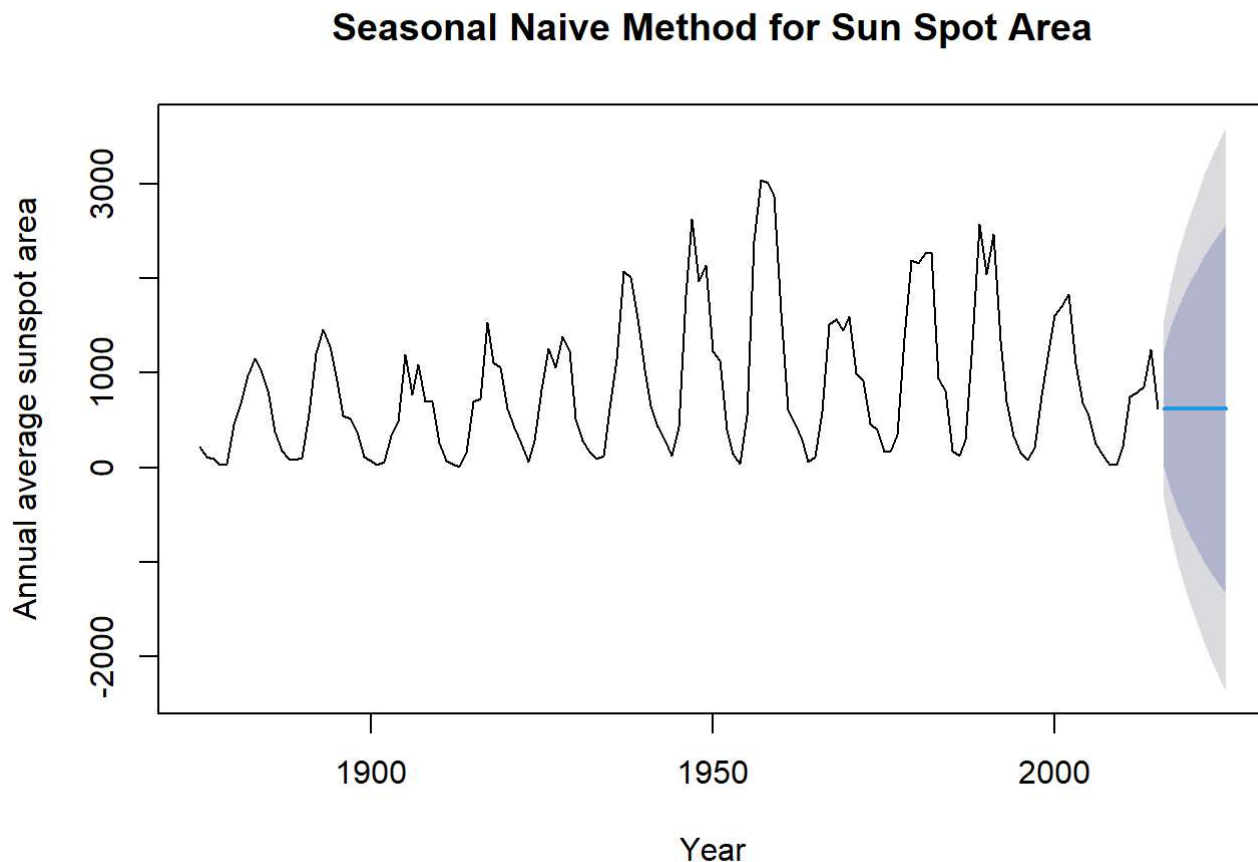
```
Acf(sn_sunspotarea$residuals, main = "Correlogram of Residuals - Seasonal Naive Method")
```

Correlogram of Residuals - Seasonal Naive Method
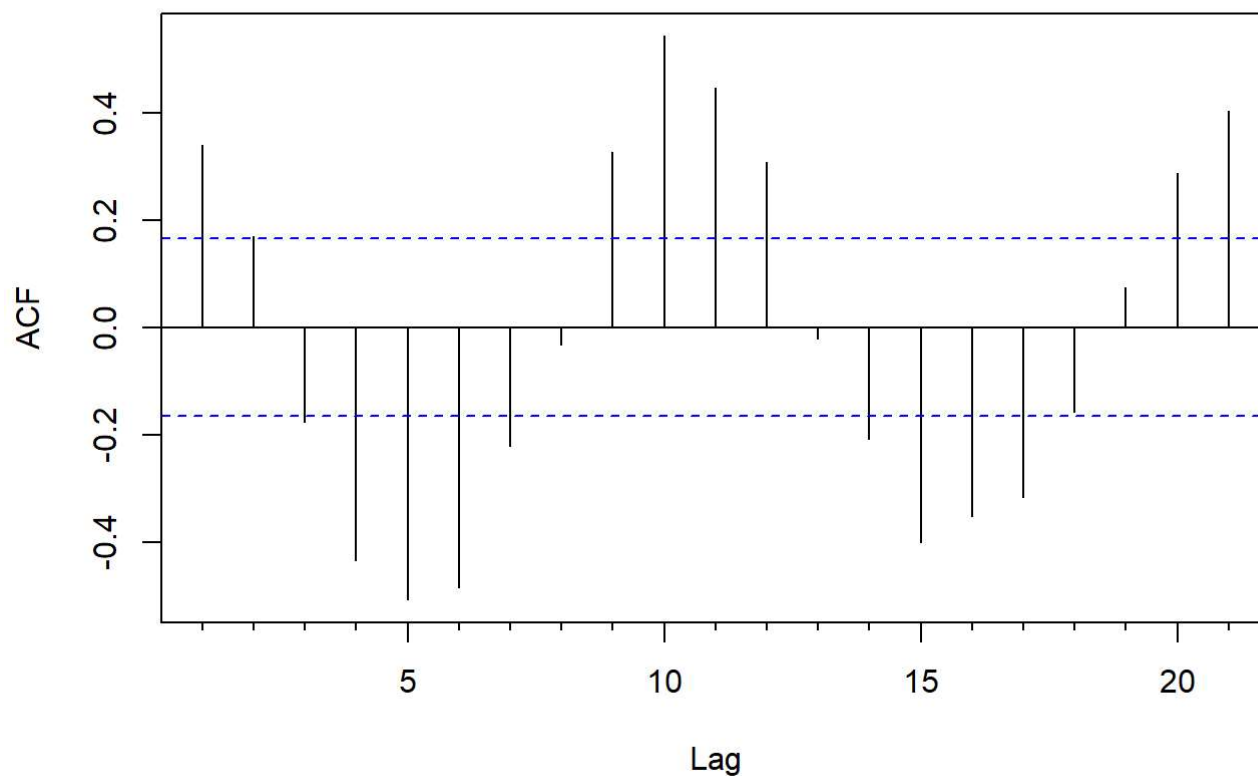

```
accuracy(sn_sunspotarea)
```

```
##                       ME      RMSE      MAE       MPE     MAPE MASE      ACF1
## Training set   2.897679 481.4744 345.6567 -36.62194 74.49806    1 0.3404685
```

```
ljung <- Box.test(sn_sunspotarea$residuals, type = "Ljung-Box", lag = 8)
print(ljung)
```

```
##
##   Box-Ljung test
##
## data:  sn_sunspotarea$residuals
## X-squared = 133.13, df = 8, p-value < 2.2e-16
```

Inference:

- Based on the accuracy (MAPE), it can derived that Seasonal Naive Method is not the appropriate method to go with.
- The L-Jung Box test highlights that the hypothesis of not having any autocorrelations between the residuals can be rejected.
- On the basis of ACF, it can be inferred that there is seasonality.
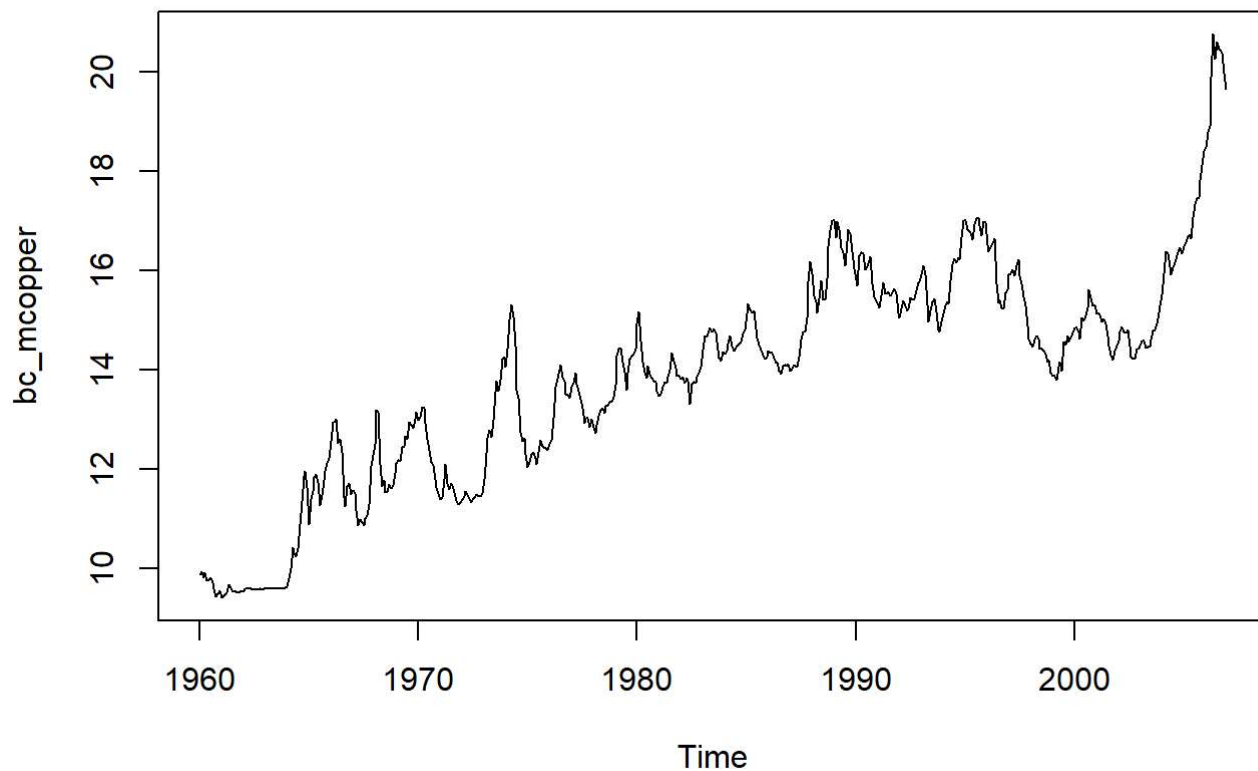
Exercise 5

mcopper

(i) Plot the time series.

```
plot(mcopper, main = "Time Series for Monthly Copper Prices", ylab= "Copper prices", xlab = "Year")
```


Time Series for Monthly Copper Prices
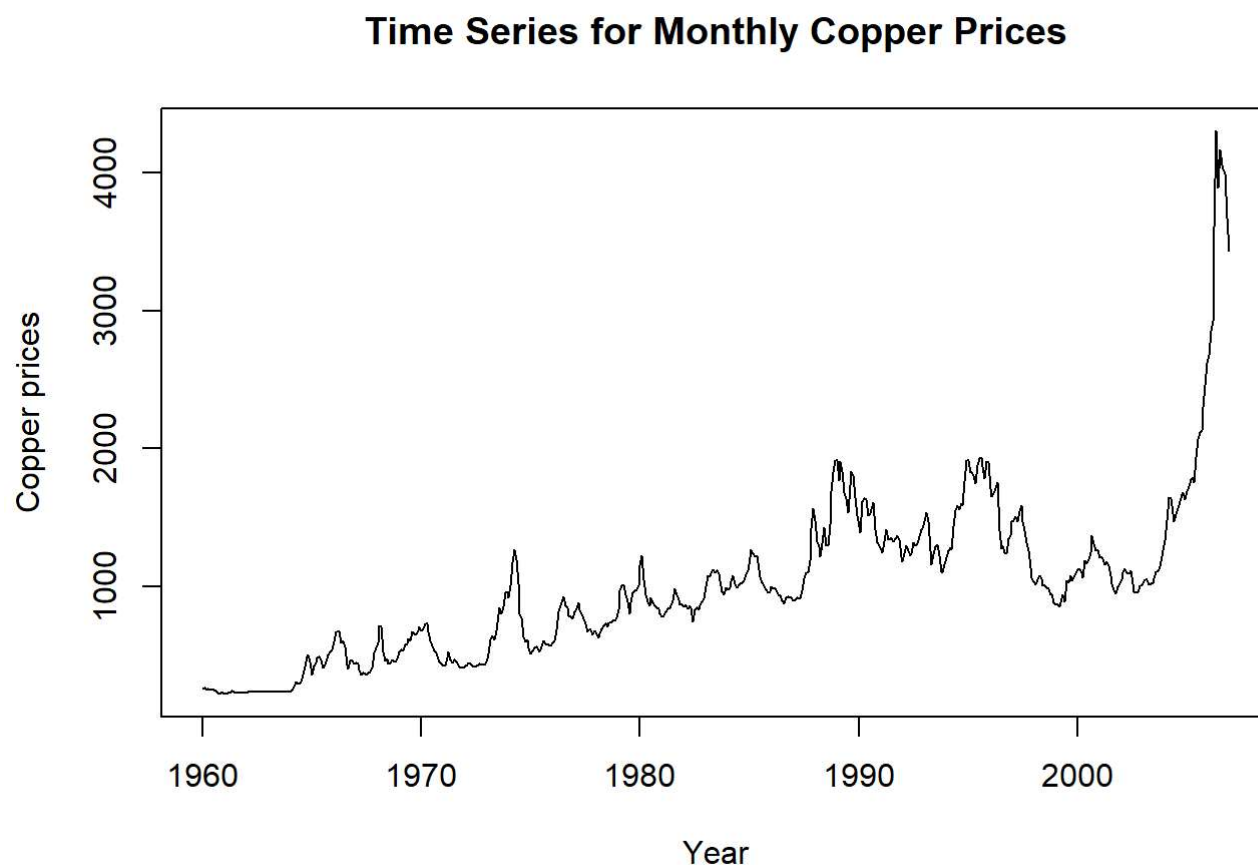
(ii) Perform an appropriate Box Cox transformation.

```
lambda <- BoxCox.lambda(mcopper)
bc_mcopper <- BoxCox(mcopper, lambda)
plot(bc_mcopper)
```

(iii) Plot the transformed data along with a five year forecast using the drift method.

```
d_mcopper = rwf(bc_mcopper,h=60, drift=TRUE)
plot(d_mcopper, main = "Five year forecast for the copper prices - Drift Method", ylab = "Copper
Prices", xlab = "Year")
```
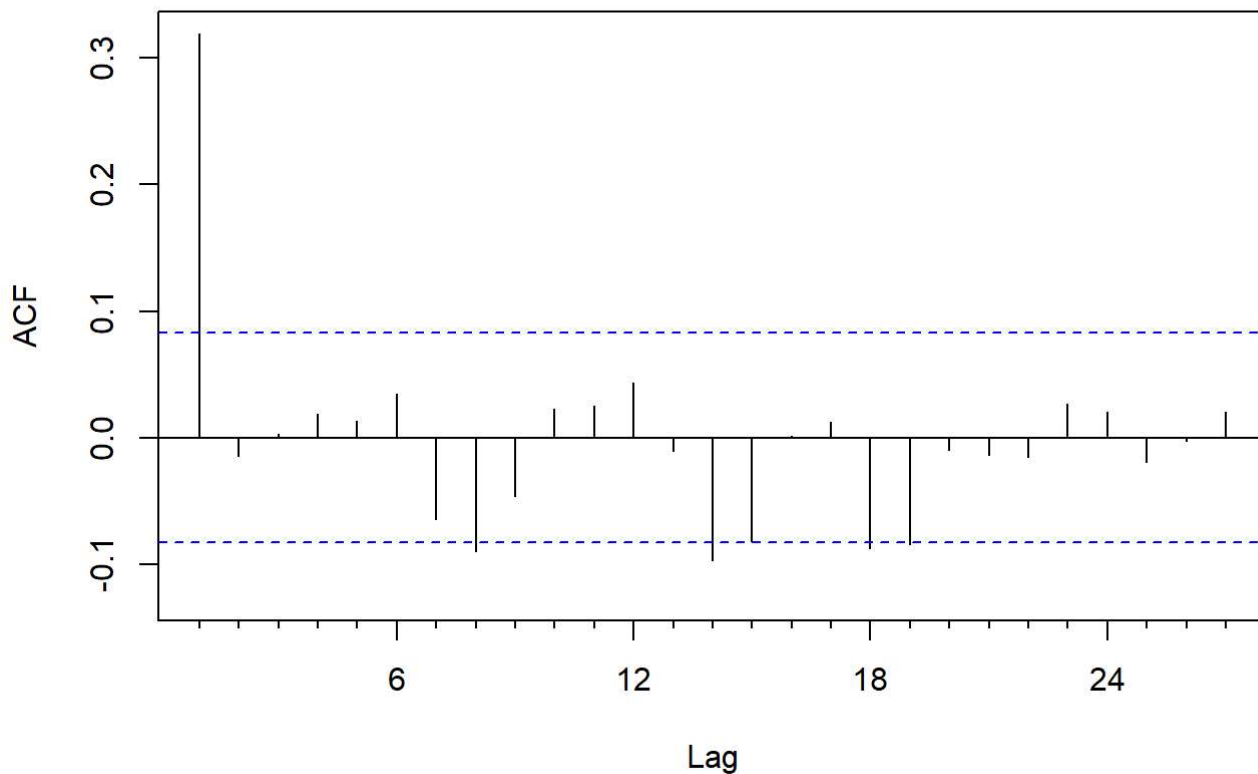
Five year forecast for the copper prices - Drift Method
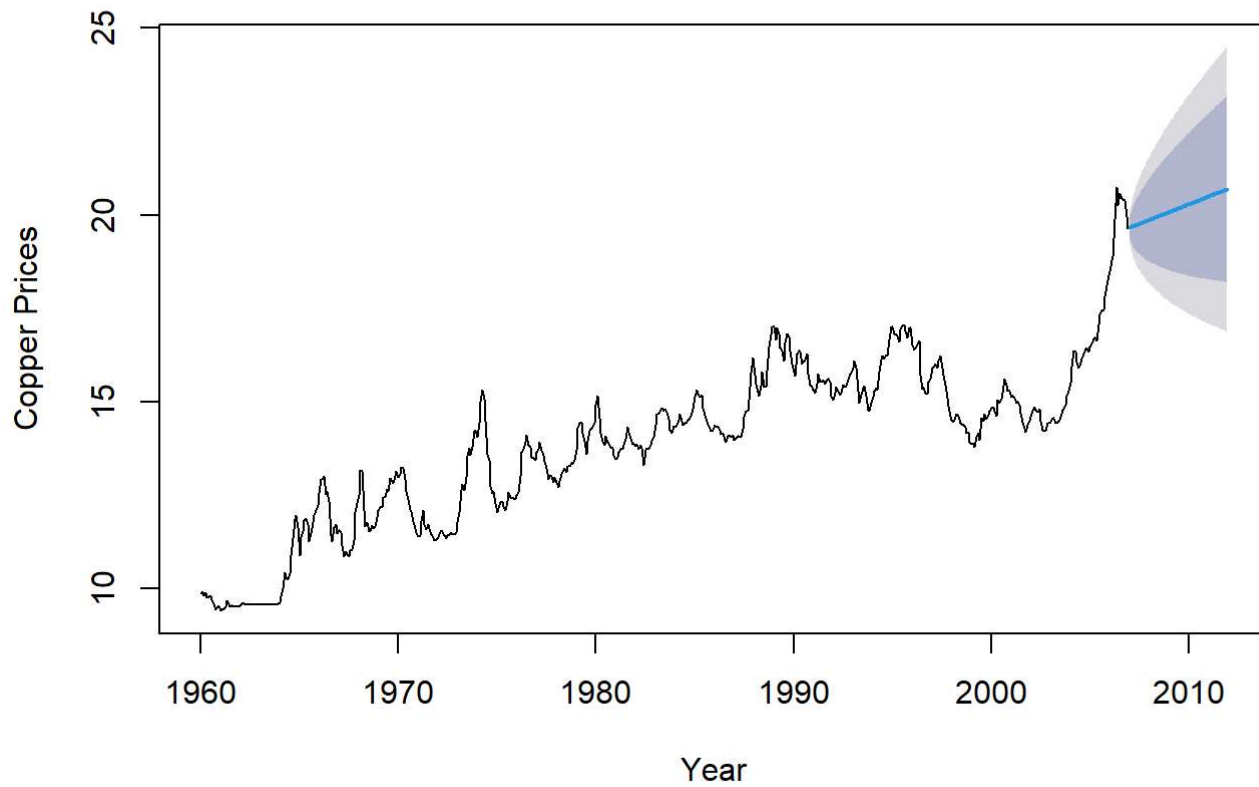
(iv) Plot a correlogram of the residuals.

```
Acf(d_mcopper$residuals, main = "Correlogram of Residuals - Drift Method")
```

Correlogram of Residuals - Drift Method

(v) Perform a Ljung-Box test with lag 8 on the residuals.

```
ljung <- Box.test(d_mcopper$residuals, type = "Ljung-Box", lag = 8)
print(ljung)
```

```
##
##   Box-Ljung test
##
## data:  d_mcopper$residuals
## X-squared = 65.603, df = 8, p-value = 3.666e-11
```

(vi) Use parts (iv) and (v) to comment on whether there is autocorrelation or seasonality.

- Deriving from (iv), it can be observed that there is *seasonality* in the data.
- On the basis of (v), it can be inferred that since the *p-value <0.05* there is autocorrelation between residuals.

Exercise 6

huron

(i) Use the window function to split the time series into a training set consisting of the data from 1875 to 1955 and the test set consisting of the data from 1956 to 1972.

```
train_huron = window(huron, start=1875, end=1955)
test_huron = window(huron, start = 1956, end = 1972)
```
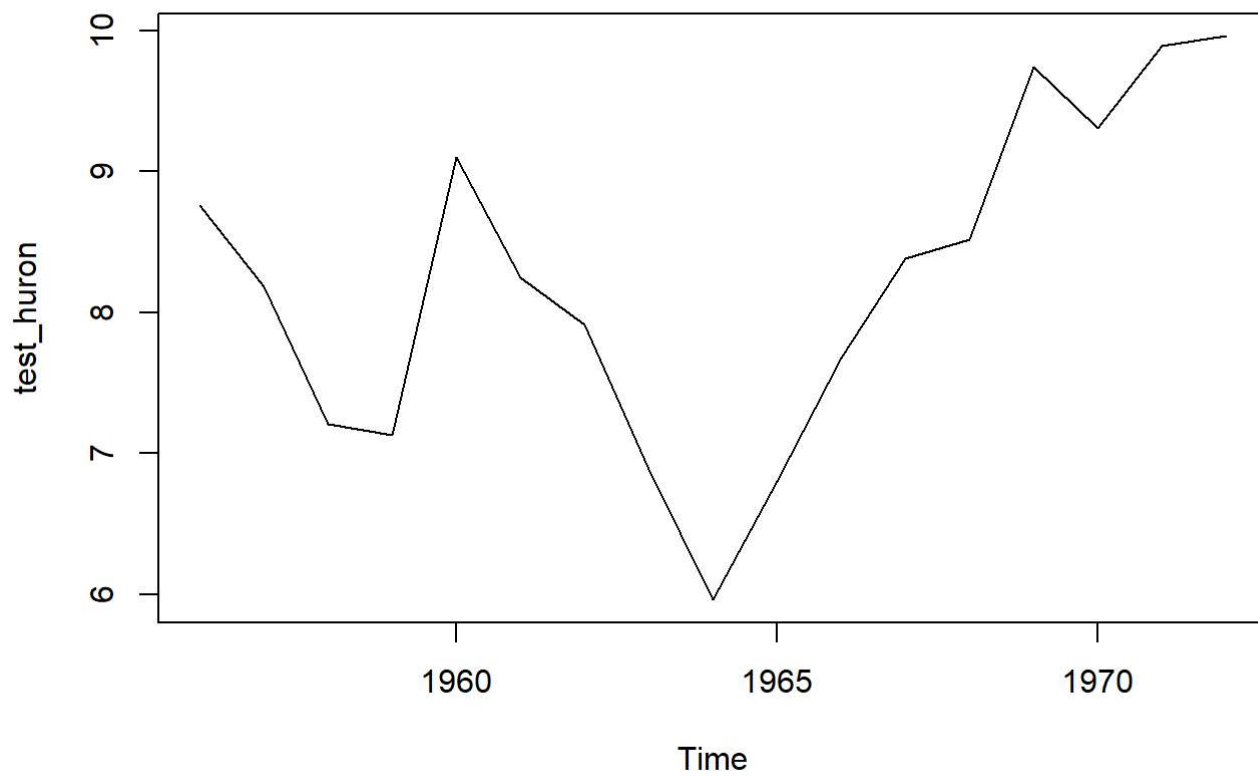
(ii) Check you have correctly split up the data by producing two separate plots

```
plot(train_huron, main = "Training set check")
```

Training set check
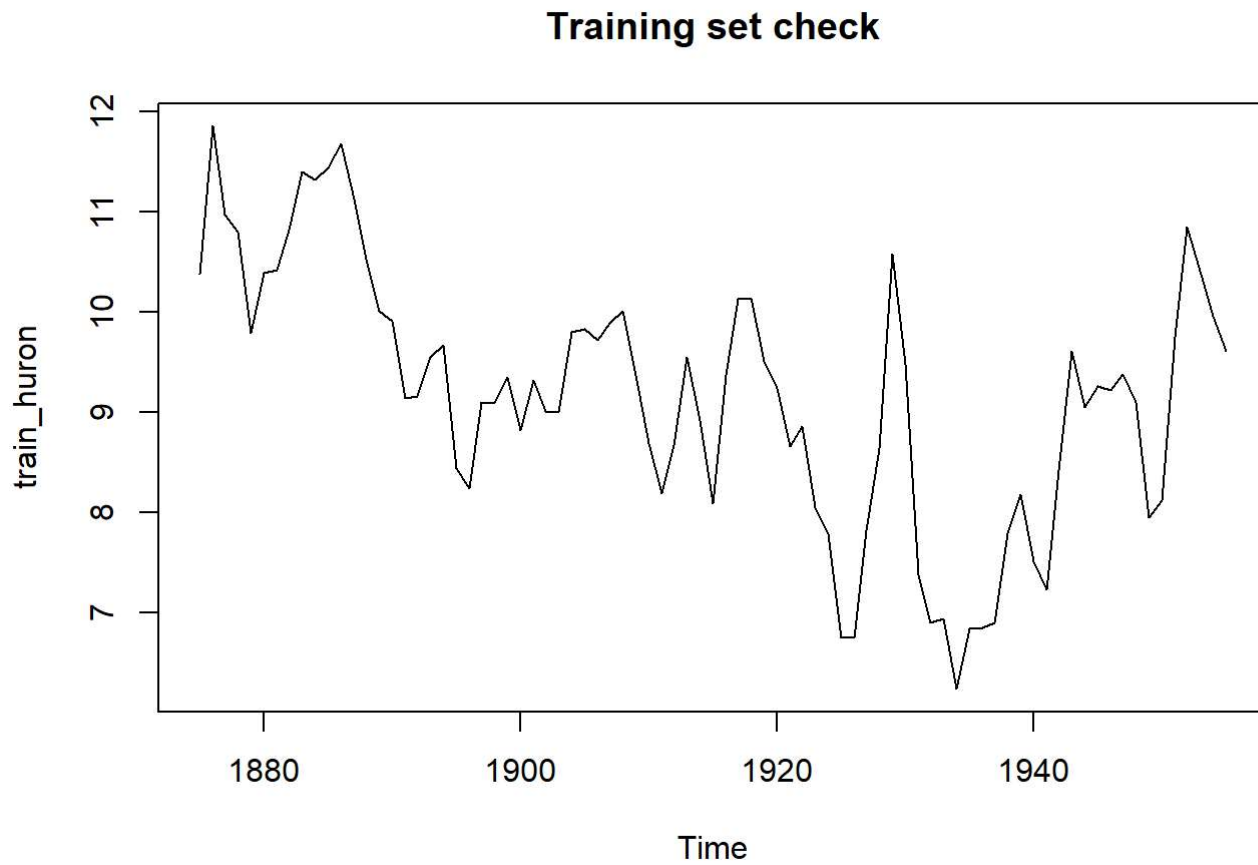

```
plot(test_huron, main = "Test set check")
```
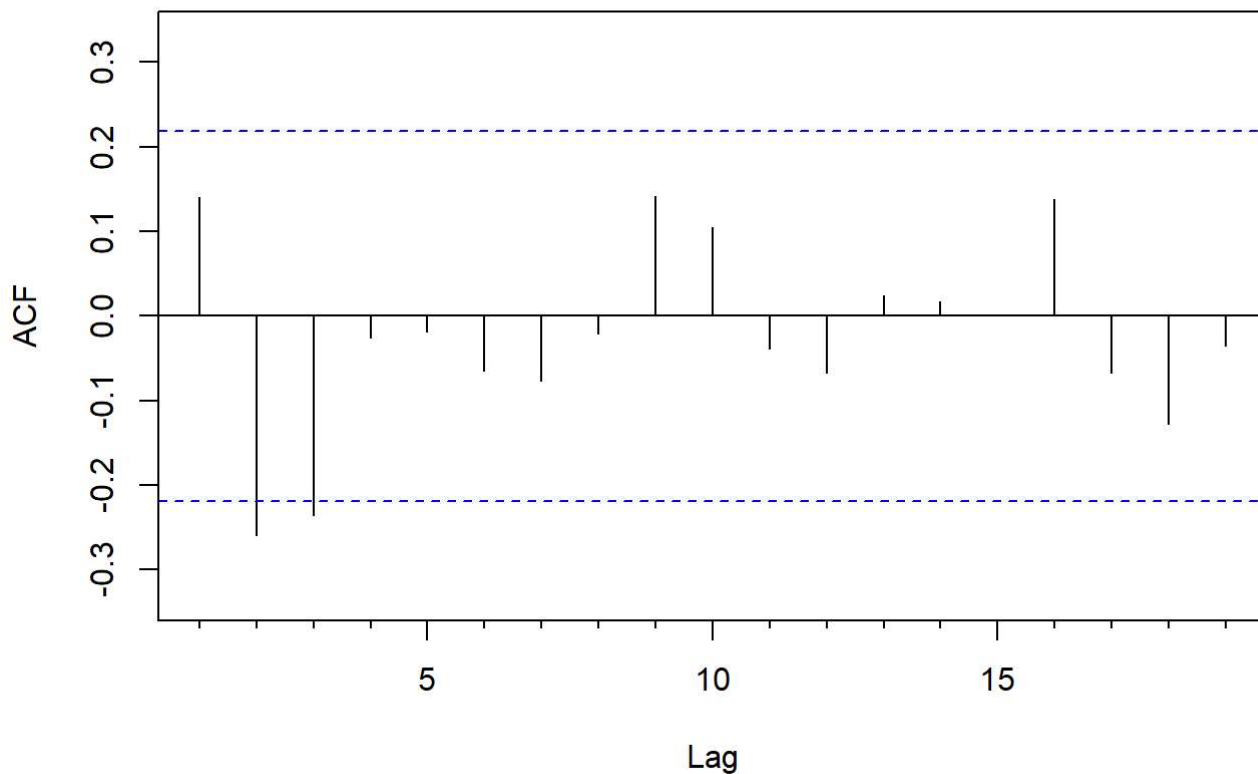
Test set check

(iii) Explain why the seasonal naïve forecast is not appropriate here.

```
sn_huron = snaive(train_huron)
Acf(sn_huron$residuals, main = "Correlogram of Residuals - Seasonal Naive Method")
```

Correlogram of Residuals - Seasonal Naive Method


```
ljung <- Box.test(sn_huron$residuals, type = "Ljung-Box", lag = 2)
print(ljung)
```

```
##
##  Box-Ljung test
##
## data:  sn_huron$residuals
## X-squared = 7.3224, df = 2, p-value = 0.0257
```

Inference:

- Deriving from the ACF and Box-Ljung test, there is autocorrelation between the residuals and therefore, isn't the right option to go with.

(iv) Use the training set to produce three forecasts for the period 1956 to 1972, using the naïve method, the driftmethod and the mean method. Plot each forecast alongside the test data.

```
n_trainhuron = naive(train_huron, h = 17)
d_trainhuron = rwf(train_huron, drift = TRUE, h = 17)
m_trainhuron = meanf(train_huron, h = 17)
```

```
plot(huron)
lines(n_trainhuron$mean, col = "green")
lines(d_trainhuron$mean, col = "blue")
lines(m_trainhuron$mean, col = "red")
```
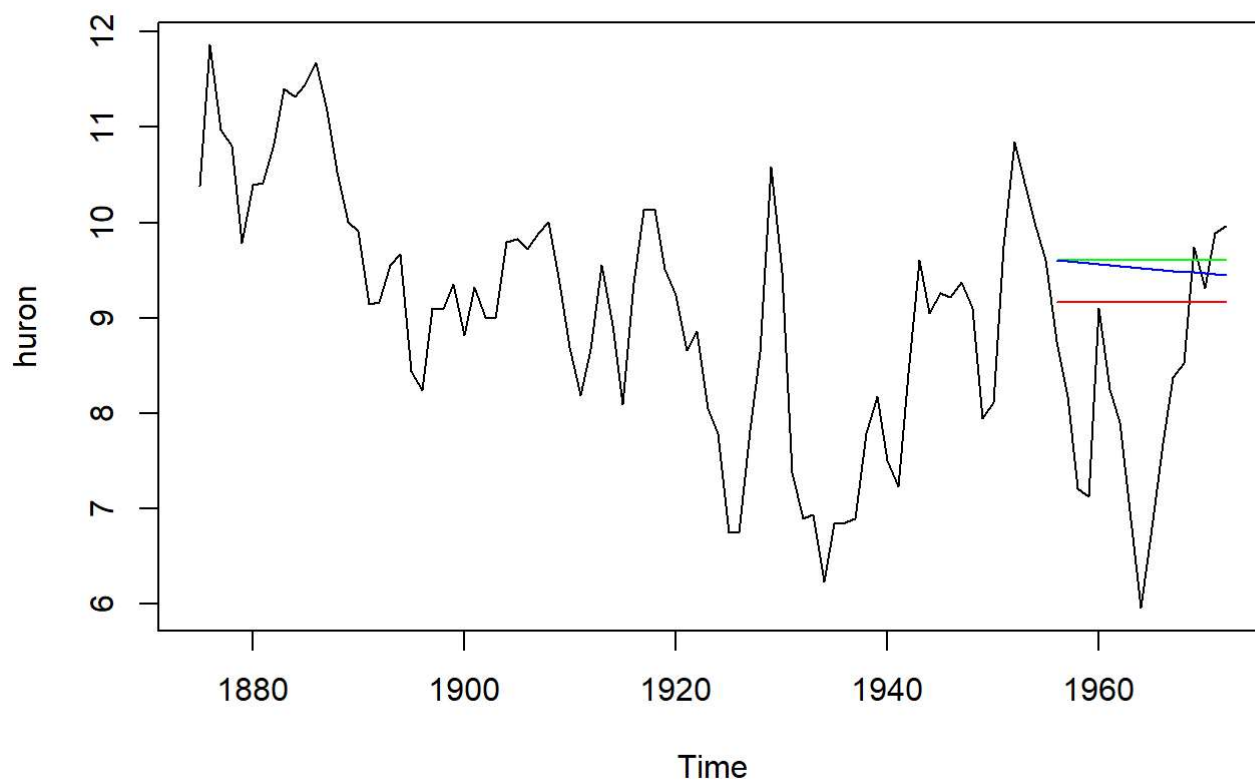

(v) Compare the accuracy of each of the forecasts you calculated in part (iv) by comparing their errors. Which forecast performs the best?

```
print("Accuracy for Naive Method")
```

```
## [1] "Accuracy for Naive Method"
```

```
accuracy(n_trainhuron)
```

```
##                     ME      RMSE      MAE        MPE      MAPE MASE      ACF1
## Training set -0.009625 0.7178988 0.554375 -0.4238952 6.188552    1 0.1407863
```

```
print("Accuracy for Drift Method")
```

```
## [1] "Accuracy for Drift Method"
```

```
accuracy(d_trainhuron)
```

```
##                          ME      RMSE      MAE        MPE     MAPE MASE      ACF1
## Training set -2.44243e-16 0.7178343 0.554375 -0.3165107 6.186713    1 0.1407863
```

```
print("Accuracy for Mean Method")
```

```
## [1] "Accuracy for Mean Method"
```

```
accuracy(m_trainhuron)
```

```
##                          ME   RMSE      MAE        MPE    MAPE     MASE     ACF1
## Training set -1.867803e-16 1.2853 1.016327 -2.130276 11.6982 1.833284 0.839739
```

- Based on the above results (MAPE), it can be inferred that the **Naive Method** and the **Drift Method** have performed well in comparison to the Mean Method.

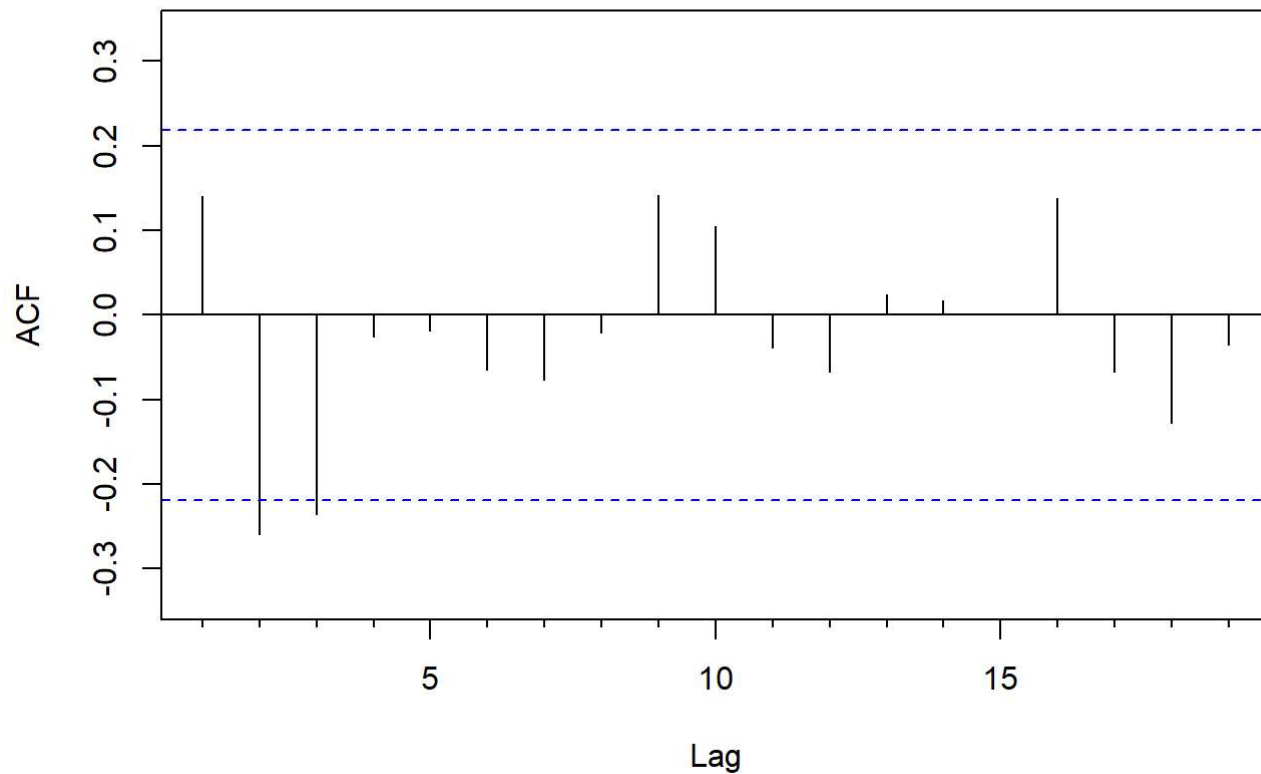(vi) For the best performing method, compute the residuals and plot them.

```
Acf(n_trainhuron$residuals, main = "Correlogram of Residuals - Naive Method")
```


Correlogram of Residuals - Naive Method
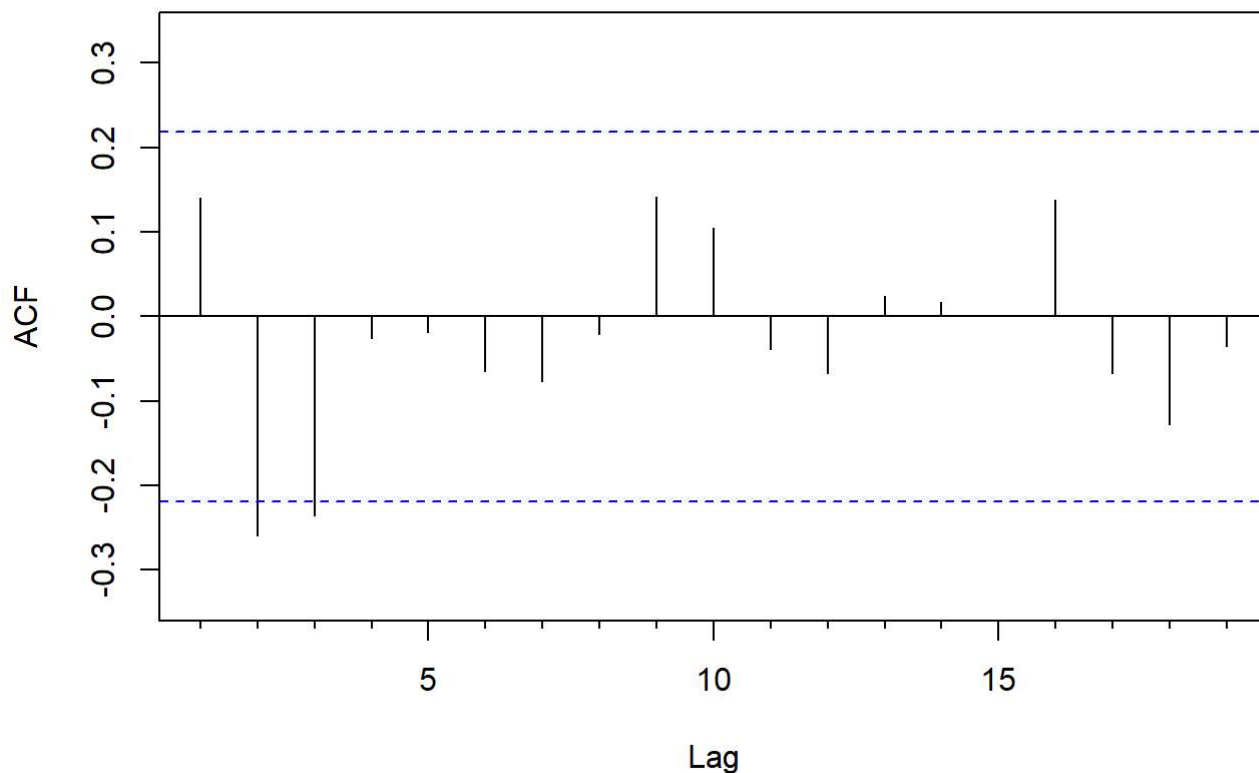
```
Acf(d_trainhuron$residuals, main = "Correlogram of Residuals - Drift Method")
```
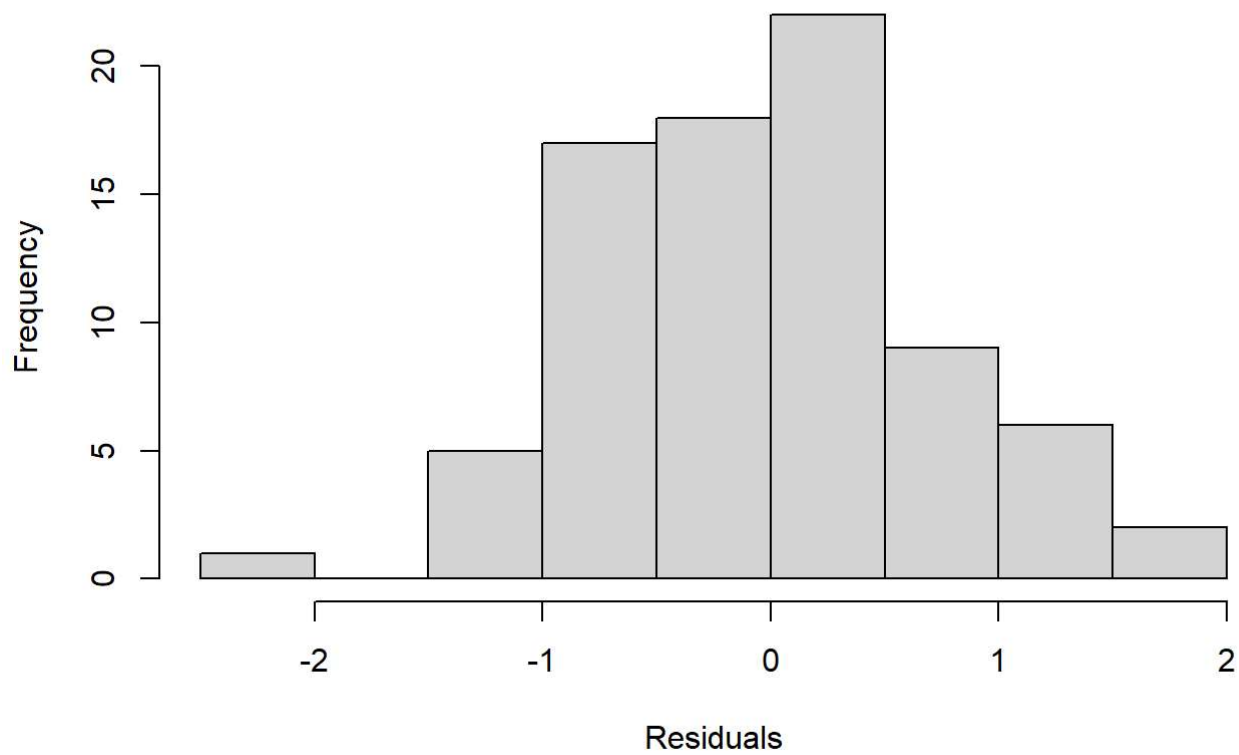
Correlogram of Residuals - Drift Method

(vii) Do the residuals appear to be uncorrelated and normally distributed?
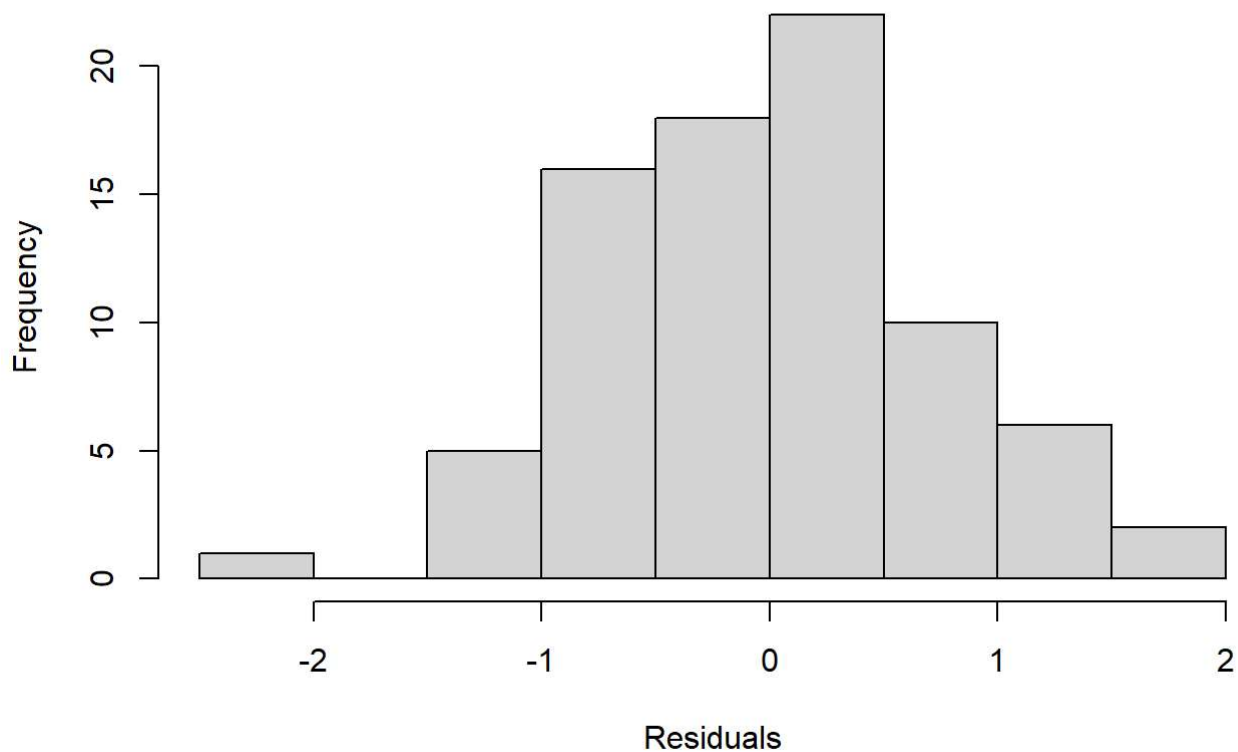
```
hist(n_trainhuron$residuals, main = "Histogram of Residuals in Naive Method",xlab="Residuals")
```

Histogram of Residuals in Naive Method


```
hist(d_trainhuron$residuals, main = "Histogram of Residuals in Drift Method", xlab="Residuals")
```

Histogram of Residuals in Drift Method

- It can be concluded that the residuals are *correlated* and *not normal*. It is right sequenced.

(viii) In light of your answer to part (vii), would you like to change your choice of forecast? If so, check the residuals of your second choice

- No. As the MAPE for Naive and Drift have a very slight difference but are significantly better than the Mean Method. Therefore, my choice of forecasting method remains the same.