CSC425 – Time series analysis and forecasting
Homework 2
YunTzu,Yu

Problem 1 [10 pts]

Continue your analysis of the weekly sales data for **ToothPaste**. Remember, it is in dataset "groceries.csv" which contains weekly units sold for three grocery items: ToothPaste (100ml container of toothpaste), PeanutButter (340g. jat of crunchy peanut butter), and Biscuits (200g., 10 finger package of shortbread cookies). The dataset contains the variable Date defined as the first day of the week for the sales period. Answer the following questions

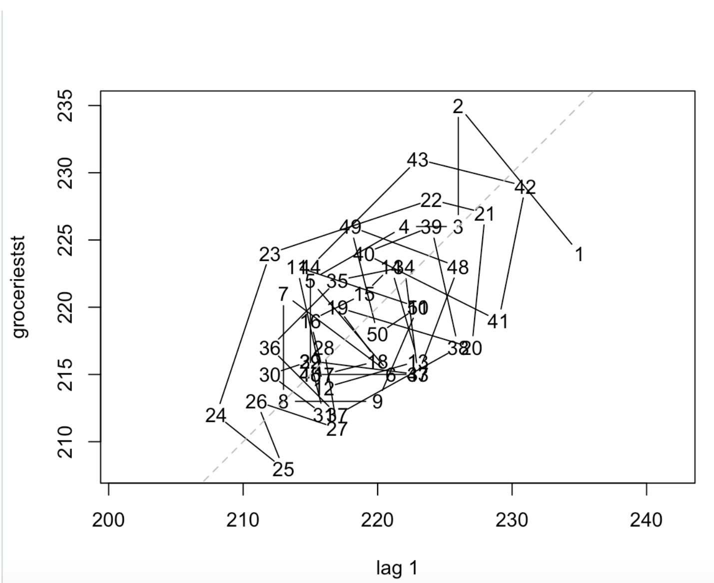
a) Graph the lag plot of the ToothPaste series and its first lag. Interpret the graph. What does it tell you about autocorrelation for this series?

We can see a clear correlation between the series of time t and the series of time t-1, so it demonstrates there is a serial dependence.

groceries <- read_csv("Desktop/dsc 425/module 1/groceries.csv")

grocerieststst <-ts(groceries$ToothPaste,start= 2008.1,frequency = 1)

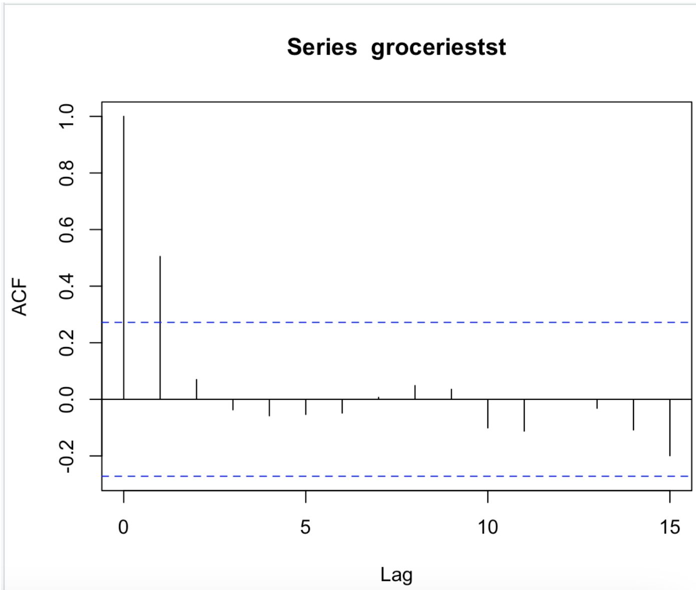
lag.plot(grocerieststst,lag = 1)



b) Compute and plot the first 15 lags of the autocorrelation function for ToothPaste weekly sales and discuss if the series shows evidence of serial correlation.

We grouped all single correlations together by applying the autocorrelation function, which measures all the dependence among observations in a time series at different lags.
From the graph, we can tell when we need to pay attention to these autocorrelations in the series.
The autocorrelation at lag 1 is 0.5, and then it falls rapidly to nearly zero.

acf(grocerieststst,lag.max=15,plot = TRUE)



c) Use the Ljung Box test to hypothesize that ToothPaste weekly sales have a significant serial correlation.

We applied the Ljung Box test to test the null hypothesis that the first 10 correlations are all zero
Here if we test it with lag =10, the p-value is 0.1049, which could almost fail to reject the hypothesis, however, if we use lag =1, the p-value is 0.0001792, and we definitely could reject the null hypothesis. This could be because only lag one has a correlation in this time series.

```
> Box.test(grocerieststst,lag =1, type = "Ljung-Box")

        Box-Ljung test

data:  grocerieststst
X-squared = 14.038, df = 1, p-value = 0.0001792

> Box.test(grocerieststst,lag =10, type = "Ljung-Box")

    Box-Ljung test
```

data: grocerieststt
X-squared = 15.819, df = 10, p-value = 0.1049

d) Discuss the importance of weak stationarity for time series analysis and describe a method to analyze whether a TS is stationary.

Weak stationarity is the mean and the standard deviation is finite and time-invariant, and autocorrelations do not depend on time, so we could use the past data to predict the future. We could see the plot of ACF to tell if it is a stationary time series. The stationary time series usually falls rather rapidly and become zero in a while. The time series of weekly sales data for ToothPaste is stationary.

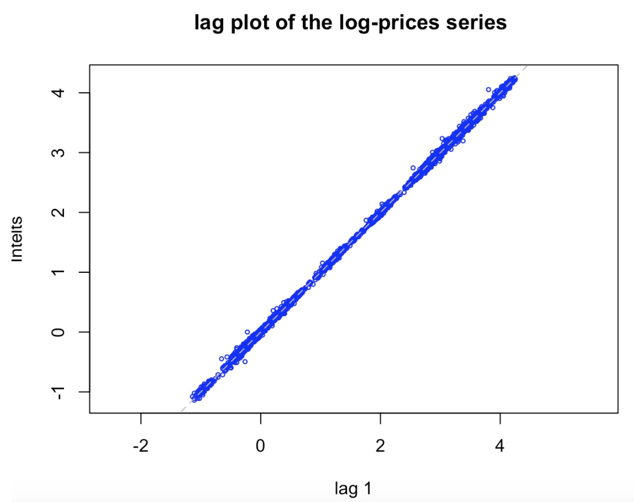
Problem 2 [10 pts]

Analyze the long-term behavior of the Intel stock price by investigating its autocorrelation behavior.

a) Graph the lag plot of the log-prices series and its first lag. Interpret the graph. What does it tell you about autocorrelation for this series?

We can see the correlation between two time series is pretty strong, which demonstrates that there is a lot of serial dependence.

```
library(lubridate)

Intel <- read_csv("Desktop/dsc 425/module 2/Intel-1986-2007.csv")

View(Intel)

Intel$Dates=as.Date(date_decimal(Intel$Date),tz = "UTC")

Intel=Intel[,-c(2)]

Intelts = ts(log(Intel$`Adj Price`),start = (1987-01-01),frequency = 1)

lag.plot(Intelts,lag=1, pch = 1, col="blue", cex=0.5,main = "lag plot of the log-prices series ")
```

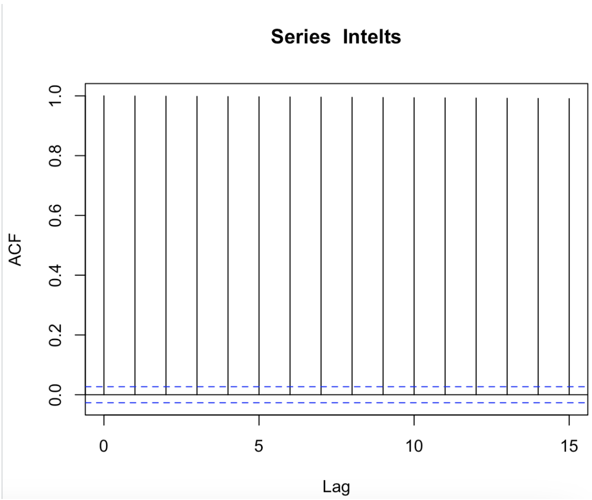


b) Compute and plot the first 15 lags of the autocorrelation function of the series of log-prices and discuss if the series shows evidence of serial correlation.

If serial correlations are not zero then future observations can be predicted from the past. We can see from the ACF that all of the serial correlations are non-zero for the first 15 lags.

```
> acf(Intelts,lag.max=15,plot =FALSE)

Autocorrelations of series ‘Intelts’, by lag

    0     1     2     3     4     5     6     7     8     9    10    11 
1.000 0.999 0.999 0.998 0.998 0.997 0.996 0.996 0.995 0.994 0.994 0.993 
   12    13    14    15 
0.993 0.992 0.991 0.991
```

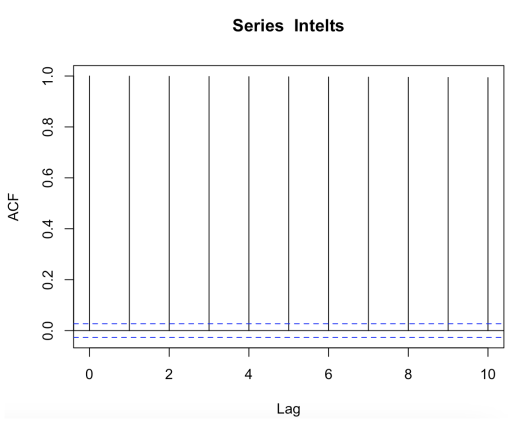
acf(Intelts,lag.max=15,plot =TRUE)



c) Explain the pattern that you see in the first 10 lags of the autocorrelation function. Does it appear that the series is stationary? Give a reason why or why not.

We can see from the time series and tell that it is a non-stationary time series because it has a clear trend in the time plot and its acf plot shows a slow decay.

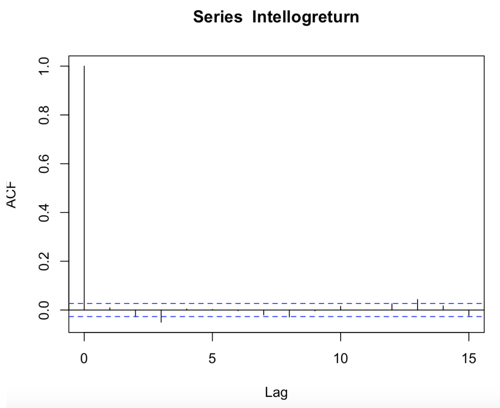
```
autoplot(Intelts)
acf(Intelts,lag.max=10,plot =TRUE)
```



d) Compute and plot the first 15 lags of the ACF of the log returns (i.e. the difference of the logs). Interpret the result.

We can see from the plot that there are only two slight serial correlations at lag 3 and lag 13.

```
Intellogreturn=diff(log(Intel$`Adj Price`))
acf(Intellogreturn,lag.max=15,plot =TRUE)
```



e) Use the Ljung Box test to hypothesize that the log-returns have a significant serial correlation.

We could see that the p-value is 0.003534, so we almost could reject the null hypothesis that all are equal to zero.

Box.test(Intellogreturn,lag =10, type = "Ljung-Box")

Box-Ljung test

data: Intellogreturn
X-squared = 26.157, df = 10, p-value = 0.003534

Problem 3 [10 pts]

Consider the following AR(1) time series process: $r_t = .9r_{t-1} + a_t$, where is a Gaussian (i.e. normally distributed) white noise series with zero mean and constant variance σ^2=0.5.

a) What is the mean of the time series r_t?

0/1-0.9 = 0

b) Determine if the AR(1) model is stationary. Explain.

The model is stationary because the mean doesn't change

c) Determine the overall variance of the series.
.5/(1-.81)= 2.63

Problem 4 [10 pts]

Consider the following MA(1) time series process: $X_t = 5 + a_t - 0.5a_{t-1}$, where $\{a_t\}$ is a Gaussian white noise series with mean zero and constant variance σ^2=0.025.

a) What is the mean of the time series X_t?
5
b) What is the variance of the series?
1+-.5 =.5
c) Discuss if the MA (1) model is stationary. Explain.
moving average model is always stationary

Problem 5 [20 pts]

The Institute for Supply Management (ISM) (http://www.ism.ws/) is responsible for maintaining the Purchasing Managers Index (PMI). The index is derived from monthly surveys of private-sector companies. Monthly data from 01/01/1980 to 12/01/2015 are saved in the NAPM.csv file. This problem asks you to apply an AR(p) autoregressive model to describe the dynamic behavior of the index.

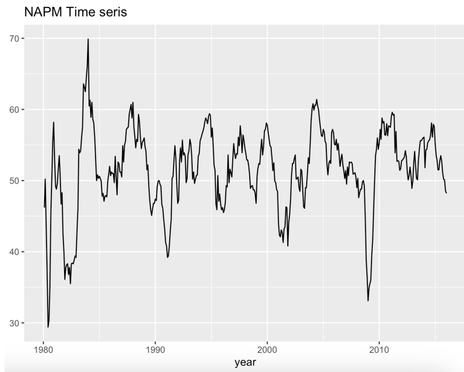
a) Import the data and create a time series object for the index using the ts() function where the starting date is the first month of 1980.

```
library(readr)

NAPM <- read_csv("Desktop/dsc 425/module 3/NAPM.csv")

NAPMts<- ts(NAPM$index, start = 1980.1,frequency = 12)
```

b) Create a time plot of the data and determine whether the series is multiplicative or additive. If it is multiplicative, compute the log of the series before proceeding to the next parts of this question.

From the graph, we could tell that it should be additive because the variance tends to stay the same no matter what size the actual value is in the long term.

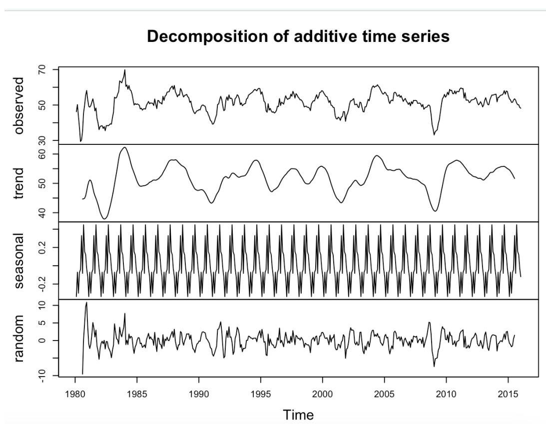
```
autoplot(NAPMts,main= "NAPM Time seris",xlab = "year")
```



c) Create a decomposition of the series and analyze the series for trends and seasonality.
Describe what you find in detail.

We can see from the graph of the original series that the pattern did not change a lot in the long term, and there is no clear trend either.
It is hard to tell if there is any pattern from the random graph, as the range of random should be as small as possible, but it is pretty large compared to the seasonal graph.

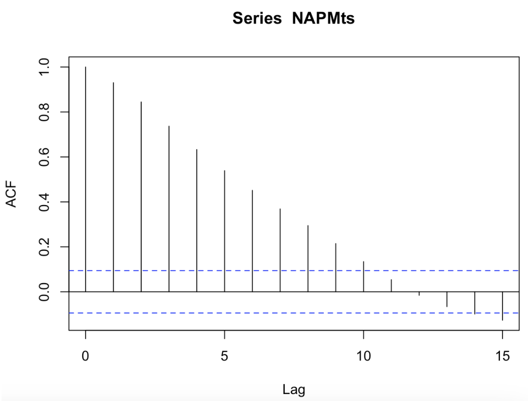
```
deNAPMts  = decompose(NAPMts)
plot(deNAPMts)
```



d) Analyze if the time series is serially correlated using the ACF plot and the Ljung Box test.

We can see from the ACF that there are serial correlations in the first 10 lags.
We got the p-value 2.2e-16 from the Ljung Box test, which means that we can reject the null hypothesis that all equal zero.

```
acf(NAPMts,lag.max=15,plot =TRUE)
```




```
> Box.test(NAPMts,lag =15, type = "Ljung-Box")

        Box-Ljung test

data:  NAPMts
X-squared = 1459.3, df = 15, p-value < 2.2e-16
```

e) Fit an AR(2) model with the Arima function as described in class and write down the expression of the estimated AR(p) model. Explain the meanings of the coefficients in the model.

```
> s = NAPMts[-length(NAPMts)]
> fit = Arima(s, order = c(2,0,0))
> fit

Series: s
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2     mean
      1.0881  -0.1686  51.3669
s.e.  0.0476   0.0476   1.2323

sigma^2 = 4.449:  log likelihood = -932.79
AIC=1873.58   AICc=1873.67   BIC=1889.84
```

f) Examine the significance of the model coefficients using the "coeftest" function from the "lmtest" package. Discuss which coefficients are significantly different from zero.

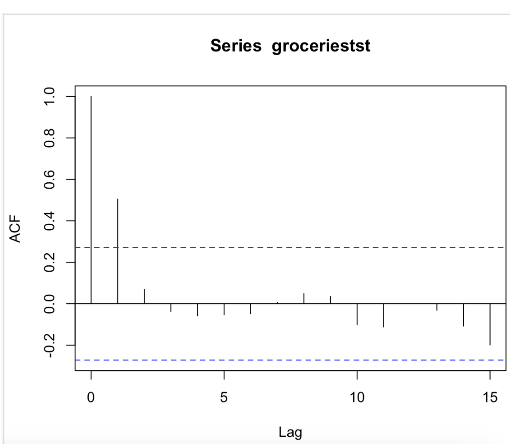
We could check the p-values for the coefficient, all the p_values are small enough that we can reject the null hypothesis that is equal to zero.

```
> coeftest(fit)

z test of coefficients:

          Estimate   Std. Error z value   Pr(>|z|)
ar1       1.088071   0.047608 22.8548   < 2.2e-16 ***
ar2      -0.168558   0.047611 -3.5403   0.0003996 ***
intercept 51.366862  1.232284 41.6843   < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 6 [10 pts]
Continue your investigation of the ToothPaste series.

a) From the ACF plot that you computed in problem 1, determine a degree for an MA model for the series. Explain your choice.

By checking ACF plot, we only see one serial correlation, so I would use 1 degree for an MA model.



b) Fit an MA(#) model with for that degree using the Arima function as described in class and write down the expression of the estimated AR(p) model. Explain the meanings of the coefficients in the model.

MA parameter is 0.6, and the mean is 219.43.
The model is v_t = a_t+.6*a_t-1+219.43 .

```
> fit = Arima(grocerieststst, order =c(0,0,1))
> fit
Series: grocerieststst
ARIMA(0,0,1) with non-zero mean

Coefficients:
       ma1      mean
    0.6000  219.4313
s.e. 0.1075    1.0399

sigma^2 = 23.17:  log likelihood = -154.7
AIC=315.4   AICc=315.9   BIC=321.26
```

c) Examine the the significance of the model coefficients using the “coeftest” function from the “lmtest” package. Discuss which coefficients are significantly different from zero.

We could check the p-values, and find that they all are small enough that we can reject the null hypothesis, so the coefficients are all significantly different from zero.

```
> library(lmtest)
> coeftest(fit)

z test of coefficients:

          Estimate Std. Error  z value  Pr(>|z|)
ma1        0.60002    0.10752   5.5807 2.395e-08 ***
intercept 219.43130   1.03994 211.0045 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

“Reflection” Problem [5 pts]

Post a message on the discussion board reflecting on the topics in week 2 and on homework 2. Indicate the assignment in this module you found to be the easiest, the one you found to be the hardest, and why. A new thread called “Homework 2 Reflection” will be created.