CSC 425 Homework 1
Due Weds 9/28
Sarah Cummings

Problem 1: Timeseries with 500 observations has the following sample autocorrelation values:
Lag(1): -0.374 Lag(2): 0.019 Lag(3): 0.143

a) Calculate the Q stat of lag(3)

$$Q(3) = T(T+2)\sum_{k=1}^{3} \frac{\hat{\rho}_k^2}{T-k}$$ $=T(T+2)\ [(p_1^2/T\text{-}1) + (p_2^2/T\text{-}2) + (p_3^2/T\text{-}3)]$

T= 500 p_1= -0.374 p_2= 0.019 p_3= 0.143

$500(500+2)\ [(-0.374^2/500\text{-}1) + (\ 0.019^2/\ 500\text{-}2) + (0.143^2/\ 500\text{-}3)]$

=80.868

b) The $\chi 2$ statistic with two degrees of freedoms equals 5.99. State the null hypothesis and form a conclusion. Note: If r_t is a finite variance iid sequence, we reject the hypothesis of zero auto-correlation if $Q(m) > x_{(1\text{-}a)}^2$ where $x_{(1\text{-}a)}^2$ is the (1-α)-th percentile of the chi-squared distribution with m degrees of freedom

Given Q(m)= 80.868 which is greater than $x_{(1\text{-}a)}^2$, we reject the null hypothesis of zero autocorrelation. We conclude that at least one autocorrelation value is not zero, and the sequence r_t is serially correlated at some lag L.

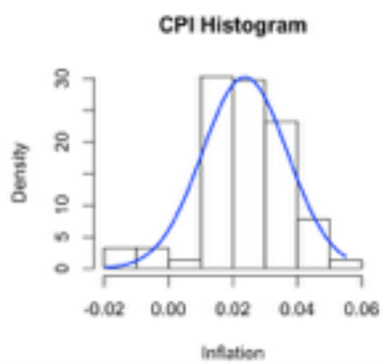
Problem 2: Consumer Price Index

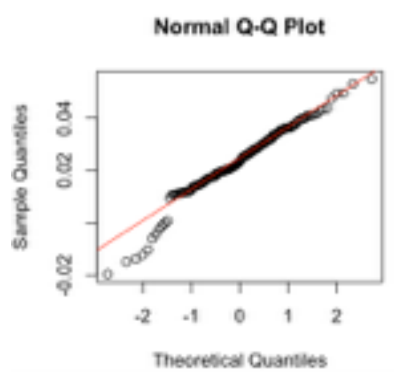
a) The following values for inflation rates:

Sample mean: **0.023670** Standard Deviation: **0.013215** Skewness: **-0.571439**

Excess Kurtosis: **0.934786** Minimum: **-0.019615** Maximum: **0.054975**

b) Analysis of distribution for inflation rates using histogram and normal quantile plot.





Looking at our histogram, it appears as though our data is not quite symmetric. We can refer back to our summary statistics and see that the mean (0.023670) and the median (0.023179) are very similar, but there is some imbalance with the left tail. Our skewness value (-0.571) is a little further from zero than we would like, so it probably won't pass the symmetry test.

The QQ plot shows the data is fairly normal in the middle, with some fatness in the lower tail.

c) Null hypothesis test of perfect symmetry for inflation rates with 5% significance.

Test statistic computed in R: -**2.904427** P value: **0.003679264**

Since our p-value is less than 0.05, we reject the null hypothesis and conclude that our data is not symmetric at the 5% significance level.

d) Null hypothesis of excess kurtosis equal to zero with 5% significance.

K stat computed in R: **2.375595** P Value: **0.01752067**

Since our p-value is greater than 0.05, we fail to reject the null hypothesis. We can conclude the tails are normal.

e) Null hypothesis of normality using Jarque Bera test with 5% significance.

Test statistic: **14.8805** P Value: **0.0005871**

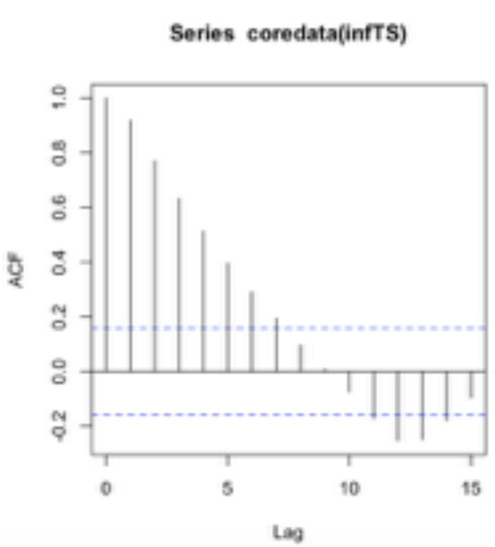
Since our p-value is less than 0.05, we reject the null hypothesis that the data comes from a normal distribution.

f) Time plot for the time series of inflation rates and analysis

Make sure the plot is correctly labeled and titled. Analyze the time trend displayed by the plot, and discuss if data show any striking pattern, such as trends or seasonality?

Looking at the time plot, it appears the inflation rate is fairly inconsistent year by year. There are no signs of seasonality in this plot. The inflation rate went down between 2000 and 2002, but the rate went back up a bit in 2003. There is no distinguishable pattern,

g) Computation and plot of first fifteen lags of the autocorrelation function for inflation. Does TS show evidence of serial correlation? Is the TS weakly stationary?



The TS is weakly stationary*— the lags show a quick decay to zero. It appears as though there is serial correlation.

Autocorrelations of series 'coredata(infTS)', by lag 1.000 0.918 0.771 0.632 0.513 0.395 0.291 0.195 0.096 0.009 -0.075 -0.172 -0.254 -0.249 -0.179 -0.096

*I'm having some trouble with this plot. Is this decay to zero "quick" enough to confirm weak stationarity? Perhaps I need to look at more examples of weakly stationarity ACFs. Would love an extra note from the grader of whether or not this is stationary

h) Ljung Box test to evaluate if inflation rates are serially correlated.

Box-Ljung test data: infTS

X-squared = 291.5, df = 3, p-value < 2.2e-16

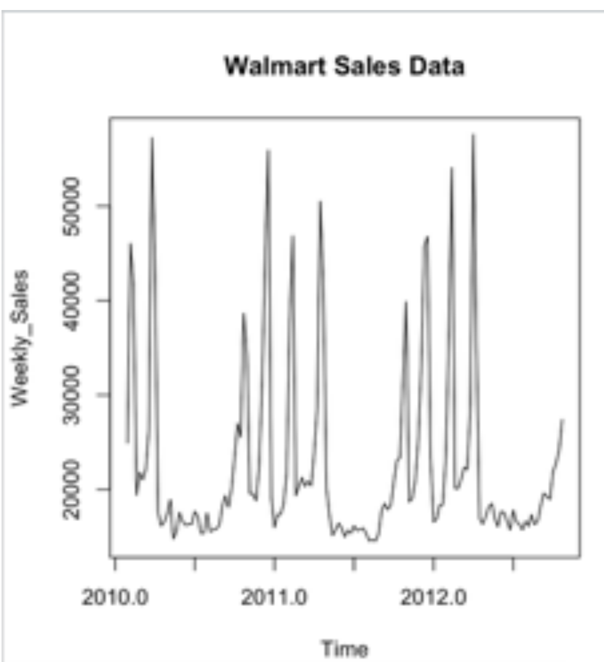
Box-Ljung test data: infTS

X-squared = 373.07, df = 6, p-value < 2.2e-16

Given such small p values, we reject the null hypothesis of independence and conclude the inflation rates are serially correlated.

Problem 3: Walmart Data

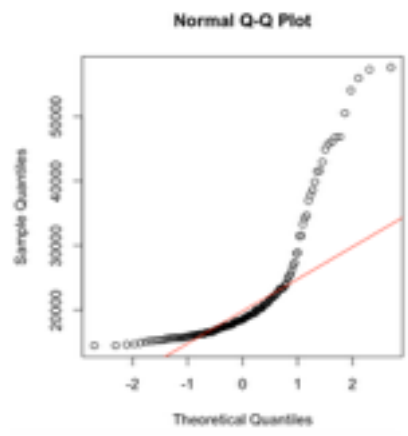
a) Time plot and analysis. Any striking behavior or upward/downward trends?



This data definitely has a seasonality component to it, with each year having similar peaks and valleys in the sales. There are a couple significant peaks in the beginning of each year, and some peaks in the end of each year.

I would expect that the peaks correspond to holidays. Walmart sales likely increase during the week of christmas, and the week of thanksgiving— given black Friday specials.

b) Histogram, normal probability plot, and analysis of distribution of weekly sales.

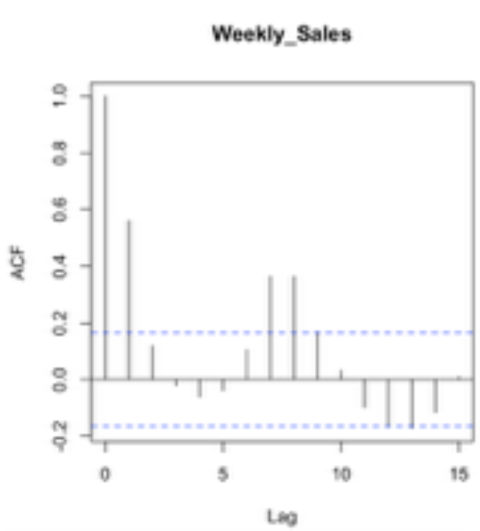


As seen in our histogram and normal probability plot, the Walmart sales data is not normally distributed. The histogram shows the data is left skewed, with a majority of the sales data falling towards the lower lend of the range. There are, however, a significant number of weeks that have much higher sales totals than the norm.

c) Five number summary for weekly sales and discussion of the values.

Minimum	**14537.37**	Maximum	**57592.12**		
1. Quartile	**16494.63**	3. Quartile	**23214.22**	Median	**18535.48**

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



Autocorrelations of series 'coredata(salesTS)', by lag

1.000 0.561 0.120 -0.022 -0.062 -0.039 0.106 0.364 0.363 0.168 0.033 -0.099 -0.167 -0.172 -0.117 0.010

The series shows evidence of serial correlation, with quick decay to zero.

e) Ljung Box test to hypothesis that weekly sales have significant serial correlation.

Box-Ljung test data: salesTS

X-squared = 48.173, df = 3, p-value = 1.957e-10

Box-Ljung test data: salesTS

X-squared = 50.681, df = 6, p-value = 3.432e-09

Given such small p values, we reject the null hypothesis of independence and conclude the inflation rates are serially correlated.

f) Discuss in general the importance of weak stationarity for time series analysis, and explain the methods that are used to analyze whether a TS is stationary.

Stationarity, or the idea that the probability laws that govern the the process don't change over time, is the most important assumption of a time series dataset (Springer, 16). With a weak stationary time series, statistical properties are constant over time. We need these properties to be constant because it allows us to predict future values— we simply assume the future data will have the same mean, variance, autocorrelation etc. as they have had in the past. Our methods to model time series data rely heavily on stationarity.

The most common way to analyze whether a TS is stationary is the ACF plot. In looking at the plot, we can see if the lags decrease quickly to zero, or if they decrease slowly over time. If they decrease quickly, the time series is weakly stationary.

Code:

```
library(fBasics)
library(tseries)
library(zoo)

#set working directory
setwd("/Users/sarahcummings/Documents/csc425")

###################
##  PROBLEM 2

#Save our data to a varaible
#Compute the basic statistics for our data
cpiData= read.csv('cpi_2001_2013.csv')
basicStats(cpiData$inflation)

#Create a histogram
hist(cpiData$inflation, xlab="Inflation", prob=TRUE, main="CPI Histogram")
# add approximating normal density curve
xfit<-seq(min(cpiData$inflation),max(cpiData$inflation),length=40)
yfit<-dnorm(xfit,mean=mean(cpiData$inflation),sd=sd(cpiData$inflation))
lines(xfit, yfit, col="blue", lwd=2)

#Create a normal prob plot
qqnorm(cpiData$inflation)
qqline(cpiData$inflation, col = 2)

inf<-cpiData$inflation

#Skewness test
skew_test = skewness(inf)/sqrt(6/length(inf))
skew_test
#p value
2* (1-pnorm(abs(skew_test)))

# Fat-tail test
k_stat = kurtosis(inf)/sqrt(24/length(inf))
k_stat
#p value
2*(1-pnorm(abs(k_stat)))

#Jarque-Bera normality test
normalTest(inf,method=c("jb"))
```

```
#Create a timeseries plot of CPI
#pull inflation column from data and make a timeseries object.

infTS2<- ts(data = cpiData[3], start = c(2001,1), frequency = 12)

plot(infTS2, main='Time Plot for CPI Inflation')

#Compute Analysis of autocorrelations up to lag 15
acf1<-acf(coredata(infTS), plot=T, lag=15)
#print lags to the console
acf1

#Box test
Box.test(infTS,lag=3,type='Ljung')
Box.test(infTS,lag=6,type='Ljung')

##############################
##  PROBLEM 3

walData= read.csv('walmart_sales.csv')
head(walData)

# Make a timeseries object and a time plot
salesTS<-ts(data=walData[2], start = c(2010,5), frequency = 52)
plot(salesTS, main="Walmart Sales Data")

#Create a histogram
hist(walData$Weekly_Sales, xlab="Sales Data", prob=TRUE, main="Walmart Sales
Histogram")
# add approximating normal density curve
xfit<-seq(min(walData$Weekly_Sales),max(walData$Weekly_Sales),length=40)
yfit<-dnorm(xfit,mean=mean(walData$Weekly_Sales),sd=sd(walData$Weekly_Sales))
lines(xfit, yfit, col="blue", lwd=2)

#Create a normal prob plot
qqnorm(walData$Weekly_Sales)
qqline(walData$Weekly_Sales, col = 2)

#Compute five number summary
basicStats(walData$Weekly_Sales)
head(walData)

#Compute Analysis of autocorrelations up to lag 15
myACF<-acf(coredata(salesTS), plot=T, lag=15)
#pring the lags to consol
myACF
```

```
#Box test
Box.test(salesTS,lag=3,type='Ljung')
Box.test(salesTS,lag=6,type='Ljung')
```