Prediction and Time Series Computer Exercise Week 3

Name: Nguyen Xuan Binh   Student ID: 887799

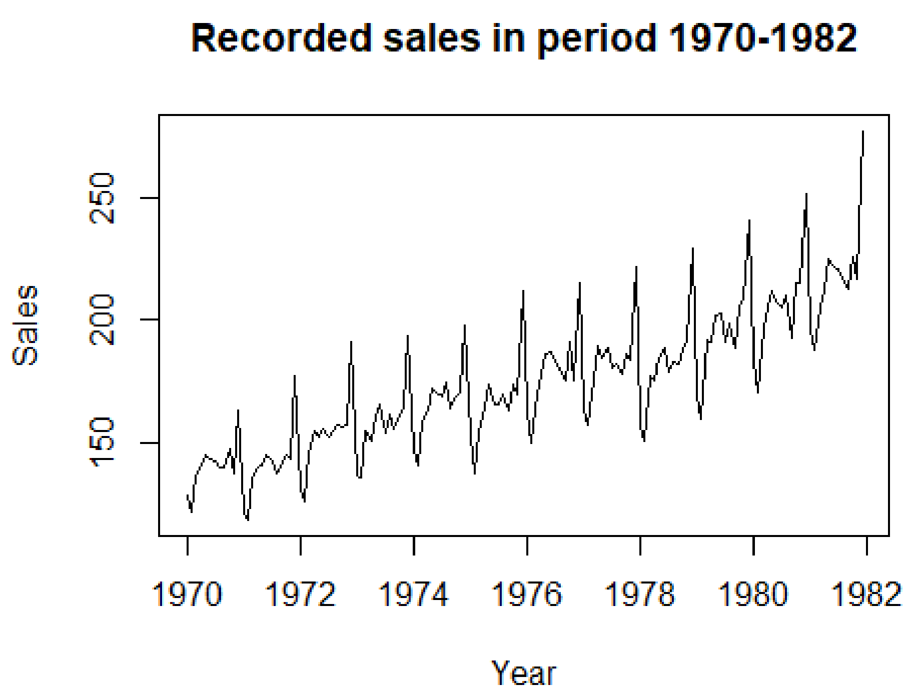
Exercise 3.4

Consider the time series Sales from the file SALES.txt. Apply differencing, seasonal differencing and logarithmic transformations to remove the trend, the seasonality and the increasing variance. Which difference operations did you apply? Visualize both the original and the transformed time series. Hint: Difference operators are given by the function diff in R.

File	Variable	Description	Interval	Length
SALES.txt	Sales	Sales volume of a wholesaler	1 month	$n = 144$

First, let's plot the time series of the sales

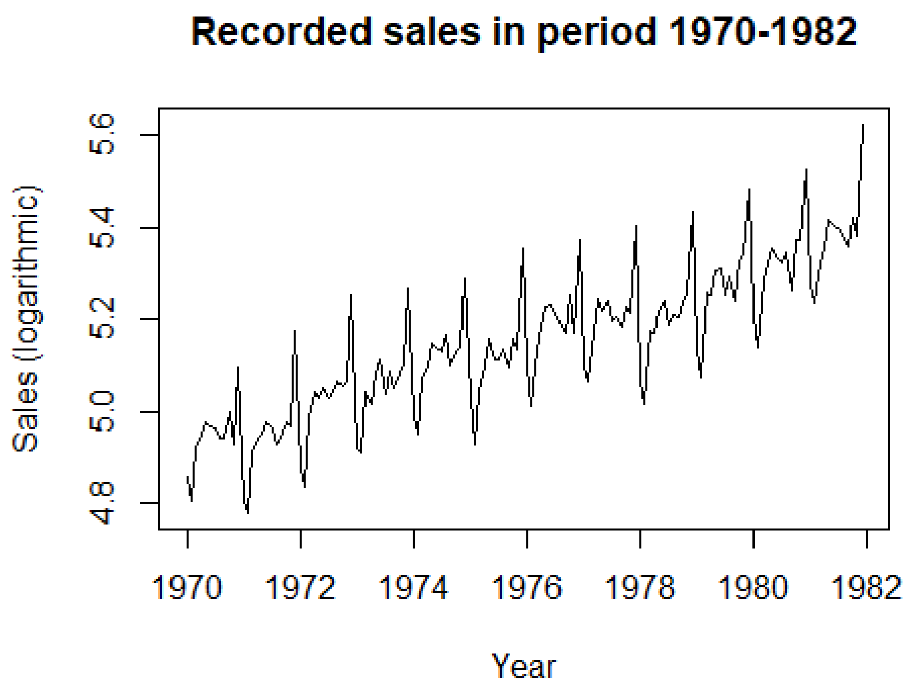
```
SALES <- read.table("SALES.TXT",header=T)
salesTS <- ts(SALES$Sales, start=SALES$Year[1], frequency=12)

plot(salesTS, xlab = "Year" ,ylab = "Sales", main = "Recorded sales in period 1970-1982")
```



Observation: the time series does not seem to be stationary. It has an upward trend as well as seasonality. We need to stationarize the model using differencing techniques Analysis: The trend is really clear in the figure: Increasing variance: the sales vary more and more widely along the time axis Upward increase: although there are varying values, the trend still goes upward Seasonality: Each two years, the sales will drop down nearly to the sales of 2 years ago and then increases again => The cycle is 2 years

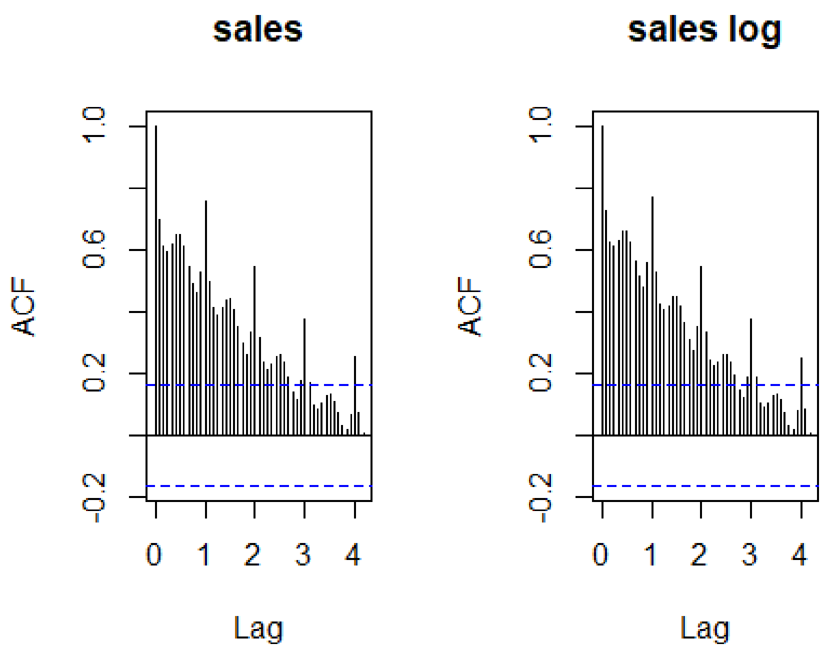
Since the data shows changing variance over time, the first thing we will do is stabilize the variance by applying log transformation. The resulting series will be a linear time series.

```
salesTSLog <- log(salesTS)

plot(salesTSLog, xlab = "Year" ,ylab = "Sales (logarithmic)", main = "Recorde
d sales in period 1970-1982")
```



Observation: the variance seems to be constant now. The trend increases linearly. Let's compare ACF of sales before and after logarithmic transformation

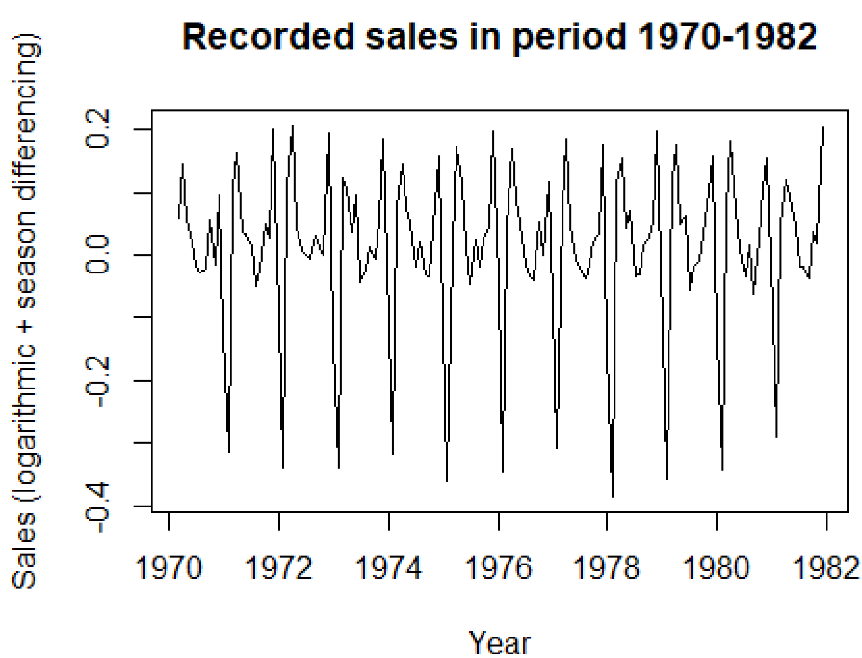
```
par(mfrow = c(1,2))
acf(salesTS, main = "sales", lag.max=50)
acf(salesTSLog, main = "sales log", lag.max=50)
```



Now we can see that seasonal change correlations decrease downwards in a more straight way in a logarithmic map

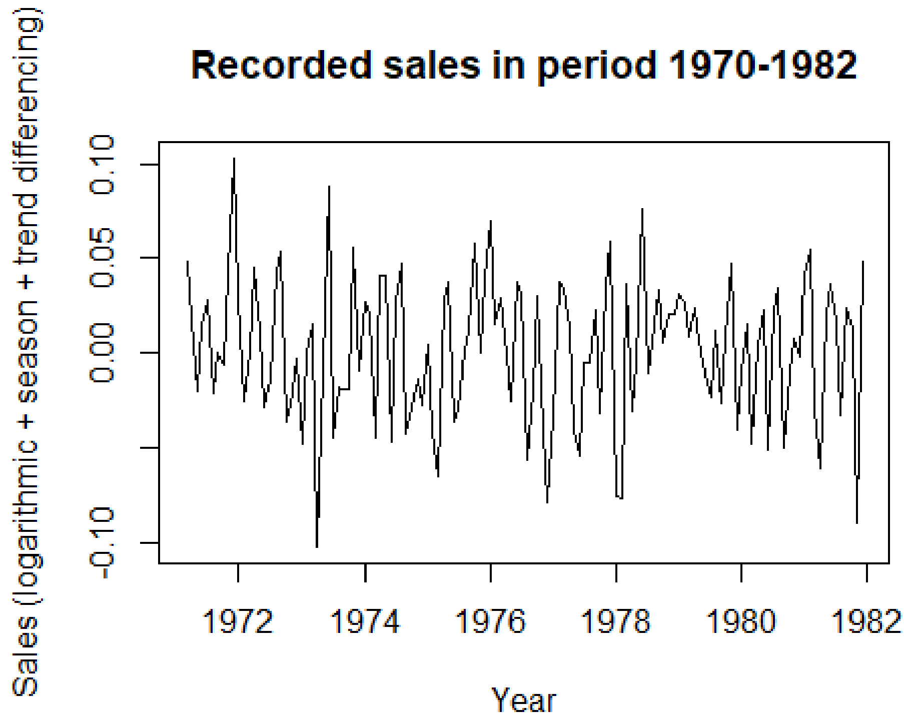
Now let's apply the seasonal diffencing to remove the seasonality trend. As stated about, a season in this data is around 2 years => we will choose lag = 2

```
salesTSLogDiff <- diff(salesTSLog, lag = 2)

plot(salesTSLogDiff, xlab = "Year" ,ylab = "Sales (logarithmic + season diffe
rencing)", main = "Recorded sales in period 1970-1982")
```



Now let's apply the trend diffencing to remove the upward trend. The basic unit is a month in the data file and there are 12 months in a year => we will choose lag = 12

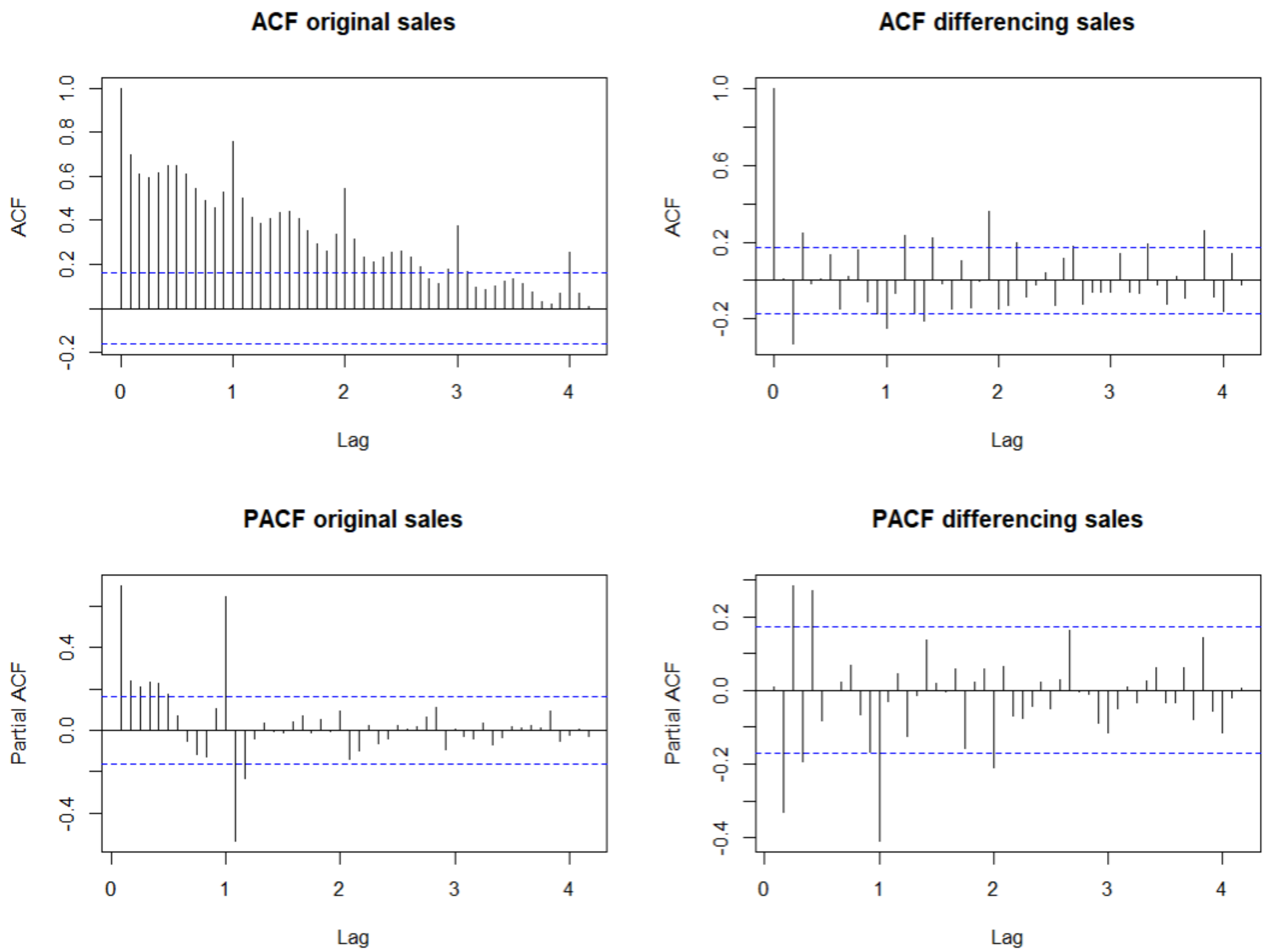
```
salesTSLogDiffSeason <- diff(salesTSLogDiff, lag = 12)

plot(salesTSLogDiffSeason, xlab = "Year" ,ylab = "Sales (logarithmic + season + trend differencing)", main = "Recorded sales in period 1970-1982")
```



Now the time series has been stationarized after being applied differencing, seasonal differencing and logarithmic transformations to remove the trend, the seasonality and the increasing variance

Let's verify with autocorrelation and partial autocorrelation functions of the original sales time series and the sales after differencing

```
par(mfrow=c(2,2))
par(mar = rep(2, 4))
acf(salesTS, lag.max=50, main = "ACF original sales")
acf(salesTSLogDiffSeason, lag.max=50, main = "ACF differencing sales")
pacf(salesTS, lag.max=50, main = "PACF original sales")
pacf(salesTSLogDiffSeason, lag.max=50, main = "PACF differencing sales")
```



The blue lines indicate the level of significance which is 5%. From our observations, the new sales time series after differencing has a higher rate of their ACF and PACF values distributed within the significance line than the original. The season patterns are not clearly visible in the differencing model as well => The sales time series model becomes stationary after removing the trends, seasons and increasing variance