Tutorial Sheet 10 Solutions

Q10.1

Consider the Alert, Canada, monthly carbon dioxide time series in the file named "co2."

(a) Fit a deterministic seasonal means plus linear time trend model to these data. Are any of the regression coefficients statistically significant?

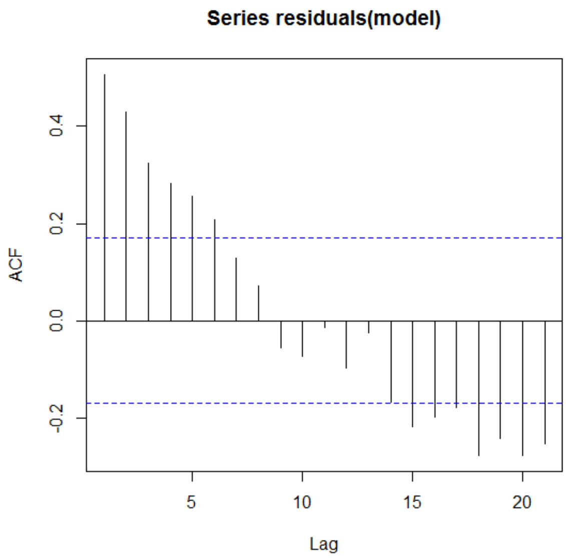
```
data(co2)
month.=season(co2)
trend=time(co2)
model=lm(co2~month.+trend)
summary(model)

#Call:
#lm(formula = co2 ~ month. + trend)

#Residuals:
#     Min       1Q   Median       3Q      Max
#-1.73874 -0.59689 -0.06947  0.54086  2.15539

#Coefficients:
#                  Estimate Std. Error t value Pr(>|t|)
#(Intercept)     -3290.5412    44.1790 -74.482  < 2e-16 ***
#month.February      0.6682     0.3424   1.952 0.053320 .
#month.March         0.9637     0.3424   2.815 0.005715 **
#month.April         1.2311     0.3424   3.595 0.000473 ***
#month.May           1.5275     0.3424   4.460 1.87e-05 ***
#month.June         -0.6761     0.3425  -1.974 0.050696 .
#month.July         -7.2851     0.3426 -21.267  < 2e-16 ***
#month.August      -13.4414     0.3426 -39.232  < 2e-16 ***
#month.September   -12.8205     0.3427 -37.411  < 2e-16 ***
#month.October      -8.2604     0.3428 -24.099  < 2e-16 ***
#month.November     -3.9277     0.3429 -11.455  < 2e-16 ***
#month.December     -1.3367     0.3430  -3.897 0.000161 ***
```

```
#trend                1.8321     0.0221  82.899  < 2e-16 ***

#Residual standard error: 0.8029 on 119 degrees of freedom
#Multiple R-squared: 0.9902,     Adjusted R-squared: 0.9892
#F-statistic: 997.7 on 12 and 119 DF,  p-value: < 2.2e-16
```

All the regression coefficients are statistically significant except for the seasonal effects for February and June whose $p-$ values are > 0.05.

(b) Calculate and interpret the sample auto-correlation function (acf) of the residuals from this model.

```
acf(residuals(model))
```



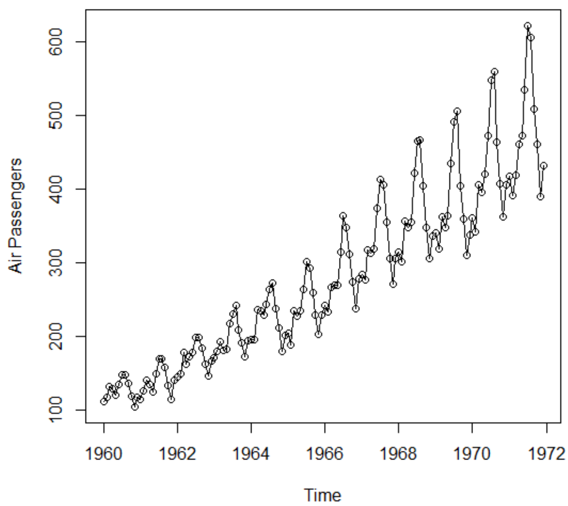
This deterministic trend model has not captured the auto-correlation in this time series. The seasonal Arima model shown in TSLecture10 is a better model for this series.

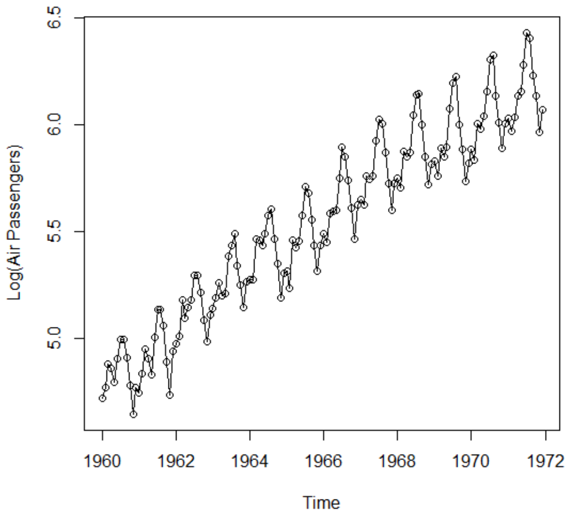
Q10.2

The monthly airline passenger time series are in the file named “airpass.”

(a) Display and interpret the time series plots of both the original series and the logarithms of the series.

```
data(airpass)
plot(airpass,type="o",
ylab="Air Passengers")
plot(log(airpass),type="o",
ylab="Log(Air Passengers)")
```

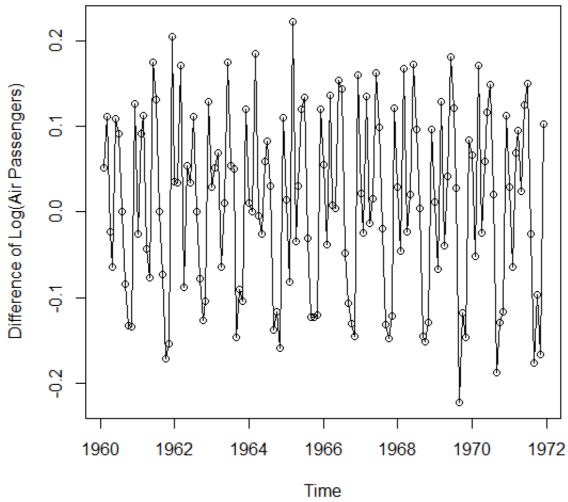




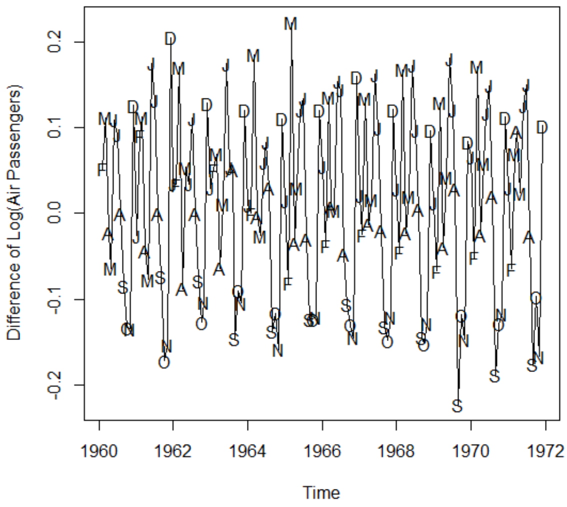
The graph of the logarithms displays a much more constant variation around the upward"trend".

(b) Display and interpret the time series plot of the difference of the logarithms of the series.

```
plot(diff(log(airpass)),type="o",
ylab="Difference of Log(Air Passengers)")
```



The series appears to be stationary, but the plot could be hiding seasonality.

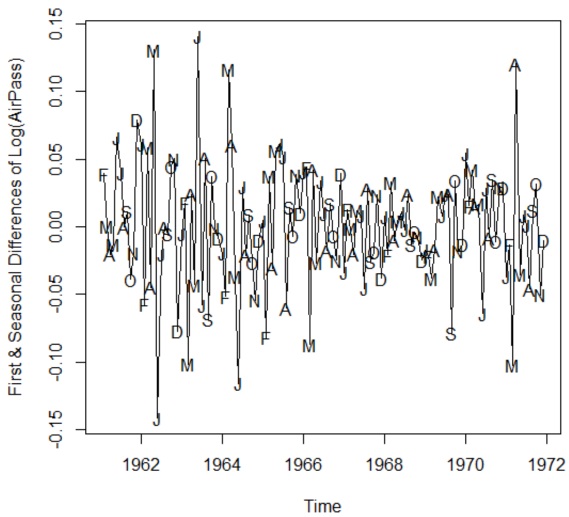



```
plot(diff(log(airpass)),type="l",
ylab="Difference of Log(Air Passengers)")
points(diff(log(airpass)),x=time(diff(log(airpass))),
pch=as.vector(season(diff(log(airpass)))))
```

Seasonality can be seen with Septembers, Octobers and Novembers being mostly low and Decembers mostly high.

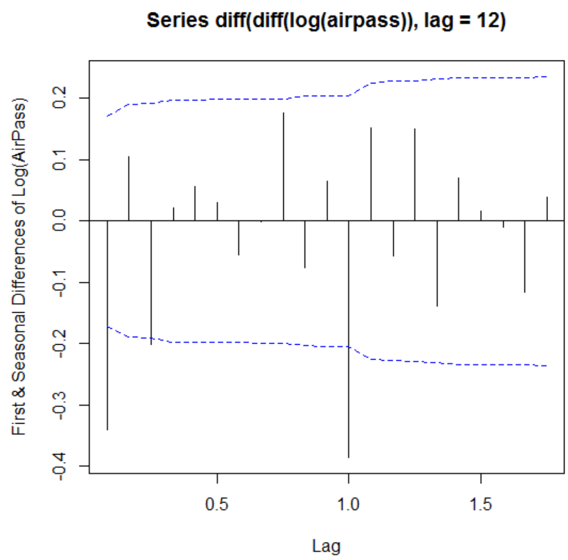
(c) Display and interpret the time series plot of the seasonal difference of the first difference of the logarithms of the series.

```
plot(diff(diff(log(airpass)),lag=12),type="l",
ylab="First & Seasonal Differences of Log(AirPass)")
points(diff(log(airpass)),x=time(diff(diff(log(airpass)),lag=12)),
pch=as.vector(season(diff(diff(log(airpass)),lag=12))))
```



The plot is done with plotting symbols. The seasonality is much less obvious now.

(d) Calculate and interpret the sample acf of the seasonal difference of the first difference of the logged series.




```
acf(diff(diff(log(airpass)),lag=12),ci.type="ma",
ylab="First & Seasonal Differences of Log(AirPass)")
```

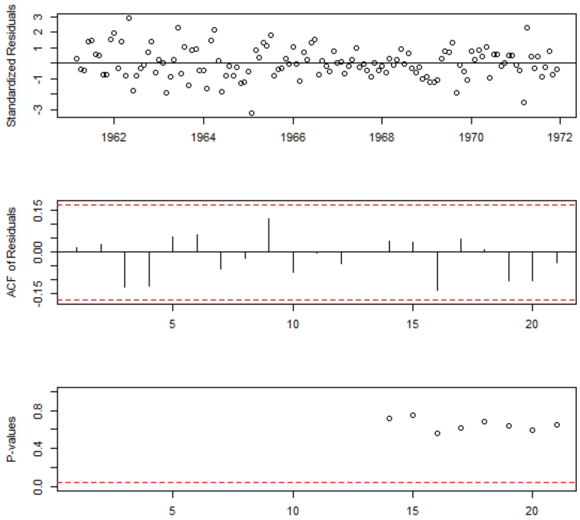
Although there is a significant auto-correlation at lag 3, the most prominent auto-correlations are at lags 1 and 12 the chosen model seems like a reasonable choice to investigate.

(e) Fit an Arima$(0, 1, 1) \times (0, 1, 1)_{12}$ model to the logged series.

```
model=arima(log(airpass),order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=12))
model

#Call:
#arima(x = log(airpass), order = c(0, 1, 1),
# seasonal = list(order = c(0, 1,1), period = 12))

#Coefficients:
#          ma1     sma1
#      -0.4018  -0.5569
#s.e.   0.0896   0.0731

#sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -485.4
```

Both the seasonal and non-seasonal ma parameters are significant.

(f) Investigate diagnostics for this model, including auto-correlation and normality of residuals.

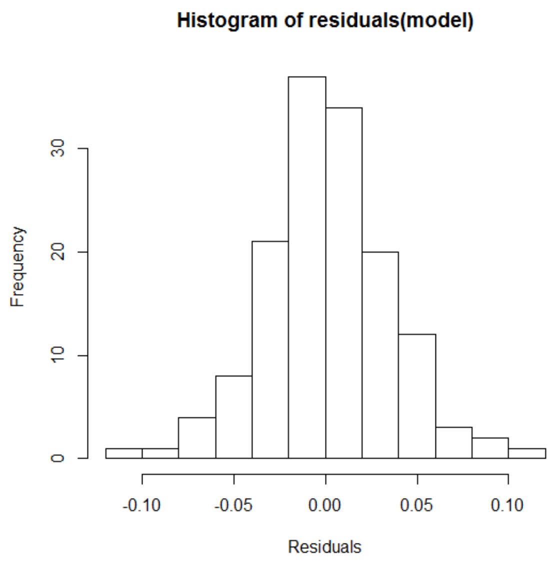


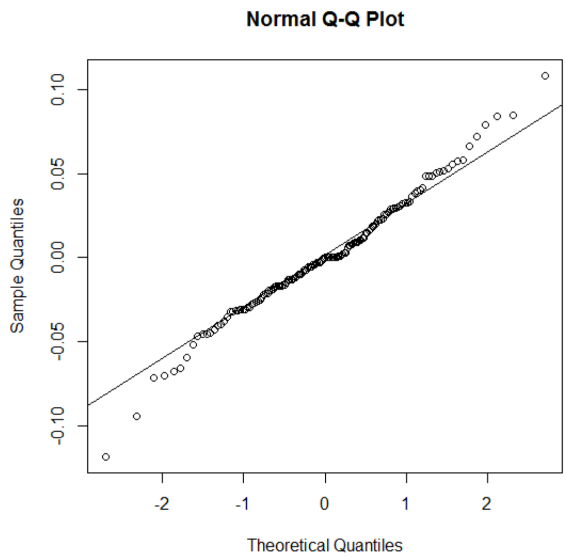

```
tsdiag(model)
```

None of these three plots indicate difficulties with the model.

There are no outliers and little auto-correlation in the residuals, individually and jointly.

```
hist(residuals(model), xlab="Residuals")
qqline(residuals(model))
qqnorm(residuals(model))
shapiro.test(residuals(model))

#        Shapiro-Wilk normality test

#data:  residuals(model)
#W = 0.9864, p-value = 0.1674
```

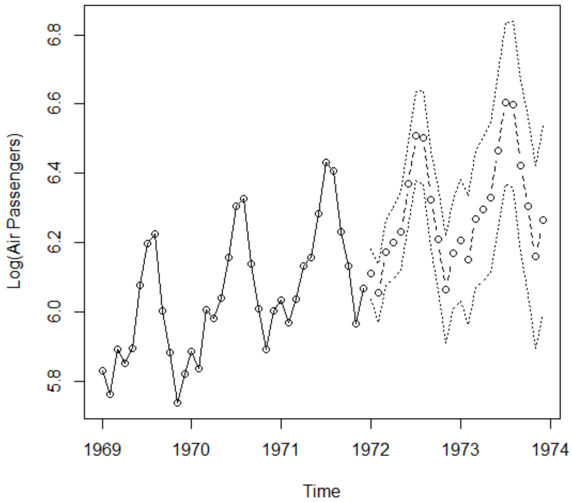




The distribution of the residuals is quite symmetric but the QQ plot indicates that the tails are lighter than a Normal distribution.

Based on the Shapiro-Wilk test, Normality cannot be rejected.

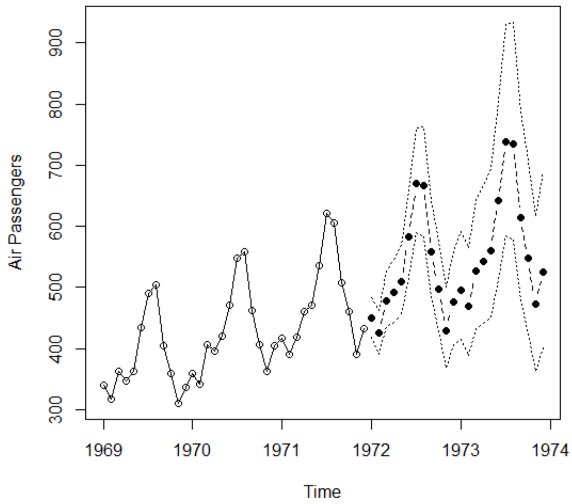
(g) Produce forecast, with limits, for this series with a lead time of two years.




```
plot(model,n1=c(1969,1),n.ahead=24,pch=19,ylab="Log(Air Passengers)")
```

The forecasts follow the seasonal and upward "trend" of the time series well. The forecast limits provide a clear measure of the uncertainty in the forecasts.

For completeness, here are the forecasts and limits in original terms.



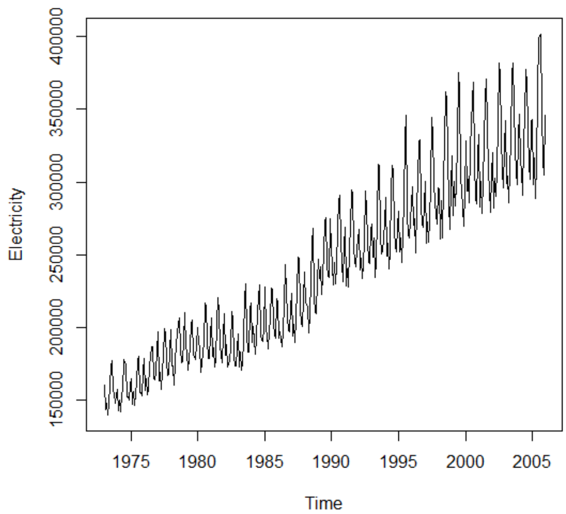

```
plot(model,n1=c(1969,1),n.ahead=24,pch=19,
ylab="Air Passengers",transform=exp)
```

Q10.3

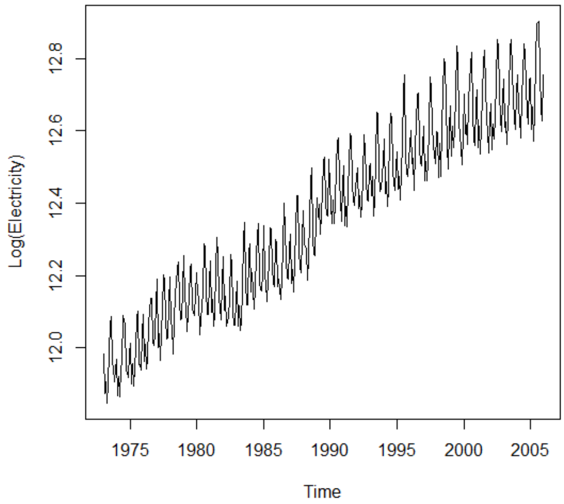
Consider the monthly electricity generation in the U.S. time series in the file named “electricity.”

(a) Display and interpret the time series plots of both the original series and the logarithms of the series.

```
data(electricity)
plot(electricity,ylab="Electricity")
plot(log(electricity),ylab="Log(Electricity)")
```

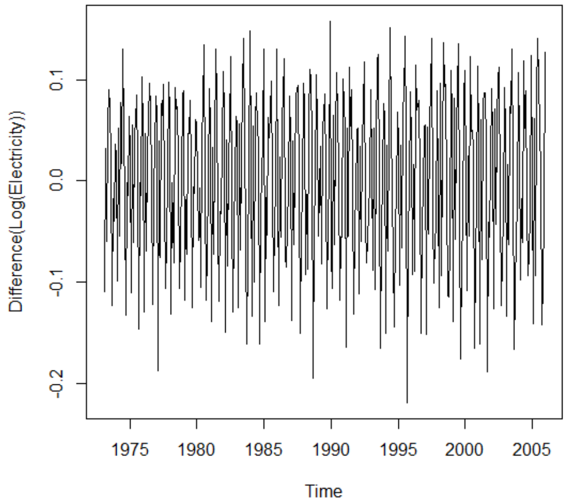


The increasing variance suggests the taking of logs.



The variance is now constant across the time series.

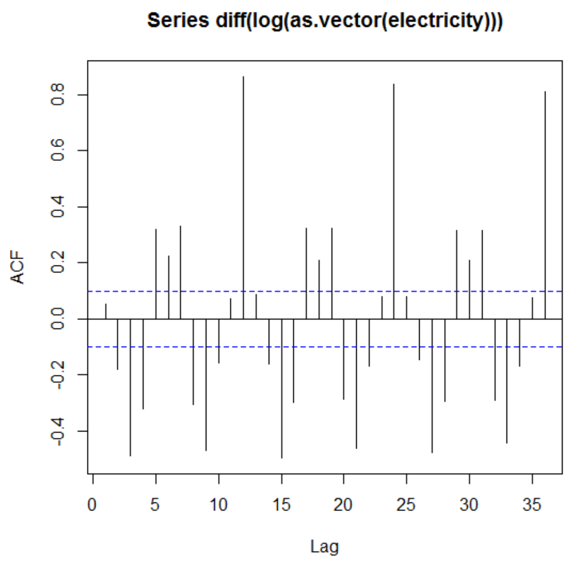
(b) Display and interpret the time series plot of the difference of the logarithms of the series.




```
plot(diff(log(electricity)),
ylab="First Differences of Log(Electricity)")
```

The data look stationary, but the seasonality will still have to be investigated further and modelled.

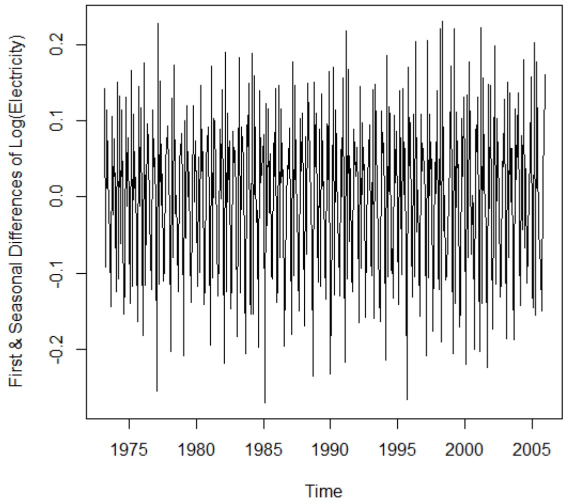
(c) Calculate and interpret the sample acf of the first difference of the logged series.




```
acf(diff(log(as.vector(electricity))),lag.max=36)
```

The very strong auto-correlations at lags 12, 24 and 36 point out the substantial seasonality in this time series.

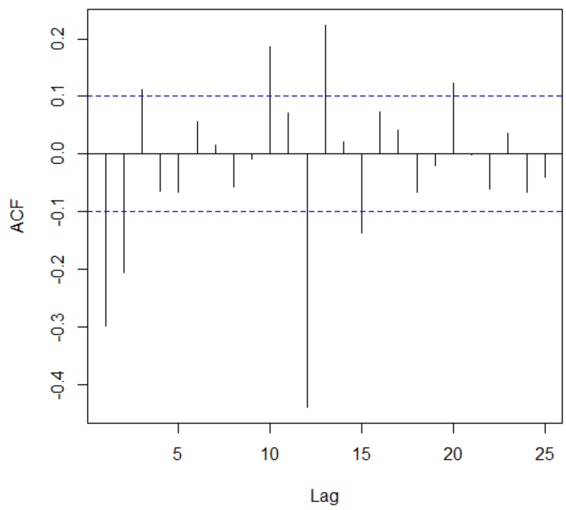
(d) Display and interpret the time series plot of the seasonal difference of the first difference of the logarithms of the series.




```
plot(diff(diff(log(electricity)),lag=12),
ylab="First & Seasonal Differences of Log(Electricity)")
```

The time series plot appears stationary.

(e) Calculate and interpret the sample acf of the seasonal difference of the first difference of the logged series. What model would you suggest for this transformed series?




```
acf(diff(diff(log(as.vector(electricity)),lag=12),lag.max=36)
```

After seasonal differencing, the strong auto-correlations at lags 24 and 36 have disappeared.

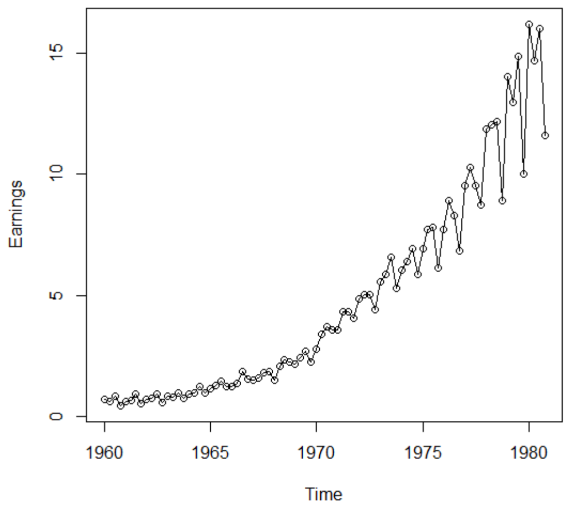
Perhaps a stationary model can now be entertained.

The model Arima$(0,1,2) \times (0,1,1)_{12}$ to the logged series could be tried.

Q10.4

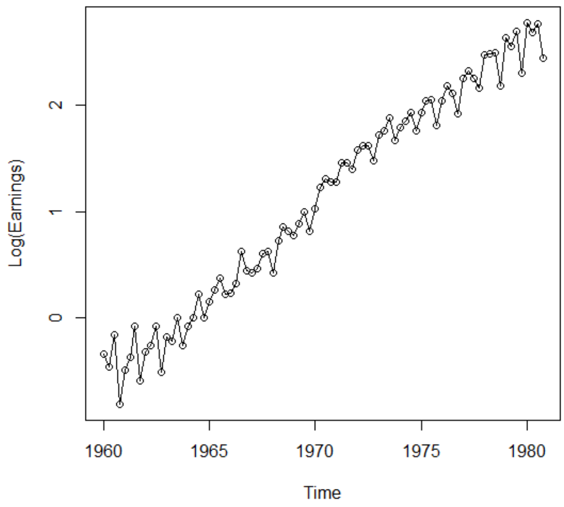
Consider the quarterly earnings per share for 1960-1980 of the U.S. company Johnson & Johnson in the file named “JJ”.

(a) Display and interpret the time series plots of both the original series and the logarithms of the series.




```
data(JJ)
plot(JJ,ylab="Earnings", type="o")
```

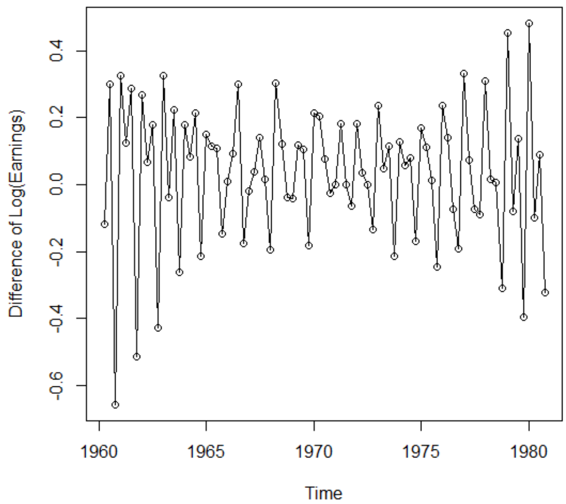
The increased variability means we should try to model the log of the series.




```
plot(log(JJ),ylab="Log(Earnings)",type="o")
```

The variance looks more constant.

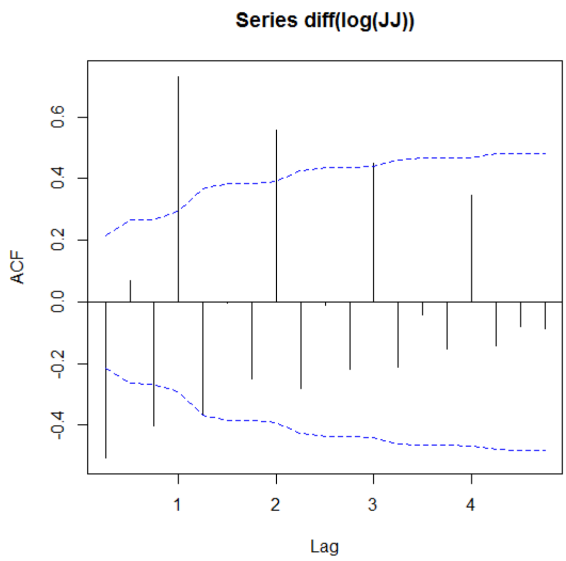
(b) Display and interpret the time series plot of the difference of the logarithms of the series.




```
plot(diff(log(JJ)),ylab="Difference of Log(Earnings)",type="o")
```

There is less variability in the middle of the series than might be expected for a stationary series.

(c) Calculate and interpret the sample acf of the first difference of the logged series.

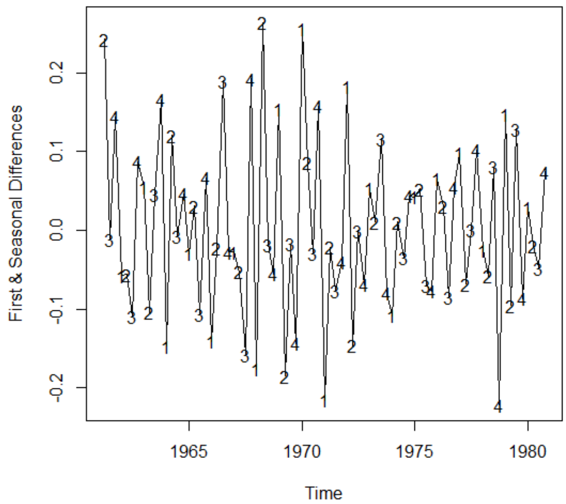



```
acf(diff(log(JJ)),ci.type="ma")
```

In this quarterly series, the strongest auto-correlations are at seasonal lags 4, 8, 12 and 16.

We thus need to address the seasonality in this series.

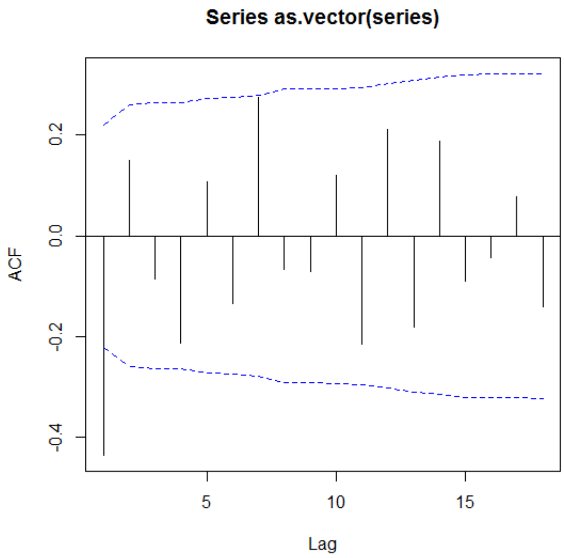
(d) Display and interpret the time series plot of the seasonal difference of the first difference of the logarithms of the series. For quarterly data, a season is of length 4.




```
series=diff(diff(log(JJ)),lag=4)
plot(series,ylab="First & Seasonal Differences", type="l")
points(y=series,x=time(series),pch=as.vector(season(series)))
```

The various quarters seem to be quite randomly distributed among high, middle and low values, so that most of the seasonality is accounted for in the seasonal difference.

(e) Calculate and interpret the sample acf of the seasonal difference of the first difference of the logged series.




```
acf(as.vector(series),ci.type="ma")
```

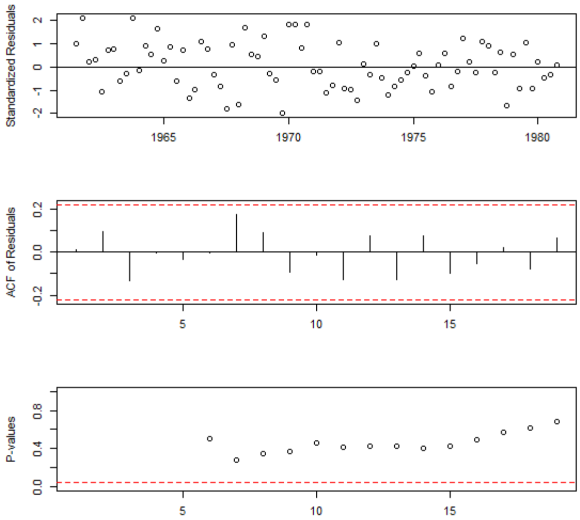
The only significant auto-correlations are at lags 1 and 7. Lag 4 (the quarterly lag) is nearly significant.

(f) Fit an Arima$(0,1,1) \times (0,1,1,)_4$ model to the logged series.

```
model=arima(log(JJ),
            order=c(0,1,1),
            seasonal=list(order=c(0,1,1), period=4)
            )
model

#Call:
#arima(x = log(JJ), order = c(0, 1, 1),
                    seasonal = list(order = c(0, 1, 1),
                    period = 4))

#Coefficients:
#          ma1     sma1
#      -0.6809  -0.3146
#s.e.   0.0982   0.1070

#sigma^2 estimated as 0.007931:  log likelihood = 78.38,  aic = -152.75
```

Both the seasonal and non-seasonal ma parameters are significant in this model.

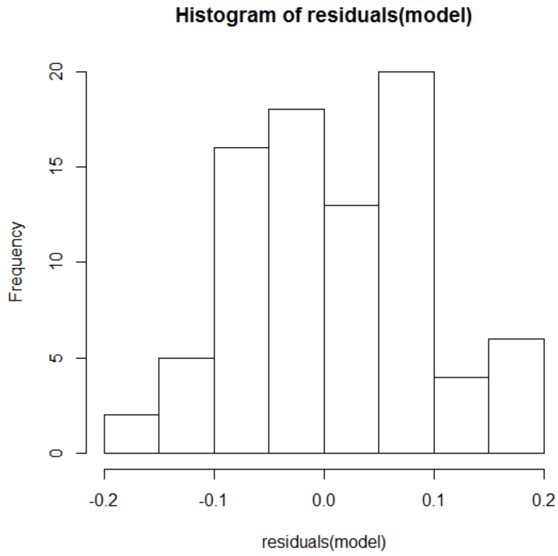
(g) Investigate diagnostics for this model, including auto-correlation and normality of residuals.

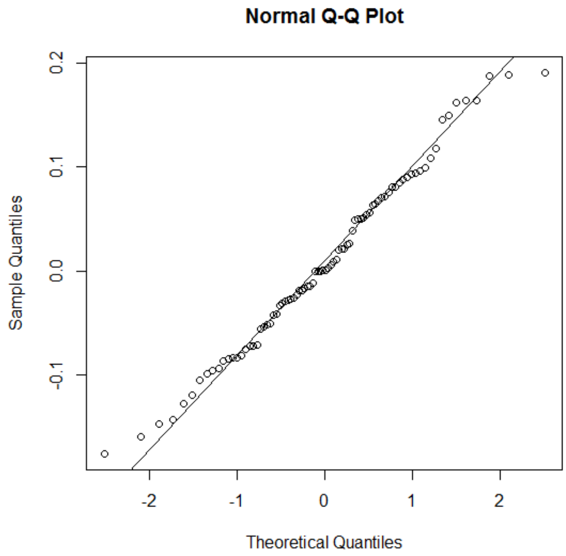


`tsdiag(model)`

These diagnostic plots do not show any inadequacies with the model.

No outliers are detected and there is little auto-correlation in the residuals.



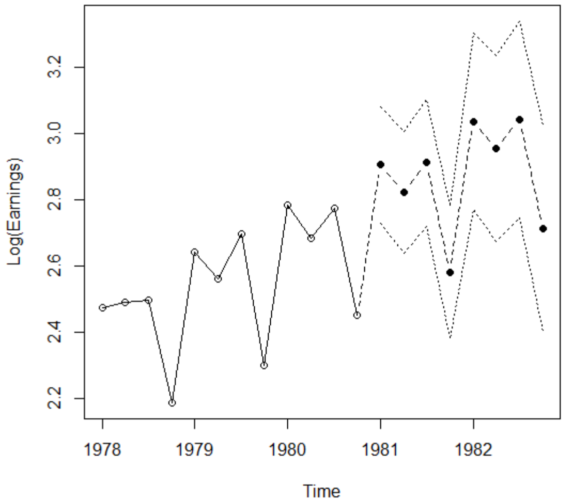



```
hist(residuals(model))
qqline(residuals(model))
qqnorm(residuals(model))
shapiro.test(residuals(model))
```

```
#        Shapiro-Wilk normality test

#data:  residuals(model)
#W = 0.9858, p-value = 0.4891
```

Normality of the residuals looks like a good assumption.

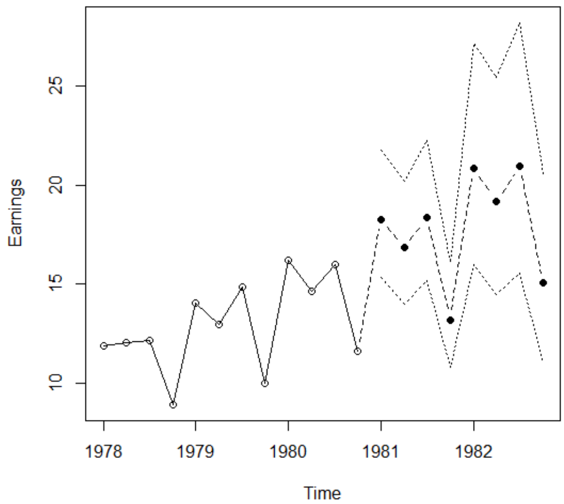
(h) Produce forecast, with limits, for this series with a lead time of two years.




```
plot(model,n1=c(1978,1),n.ahead=8,pch=19,ylab="Log(Earnings)")
```

The forecasts follow the general pattern of seasonality and "trend" in the earnings series and the forecast limits give a good indication of the confidence in these forecasts.

Here is the plot of the forecasts in original terms which enables the forecasts to be more easily understood.



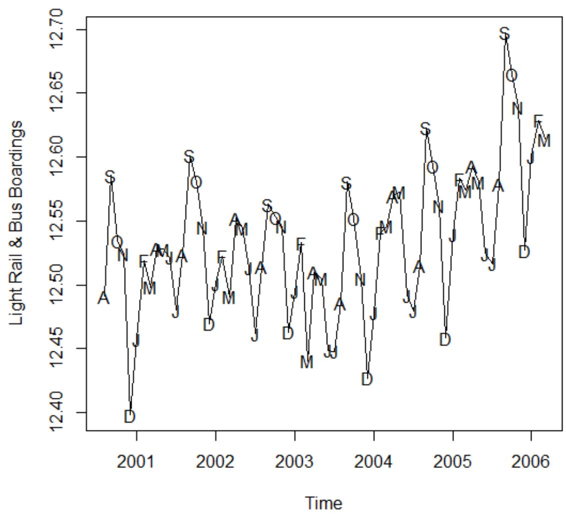

```
plot(model,n1=c(1978,1),n.ahead=8,pch=19,
ylab="Earnings",transform=exp)
```

Q10.5

The file named "boardings" contains monthly data on the number of people who boarded transit vehicles (mostly light rail trains and city buses) in Denver, Colorado for August 2000 through December 2005.

(a) Produce the time series plot for these data using plotting symbols to help assessment of seasonality. Is a stationary model reasonable?

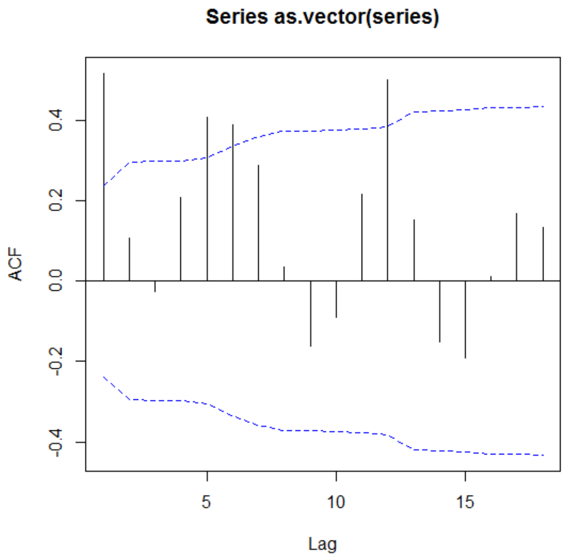
```
data(boardings)
boardings
# only the first column is required
series=boardings[,1]
plot(series,type="l", ylab="Light Rail & Bus Boardings")
points(series,x=time(series,pch=as.vector(season(series)))
```



There is substantial seasonality in this series. Decembers are generally low due to the holidays and Septembers are usually quite high due to the start up of school term.

There may also be a gradual upward "trend" that may need to be modelled with some kind of non-stationarity.

(b) Calculate and plot the sample acf for this series. At which lags are there significant auto-correlations?




```
acf(as.vector(series),ci.type="ma")
```

There are significant auto-correlations at lags 1, 5, 6 and 12.

Perhaps the model will incorporate these effects.

(c) Fit an ARMA(0,3)$\times(1,0)_{12}$ model to these data and assess the significance of the estimated coefficients.

```
model=arima(series,
            order=c(0,0,3),
            seasonal=list(order=c(1,0,0),period=12)
            )
model

#Call:
#arima(x = series, order = c(0, 0, 3),
                   seasonal = list(order = c(1, 0, 0),
                   period = 12))

#Coefficients:
#         ma1     ma2     ma3    sar1  intercept
#      0.7290  0.6116  0.2950  0.8776    12.5455
#s.e.  0.1186  0.1172  0.1118  0.0507     0.0354

#sigma^2 estimated as 0.0006542:  log likelihood = 143.54,  aic = -277.09
```

All of these coefficients are significant at the usual significance levels.

(d) Overfit an ARMA(0,4)$\times(1,0)_{12}$ model to these data and interpret the results.

```
model2=arima(series,
            order=c(0,0,4),
            seasonal=list(order=c(1,0,0),period=12)
            )
model2

#Call:
#arima(x = series, order = c(0, 0, 4),
                   seasonal = list(order = c(1, 0, 0),
                   period = 12))

#Coefficients:
#         ma1     ma2     ma3     ma4    sar1  intercept
#      0.7277  0.6686  0.4244  0.1414  0.8918    12.5459
#s.e.  0.1212  0.1327  0.1681  0.1228  0.0445     0.0419

#sigma^2 estimated as 0.0006279:  log likelihood = 144.22,  aic = -276.45
```

In this model, the added coefficient, ma4, is not statistically significant.

Also, the coefficients in common have changed very little, especially in light of their standard errors.

Finally, the AIC is slightly better(lower) for the simpler model.