Q.1) ① $Y_t = \mu + \frac{1}{2}e_t - \frac{1}{2}e_{t-1}$

$\bar{Y} = \mu + \frac{1}{2n}\sum_{t=1}^{n}(e_t - e_{t-1})$

$= \mu + \frac{1}{2n}(e_n - e_{n-1} + e_{n-1} - e_{-2} + \dots + e_1 - e_0)$

$= \mu + \frac{1}{2n}(e_n - e_0)$

$\rightarrow Var(\bar{Y}) = \frac{1}{4n^2}Var(e_n - e_0) = \frac{1}{4n^2}(Var(e_n) + Var(e_0))$

$= \frac{1}{4n^2} \times 2\sigma_e^2 = \frac{\sigma_e^2}{2n^2}$

② $Y_t = \mu + e_t$

$\bar{Y} = \mu + \frac{1}{n}\sum_{t=1}^{n} e_t$

$Var(\bar{Y}) = \frac{1}{n^2}(Var(e_0) \times n) = \frac{1}{n}\sigma_e^2$

$\rightarrow \frac{\sigma^2}{2n^2}$: denominator is $2n^2$, not n. The negative autocorrelation at lag one makes it easier to estimate the process mean when compared with estimating the mean of a white noise process.

Q.2) • $Y = \mu + e_t + 2e_{t-1}$

$\bar{Y} = \mu + \frac{1}{n}\left[\sum_{t=1}^{n}(e_t + 2e_{t-1})\right]$

$= \mu + \frac{1}{n}[2e_0 + 3e_1 + 3e_2 + \dots + 3e_{n-1} + e_n]$

$= \mu + \frac{1}{n}[e_n + 2e_0 + 3\sum_{t=1}^{n-1} e_t]$

$Var(\bar{Y}) = \frac{1}{n^2}[\sigma_e^2 + 4\sigma_e^2 + 9\cdot(n-1)\sigma_e^2] = \left(\frac{9n-4}{n^2}\right)\sigma_e^2 \approx \frac{9\cancel{n}}{n^{\cancel{2}}}\sigma_e^2$

• $Y = \mu + e_t$

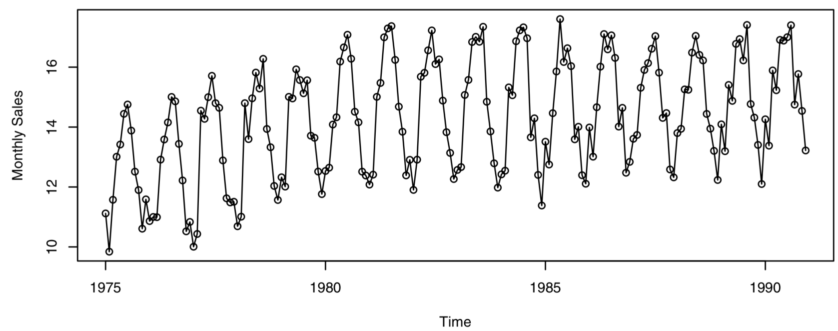
$\bar{Y} = \mu + \frac{1}{n}\sum_{t=1}^{n} e_t$

$Var(\bar{Y}) = \frac{1}{n^2}n\sigma_e^2 = \frac{1}{n}\sigma_e^2$

$\rightarrow Var(\bar{Y}) \approx \frac{9}{n}\sigma_e^2$ is approximately 9 times larger. The positive autocorrelation at lag one makes it more difficult to estimate the process mean compared with estimating the mean of a white noise process.

Q.3)

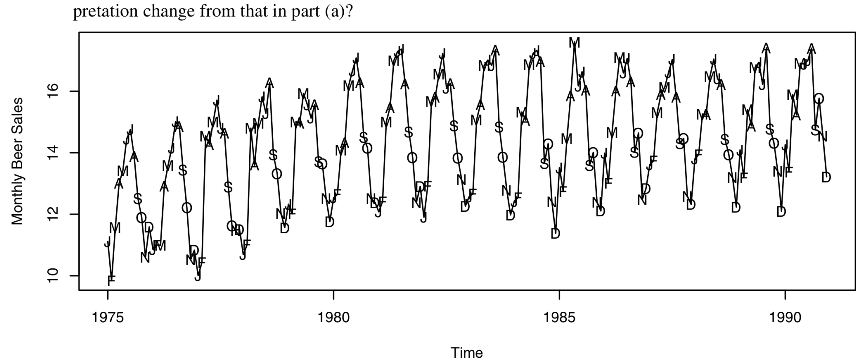
(a) Display the time series plot for these data and interpret the plot.




```
> data(beersales); plot(beersales,ylab='Monthly Sales',type='o')
```

In addition to a possible seasonality in the series, there is a general upward "trend" in the first part of the series. However, this effect "levels off" in the latter years.

(b) Now construct a time series plot that uses separate plotting symbols for the various months. Does your interpretation change from that in part (a)?




```
> plot(beersales,ylab='Monthly Beer Sales',type='l')
> points(y=beersales,x=time(beersales), pch=as.vector(season(beersales)))
```

Now the seasonality is quite clear with higher sales in the summer months and lower sales in the winter.

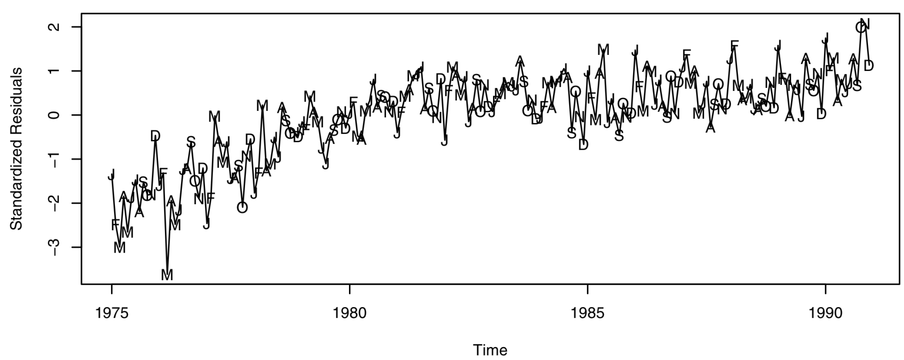
(c) Use least squares to fit a seasonal-means trend to this time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.

```
> month.=season(beersales);beersales.lm=lm(beersales~month.);summary(beersales.lm)
```

```
Call:
lm(formula = beersales ~ month.)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5745 -0.4772  0.1759  0.7312  2.1023

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     12.48568    0.26392  47.309  < 2e-16 ***
month.February  -0.14259    0.37324  -0.382 0.702879
month.March      2.08219    0.37324   5.579 8.77e-08 ***
month.April      2.39760    0.37324   6.424 1.15e-09 ***
month.May        3.59896    0.37324   9.643  < 2e-16 ***
month.June       3.84976    0.37324  10.314  < 2e-16 ***
month.July       3.76866    0.37324  10.097  < 2e-16 ***
month.August     3.60877    0.37324   9.669  < 2e-16 ***
month.September  1.57282    0.37324   4.214 3.96e-05 ***
month.October    1.25444    0.37324   3.361 0.000948 ***
month.November  -0.04797    0.37324  -0.129 0.897881
month.December  -0.42309    0.37324  -1.134 0.258487
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.056 on 180 degrees of freedom
Multiple R-Squared: 0.7103,     Adjusted R-squared: 0.6926
F-statistic: 40.12 on 11 and 180 DF,  p-value: < 2.2e-16
```

This model leaves out the January term so all of the other effects are in comparison to January. The multiple R-squared is rather large at 71% and all the terms except November, December, and February are significantly different from January.

(d) Construct and interpret the time series plot of the standardized residuals from part (c). Be sure to use proper plotting symbols to check on seasonality in the standardized residuals.




```
> plot(y=rstudent(beersales.lm),x=as.vector(time(beersales)),type='l',
   ylab='Standardized Residuals')
> points(y=rstudent(beersales.lm),x=as.vector(time(beersales)),
   pch=as.vector(season(beersales)))
```

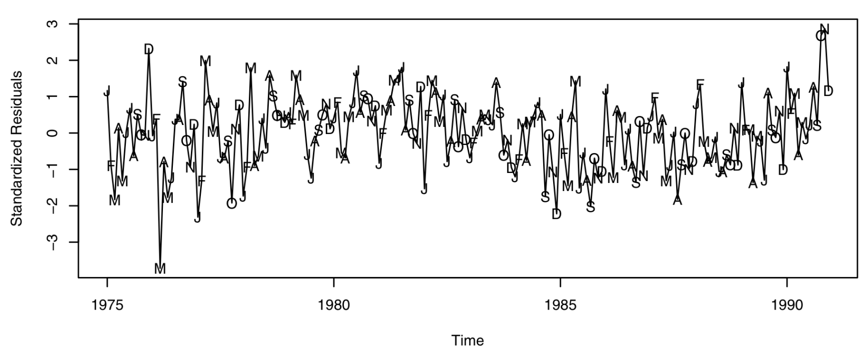
Display this plot full screen to see the detail. However, it is clear that this model does not capture the structure of this time series and we proceed to look for a more adequate model.

(e) Use least squares to fit seasonal-means plus quadratic time trend to the beer sales time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.

```
> beersales.lm2=lm(beersales~month.+time(beersales)+I(time(beersales)^2))
> summary(beersales.lm2)
```

```
Call:
lm(formula = beersales ~ month. + time(beersales) + I(time(beersales)^2))

Residuals:
     Min       1Q   Median       3Q      Max
-2.03203 -0.43118  0.04977  0.34509  1.57572

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -7.150e+04  8.791e+03  -8.133 6.93e-14 ***
month.February       -1.579e-01  2.090e-01  -0.755  0.45099
month.March           2.052e+00  2.090e-01   9.818  < 2e-16 ***
month.April           2.353e+00  2.090e-01  11.256  < 2e-16 ***
month.May             3.539e+00  2.090e-01  16.934  < 2e-16 ***
month.June            3.776e+00  2.090e-01  18.065  < 2e-16 ***
month.July            3.681e+00  2.090e-01  17.608  < 2e-16 ***
month.August          3.507e+00  2.091e-01  16.776  < 2e-16 ***
month.September       1.458e+00  2.091e-01   6.972 5.89e-11 ***
month.October         1.126e+00  2.091e-01   5.385 2.27e-07 ***
month.November       -1.894e-01  2.091e-01  -0.906  0.36622
month.December       -5.773e-01  2.092e-01  -2.760  0.00638 **
time(beersales)       7.196e+01  8.867e+00   8.115 7.70e-14 ***
I(time(beersales)^2) -1.810e-02  2.236e-03  -8.096 8.63e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5911 on 178 degrees of freedom
Multiple R-Squared: 0.9102,     Adjusted R-squared: 0.9036
F-statistic: 138.8 on 13 and 178 DF,  p-value: < 2.2e-16
```

(f) Construct and interpret the time series plot of the standardized residuals from part (e). Again use proper plotting symbols to check for any remaining seasonality in the residuals.

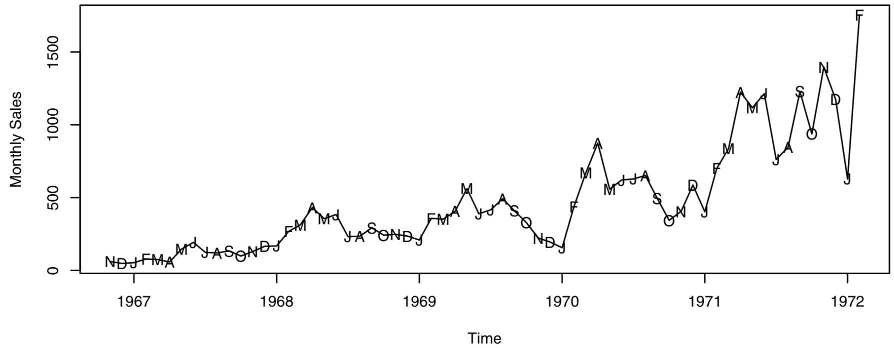



```
> plot(y=rstudent(beersales.lm2),x=as.vector(time(beersales)),type='l',
>   ylab='Standardized Residuals')
> points(y=rstudent(beersales.lm2),x=as.vector(time(beersales)),
   pch=as.vector(season(beersales)))
```

This model does much better than the previous one but we would be hard pressed to convince anyone that the underlying quadratic "trend" makes sense. Notice that the coefficient on the square term is negative so that in the future sales will decrease substantially and even eventually go negative!

Q.4)

(a) Display and interpret the time series plot for these data.



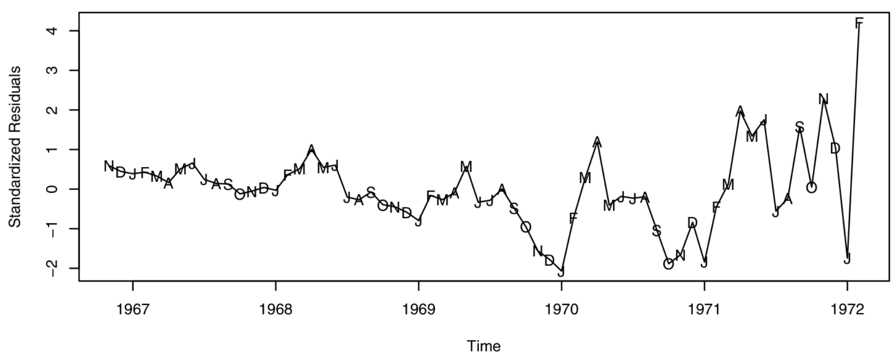

```
> data(winnebago); plot(winnebago,ylab='Monthly Sales',type='l')
> points(y=winnebago,x=time(winnebago), pch=as.vector(season(winnebago)))
```

As we would expect with recreational vehicles in the U.S., there is substantial seasonality in the series. However, there is also a general upward "trend" with increasing variation at the higher levels of the series.

(b) Use least squares to fit a line to these data. Interpret the regression output. Plot the standardized residuals from the fit as a time series. Interpret the plot.

```
> winnebago.lm=lm(winnebago~time(winnebago)); summary(winnebago.lm)
```

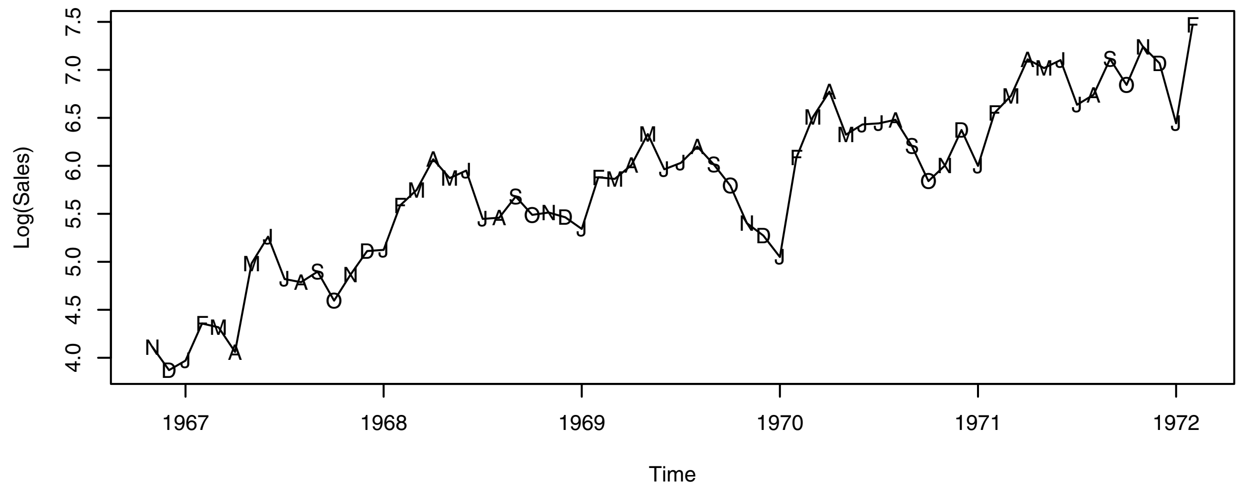
```
Call:
lm(formula = winnebago ~ time(winnebago))

Residuals:
    Min      1Q  Median      3Q     Max
-419.58  -93.13  -12.78   94.96  759.21

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)      -394885.68   33539.77  -11.77   <2e-16 ***
time(winnebago)      200.74      17.03   11.79   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 209.7 on 62 degrees of freedom
Multiple R-Squared: 0.6915,     Adjusted R-squared: 0.6865
F-statistic: 138.9 on 1 and 62 DF,  p-value: < 2.2e-16
```




```
> plot(y=rstudent(winnebago.lm),x=as.vector(time(winnebago)),type='l',
  ylab='Standardized Residuals')
> points(y=rstudent(winnebago.lm),x=as.vector(time(winnebago)),
  pch=as.vector(season(winnebago)))
```

Although the "trend" has been removed, this clearly is not an acceptable model and we move on.

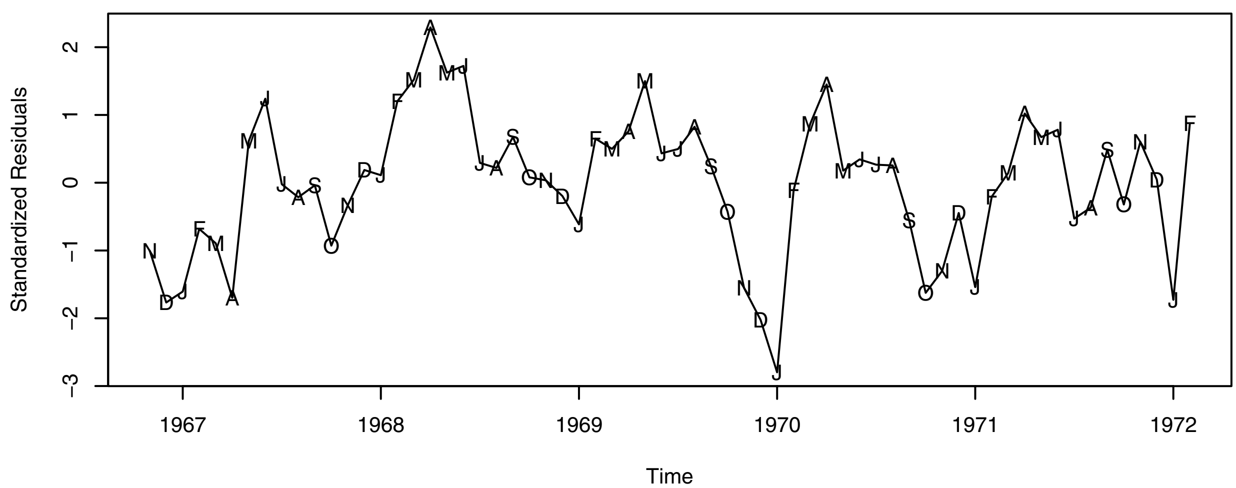
(c) Now take natural logarithms of the monthly sales figures and display and interpret the time series plot of the transformed values.




```
> plot(log(winnebago),ylab='Log(Sales)',type='l')
> points(y=log(winnebago),x=time(winnebago), pch=as.vector(season(winnebago)))
```

In this we see that the seasonality is still present but that now the upward trend is accompanied by much more equal variation around that trend.

(d) Use least squares to fit a line to the logged data. Display the time series plot of the standardized residuals from this fit and interpret.




```
> logwinnebago.lm=lm(log(winnebago)~time(log(winnebago))); summary(logwinnebago.lm)
> plot(y=rstudent(logwinnebago.lm),x=as.vector(time(winnebago)),type='l',
   ylab='Standardized Residuals')
> points(y=rstudent(logwinnebago.lm),x=as.vector(time(winnebago)),
   pch=as.vector(season(winnebago)))
```

The residual plot looks much more acceptable now but we still need to model the seasonality.

(e) Now use least squares to fit a seasonal-means plus linear time trend to the logged sales time series and save the standardized residuals for further analysis. Check the statistical significance of each of the regression coefficients in the model.

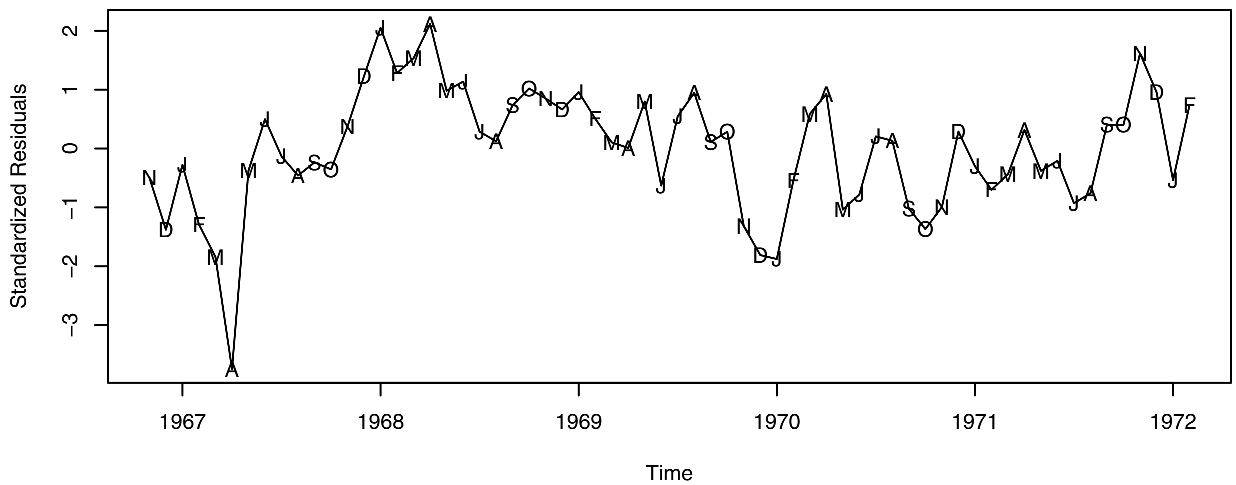
```
> month.=season(winnebago)
> logwinnebago.lm2=lm(log(winnebago)~month.+time(log(winnebago)))
```

```
> summary(logwinnebago.lm2)
```

```
Call:
lm(formula = log(winnebago) ~ month. + time(log(winnebago)))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.92501 -0.16328  0.03344  0.20757  0.57388 

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)          -997.33061   50.63995 -19.695  < 2e-16 ***
month.February          0.62445    0.18182   3.434  0.00119 ** 
month.March             0.68220    0.19088   3.574  0.00078 ***
month.April             0.80959    0.19079   4.243 9.30e-05 ***
month.May               0.86953    0.19073   4.559 3.25e-05 ***
month.June              0.86309    0.19070   4.526 3.63e-05 ***
month.July              0.55392    0.19069   2.905  0.00542 ** 
month.August            0.56989    0.19070   2.988  0.00431 ** 
month.September         0.57572    0.19073   3.018  0.00396 ** 
month.October           0.26349    0.19079   1.381  0.17330    
month.November          0.28682    0.18186   1.577  0.12095    
month.December          0.24802    0.18182   1.364  0.17853    
time(log(winnebago))    0.50909    0.02571  19.800  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.3149 on 51 degrees of freedom
Multiple R-Squared: 0.8946,     Adjusted R-squared: 0.8699 
F-statistic: 36.09 on 12 and 51 DF,  p-value: < 2.2e-16
```

This model explains a large percentage of the variation in sales but, as always, we should also look at the residuals.

(f) Display the time series plot of the standardized residuals obtained in part (e). Interpret the plot.




```
> plot(y=rstudent(logwinnebago.lm2),x=as.vector(time(winnebago)),type='l',
>  ylab='Standardized Residuals')
> points(y=rstudent(logwinnebago.lm2),x=as.vector(time(winnebago)),
>  pch=as.vector(season(winnebago)))
```

This residual plot is the best we have seen for models of this series but perhaps there are better models to be explored later.

Q.5)

(a) Obtain the residuals from the least squares fit of the seasonal-means plus quadratic time trend model.

```
> beersales.lm2=lm(beersales~month.+time(beersales)+I(time(beersales)^2))
```

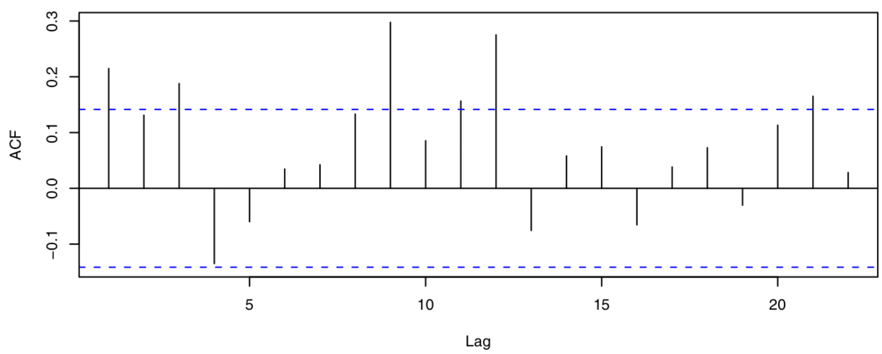
(b) Perform a Runs test on the standardized residuals and interpret the results.

```
> runs(rstudent(beersales.lm2))
```

```
$pvalue
[1] 0.0127
$observed.runs
[1] 79
$expected.runs
[1] 96.625
$n1
[1] 90
$n2
[1] 102
$k
[1] 0
```

We would reject independence of the error terms on the basis of these results.

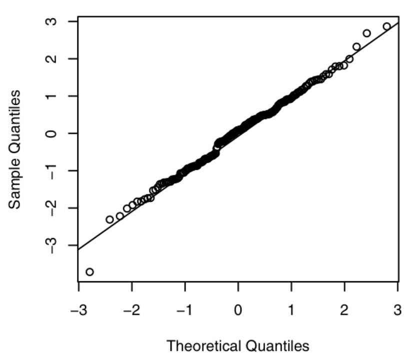
(c) Calculate the sample autocorrelations for the standardized residuals and interpret.

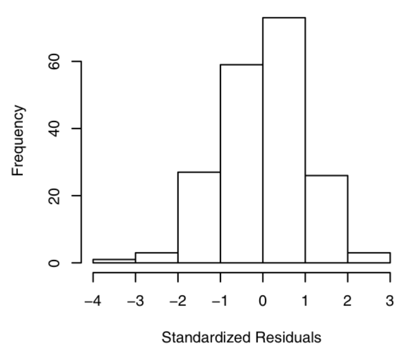



```
> acf(rstudent(beersales.lm2))
```

These results also show the lack of independence in the error terms of this model.

(d) Investigate the normality of the standardized residuals (error terms). Consider histograms and normal probability plots. Interpret the plots.






```
> qqnorm(rstudent(beersales.lm2)); qqline(rstudent(beersales.lm2))
> hist(rstudent(beersales.lm2),xlab='Standardized Residuals')
> shapiro.test(rstudent(beersales.lm2))
```

```
Shapiro-Wilk normality test
data:  rstudent(beersales.lm2)
W = 0.9924, p-value = 0.4139
```

All of these results provide good support for the assumption of normal error terms.