MA 4780 Homework 2

Benjamin Hendrick

February 10, 2016

Problem 3.2

$$
\begin{aligned}
\bar{Y} &= [\mu + \frac{1}{n}\sum_{t=1}^{n}(e_t - e_{t-1})] \\
&= \mu + \frac{1}{n}(e_n - e_0)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
Var(\bar{Y}) &= \frac{1}{n^2}Var(e_n - e_0) \\
&= \frac{2}{n^2}\sigma_e^2
\end{aligned}
$$

Problem 3.3

$$
\sum_{t=1}^{n}(e_t + e_{t-1}) = e_n + e_0 + 2\sum_{t=1}^{n-1}e_t
$$

Therefore,

$$
\begin{aligned}
Var(\bar{Y}) &= \frac{1}{n^2}[\sigma_e^2 + \sigma_e^2 + 4(n-1)\sigma_e^2] \\
&= \frac{2(2n-1)}{n^2}\sigma_e^2
\end{aligned}
$$

Problem 3.7

Load the data.

```r
library(TSA)
```

```
## Loading required package: leaps
## Loading required package: locfit
## locfit 1.5-9.1    2013-03-22
## Loading required package: mgcv
## Loading required package: nlme
## This is mgcv 1.8-7. For overview type 'help("mgcv-package")'.
## Loading required package: tseries
##
```

```
## Attaching package: 'TSA'
##
## The following objects are masked from 'package:stats':
##
##     acf, arima
##
## The following object is masked from 'package:utils':
##
##     tar
```

```
data("winnebago")
```

Part A

Figure 1 shows the time series plot of the data `winnebago`.

```
plot(winnebago,
     type="o",
     ylab="Unit Sales",
     main="Monthly Unit Sales over Time")
model <- lm(winnebago~time(winnebago))
abline(model)
```

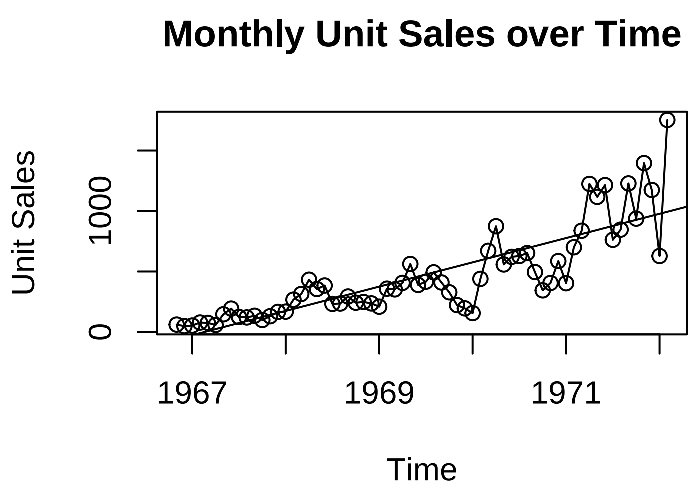

Figure 1: Time series plot of monthly unit sales of recreational vehicles from Winnebago, Inc.

Part B

The fitted regression line fits well to the time series data. The line data is not heteroscedastic, but the line still fits well enough. Figure 2 shows the standardized residuals from the fit as a time series. The residuals appear to be centered around mean zero.

```
plot(rstudent(model),
     type = "o",
     ylab = "Standard Residuals",
     main = "Monthly Unit Sales")
```

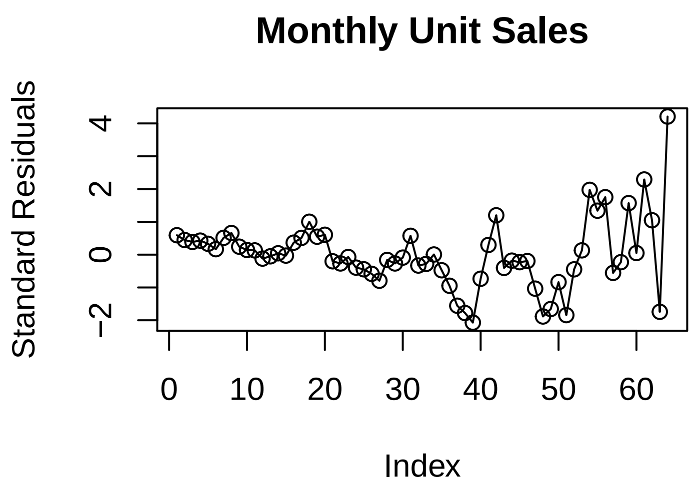

Figure 2: Plot of standardized residuals from the fit as a time series.

Part C

Figure 3 time series plot of the monthly unit sales with a log scale.

```
winnLog <- log(winnebago)
plot(winnLog,
     type="o",
     ylab = "Unit Sales (Log Scale)",
     main = "Monthly Unit Sales")
logModel <- lm(winnLog~time(winnLog))
abline(logModel)
```

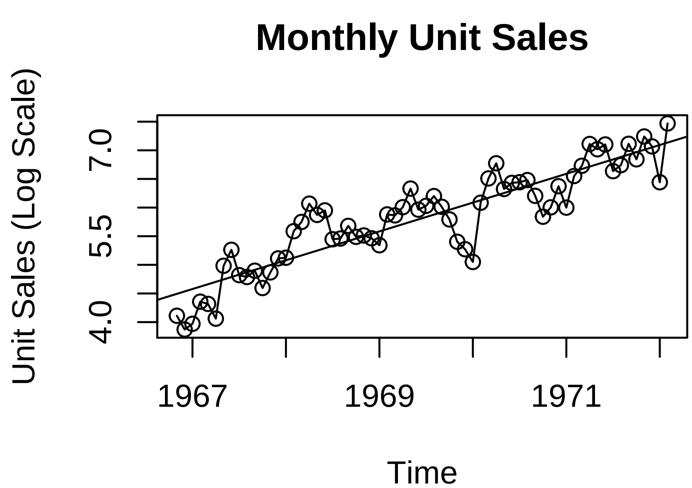

Figure 3: Time series plot of monthly unit sales of recreational vehicles from Winnebago, Inc. under a log scale

Part D

The least squares fit on the logged data fits very well. The data with the log transformation seem more heteroscedastic, yet just as linear. Figure 4 shows the standardized residuals of the logged data. The residuals appear to be centered around mean zero.

```r
plot(rstudent(model),
     type = "o",
     ylab = "Standard Residuals",
     main = "Monthly Unit Sales")
```

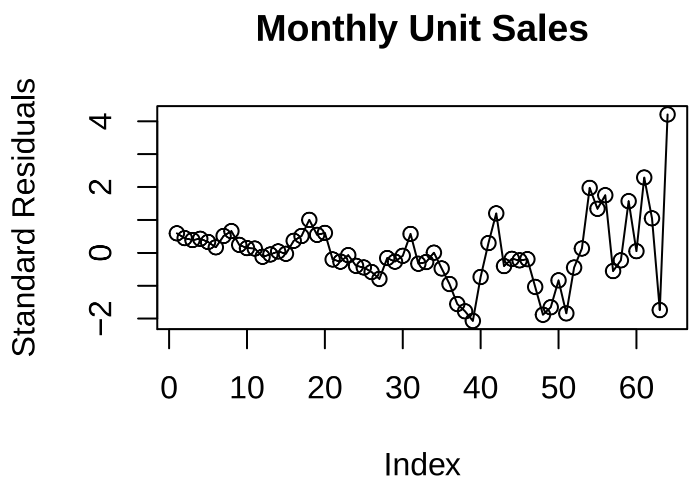

Figure 4: Plot of standardized residuals from the fit as a time series with log scale.

Part E

```r
month <- season(winnebago)
winCombo <- lm(winnLog~month + time(winnLog))
summary(winCombo)
```

```
##
## Call:
## lm(formula = winnLog ~ month + time(winnLog))
##
## 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 ***
## monthFebruary    0.62445    0.18182   3.434 0.001188 **
## monthMarch       0.68220    0.19088   3.574 0.000779 ***
## monthApril       0.80959    0.19079   4.243 9.30e-05 ***
## monthMay         0.86953    0.19073   4.559 3.25e-05 ***
## monthJune        0.86309    0.19070   4.526 3.63e-05 ***
## monthJuly        0.55392    0.19069   2.905 0.005420 **
## monthAugust      0.56989    0.19070   2.988 0.004305 **
## monthSeptember   0.57572    0.19073   3.018 0.003960 **
## monthOctober     0.26349    0.19079   1.381 0.173300
```

```
## monthNovember     0.28682    0.18186   1.577 0.120946
## monthDecember     0.24802    0.18182   1.364 0.178532
## time(winnLog)     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
```

Many of the month coefficients are statistically significant. All months but October, November, and December are not statistically significant. This is because their p-values were too high (greeter than 0.10).

Part F

Figure 5 shows the standardized residuals from the combined fit. The residuals appear to be centered around mean zero.

```
plot(rstudent(winCombo),
     type = "o",
     ylab = "Standard Residuals",
     main ="Monthly Unit Sales")
```

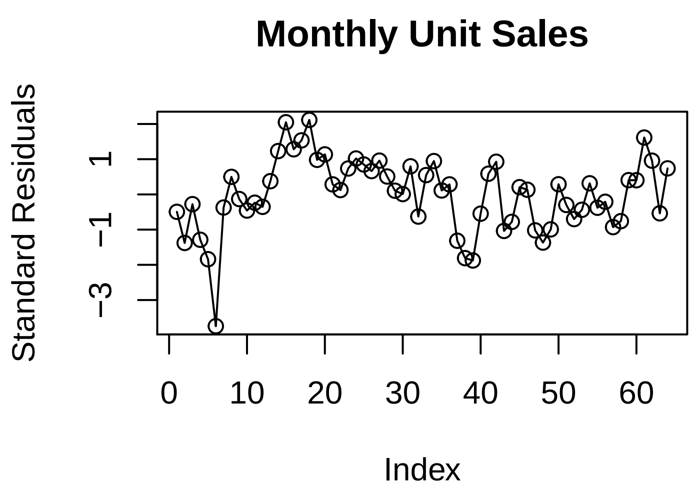

Figure 5: Plot of standardized residuals from the combined fit as a time series.

Problem 3.8

Load the data.

```
data(retail)
```

Part A

Figure 6 shows the time series data with seasonal markers. There is clear seasonality where December has the highest sales. The overall sales is positively increasing.

```r
plot(retail,
     type = "l",
     ylab = "Billions of Pounds",
     main = "Retail Sales over Time")
points(y=retail,
       x=time(retail),
       pch=as.vector(season(retail)))
```

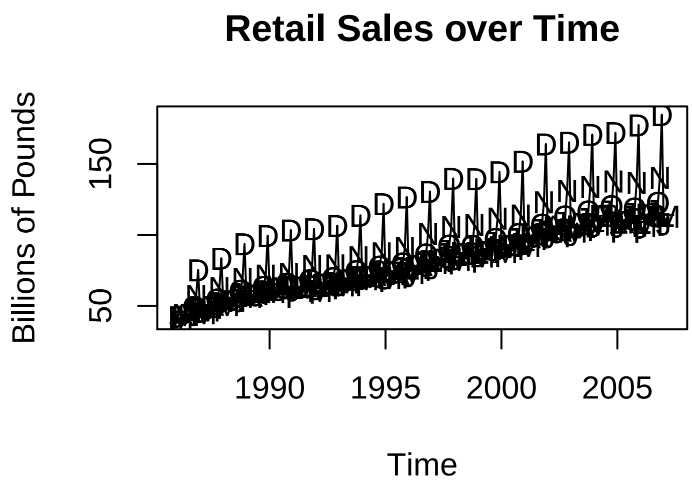

Figure 6: The time series plot of the retail data with seasonal symbols.

Part B

The linear model has many important coefficients. The `summary` of the model shows that every month is statistically significant. This is because they have small p-values (less than 0.05).

```r
retail.lm <- lm(retail~season(retail) + time(retail))
summary(retail.lm)
```

```
##
## Call:
## lm(formula = retail ~ season(retail) + time(retail))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.8950  -2.4440  -0.3518   2.1971  16.2045
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -7.249e+03  8.724e+01 -83.099  < 2e-16 ***
## season(retail)February -3.015e+00  1.290e+00  -2.337  0.02024 *
## season(retail)March     7.469e-02  1.290e+00   0.058  0.95387
## season(retail)April     3.447e+00  1.305e+00   2.641  0.00880 **
## season(retail)May       3.108e+00  1.305e+00   2.381  0.01803 *
## season(retail)June      3.074e+00  1.305e+00   2.355  0.01932 *
## season(retail)July      6.053e+00  1.305e+00   4.638 5.76e-06 ***
```

```
## season(retail)August     3.138e+00  1.305e+00   2.404  0.01695 *
## season(retail)September  3.428e+00  1.305e+00   2.626  0.00919 **
## season(retail)October    8.555e+00  1.305e+00   6.555 3.34e-10 ***
## season(retail)November   2.082e+01  1.305e+00  15.948  < 2e-16 ***
## season(retail)December   5.254e+01  1.305e+00  40.255  < 2e-16 ***
## time(retail)             3.670e+00  4.369e-02  83.995  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.278 on 242 degrees of freedom
## Multiple R-squared:  0.9767, Adjusted R-squared:  0.9755
## F-statistic:   845 on 12 and 242 DF,  p-value: < 2.2e-16
```

Part C

Figure 7 shows the standardized residuals. The residuals appear symmetric over time, implying that they are centered around mean zero.

```r
plot(rstudent(retail.lm),
     type = "l",
     ylab = "Standardized Residuals",
     main = "Retail Sales")
points(y=rstudent(retail.lm),
       x=as.vector(time(rstudent(retail.lm))),
       pch=as.vector(season(retail)))
```

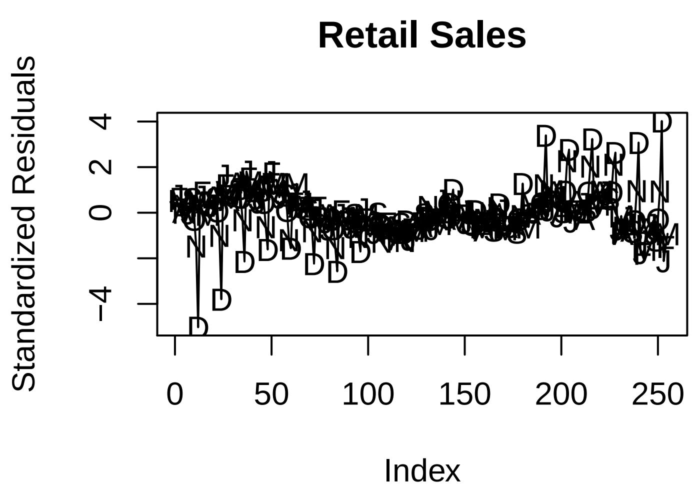

Figure 7: Plot of standardized residuals from the model in Part B

Problem 3.9

Load the data.

```r
data(prescrip)
```

Part A

Figure 8 shows the time series plot with seasonal markings. There is an apparent seasonal trend during the summer. The overall trend is positively increasing.

```
plot(prescrip,
     type = "l",
     ylab = "Presciption Costs",
     main = "Presciption Costs over Time")
points(y = prescrip,
       x = time(prescrip),
       pch = as.vector(season(prescrip)))
```

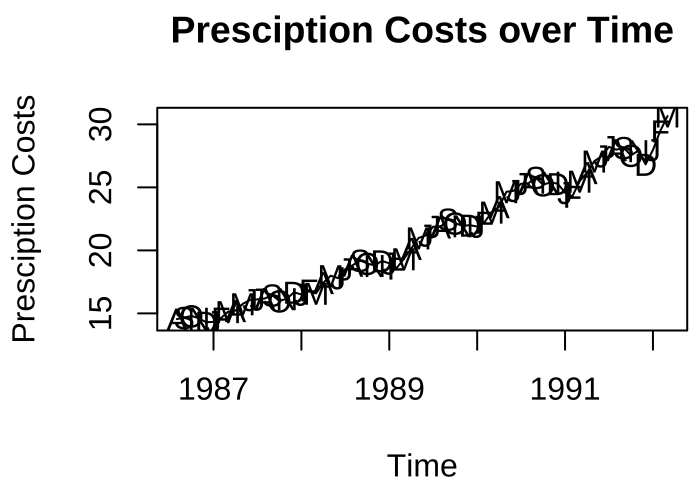

Figure 8: Time series plot of the monthly prescription costs in the United States

Part B

Figure 9 shows the time series plot of the percentage changes in prescription costs with seasonal markers.

```
percPrecrip <- na.omit(100*(prescrip - zlag(prescrip))/zlag(prescrip))
plot(percPrecrip,
     type = "l",
     ylab = "Percent Change of Prescription Costs",
     main = "Percent Change of Prescription Costs over Time")
points(y = percPrecrip,
       x = time(percPrecrip),
       pch = as.vector(season(percPrecrip)))
```

Part C

```
prescrip.lm <- lm(percPrecrip ~ harmonic(percPrecrip))
summary(prescrip.lm)
```

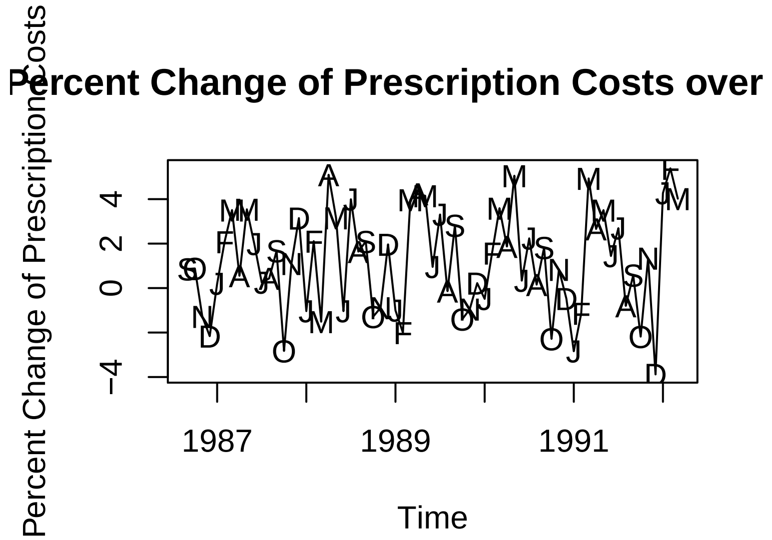

Figure 9: Time series of the sequence of month-to-month percentage changes in prescription costs.

```
##
## Call:
## lm(formula = percPrecrip ~ harmonic(percPrecrip))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8444 -1.3742  0.1697  1.4069  3.8980
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1.2217     0.2325   5.254 1.82e-06 ***
## harmonic(percPrecrip)cos(2*pi*t)  -0.6538     0.3298  -1.982   0.0518 .
## harmonic(percPrecrip)sin(2*pi*t)   1.6596     0.3269   5.077 3.54e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.897 on 64 degrees of freedom
## Multiple R-squared:  0.3148, Adjusted R-squared:  0.2933
## F-statistic:  14.7 on 2 and 64 DF,  p-value: 5.584e-06
```

The `summary` of the data shows that the harmonic cosine coefficient is not statistically significant because the p-value is greater than 0.05. If our alpha values is set to 0.10, then every coefficient would be statistically significant.

Part D

Figure 10 shows the standardized residuals of the cosine model. The residuals appear to be centered around mean zero and are fairly random.

```
plot(rstudent(prescrip.lm),
     type = "l",
     ylab = "Standardized Residuals",
     main = "Presciption Costs")
```

```r
points(y=rstudent(prescrip.lm),
       x=as.vector(time(rstudent(prescrip.lm))),
       pch=as.vector(season(prescrip)))
```

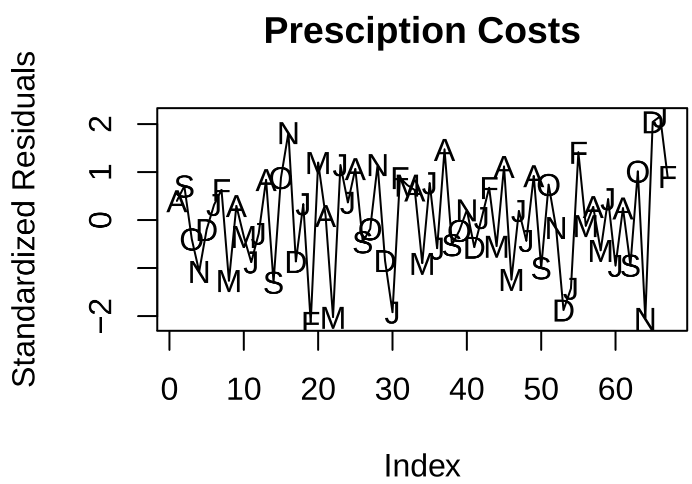

Figure 10: Plot of standardized residuals of the cosine model.

Problem 3.13

Load the data.

```r
data("retail")
```

Part A

```r
winnebago.lm <- lm(winnebago~season(log(winnebago)) + time(log(winnebago)))
```

Part B

```r
runs(rstudent(winnebago.lm))
```

```
## $pvalue
## [1] 0.0159
##
## $observed.runs
## [1] 23
##
## $expected.runs
## [1] 33
##
## $n1
```

```
## [1] 32
##
## $n2
## [1] 32
##
## $k
## [1] 0
```

The **runs** test suggest a lack of independence in the error terms in the model.

Part C

Figure 11 shows the ACF plot of the combined Winnebago model. The residuals of lags one, two, three, and five suggest significant positive auto-correlation.

```r
acf(rstudent(winnebago.lm))
```

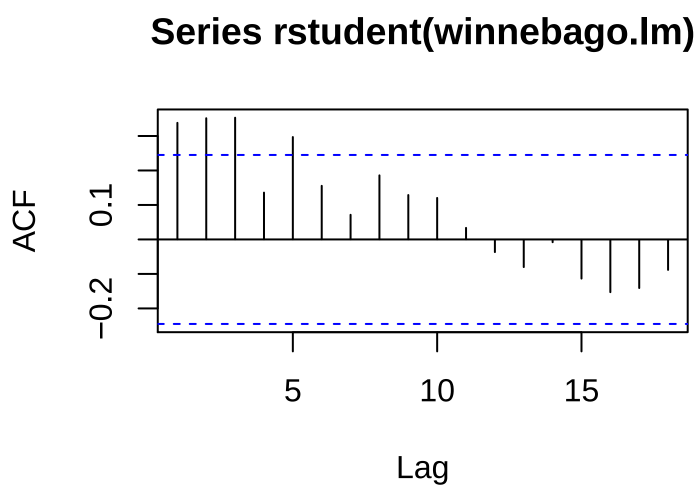

Figure 11: ACF plot of combined Winnebago model residuals

Part D

Figure 12 suggests that the distribution of the standardized residuals of the combined Winnebago model is right skewed. Figure 13 suggests the same.

```r
hist(rstudent(winnebago.lm),
     xlab = "Standardized Residuals",
     main = "Historgram of Standardized Residuals")
```

```r
qqnorm(rstudent(winnebago.lm))
qqline(rstudent(winnebago.lm))
```

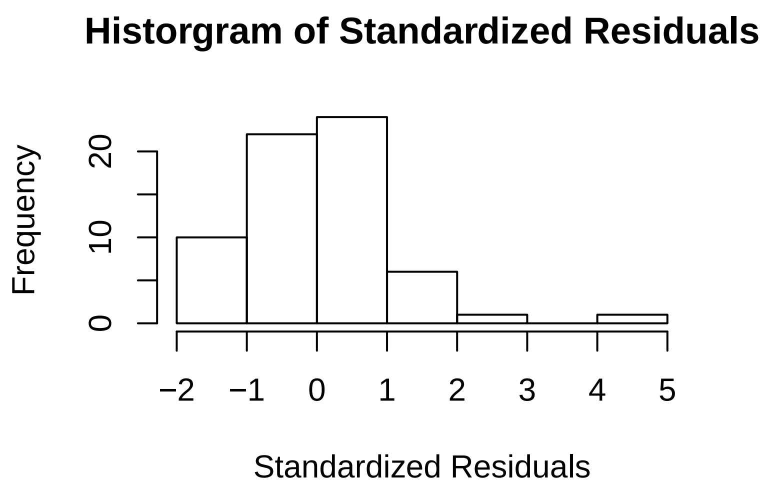

Figure 12: Histogram of combined Winnebago model residuals.

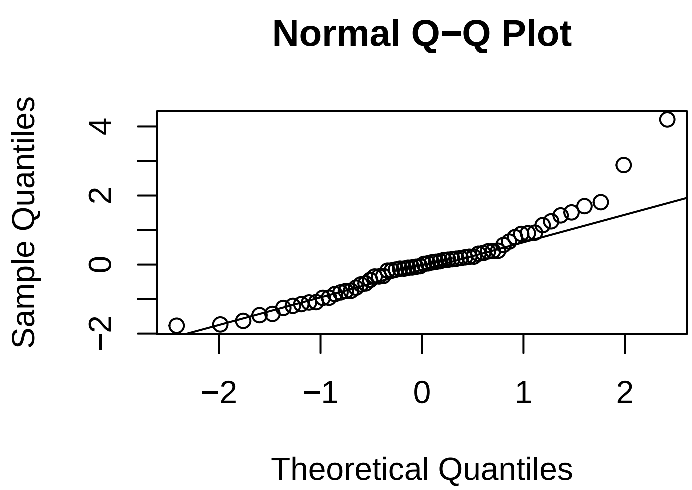

Figure 13: QQ plot of combined Winnebago model residuals.

Problem 3.14

Load the data.

```
data("retail")
```

Part A

```
retail.lm <- lm(retail~season(retail) + time(retail))
```

Part B

```
runs(rstudent(retail.lm))
```

```
## $pvalue
## [1] 9.19e-23
##
## $observed.runs
## [1] 52
##
## $expected.runs
## [1] 127.9333
##
## $n1
## [1] 136
##
## $n2
## [1] 119
##
## $k
## [1] 0
```

The `runs` test suggest a lack of independence in the error terms in the model.

Part C

```
acf(rstudent(retail.lm))
```

Figure 14 suggest significant positive auto-correlation at lags one, eleven, twelve, thirteen, and twenty-four.

Part D

Figure 15 suggests that the distribution of the standardized residuals of the combined retail model is relatively normal. Figure 16 suggests the same.

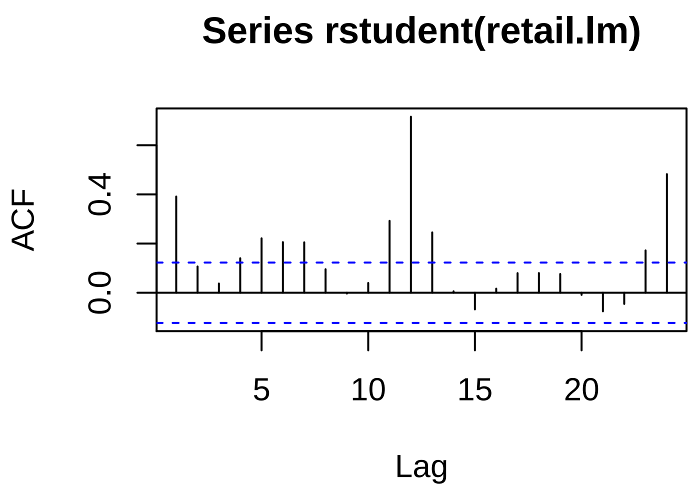

Figure 14: ACF plot for combined retail model.

```r
hist(rstudent(retail.lm),
     xlab = "Standardized Residuals",
     main = "Historgram of Standardized Residuals")
```

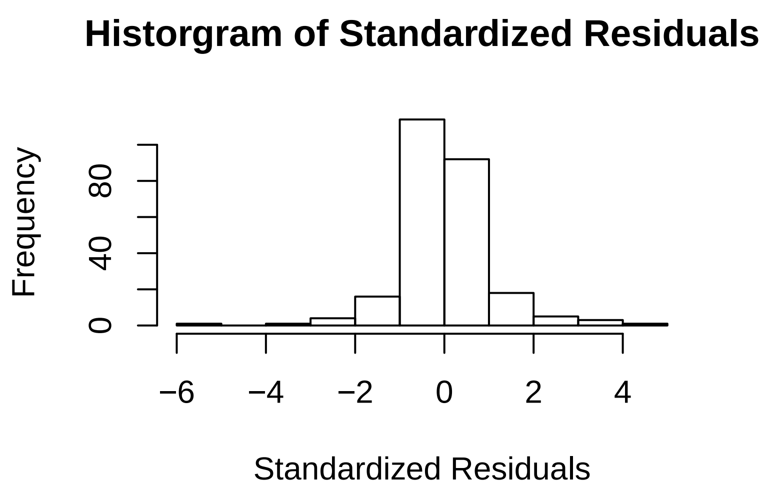

Figure 15: Histogram of standardized residuals of combined retail model.

```r
qqnorm(rstudent(retail.lm))
qqline(rstudent(retail.lm))
```

Problem 3.15

Load the data.

```r
data("prescrip")
```

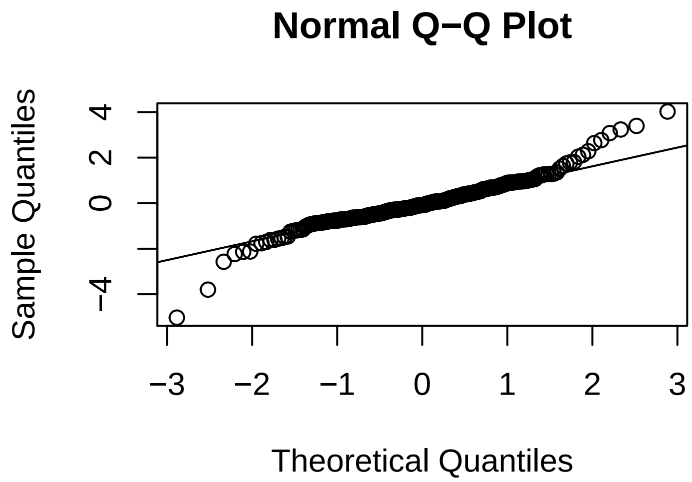

Figure 16: QQ plot of standardized residuals of combined retail model.

Part A

```
percPrecrip <- na.omit(100*(prescrip - zlag(prescrip))/zlag(prescrip))
prescrip.lm <- lm(percPrecrip ~ harmonic(percPrecrip))
```

Part B

```
runs(rstudent(prescrip.lm))
```

```
## $pvalue
## [1] 0.0026
##
## $observed.runs
## [1] 47
##
## $expected.runs
## [1] 34.43284
##
## $n1
## [1] 32
##
## $n2
## [1] 35
##
## $k
## [1] 0
```

The `runs` test suggest a lack of independence in the error terms in the model.

Part C

Figure 17 suggests significant negative auto-correlation at lags one and thirteen.

```r
acf(rstudent(prescrip.lm))
```

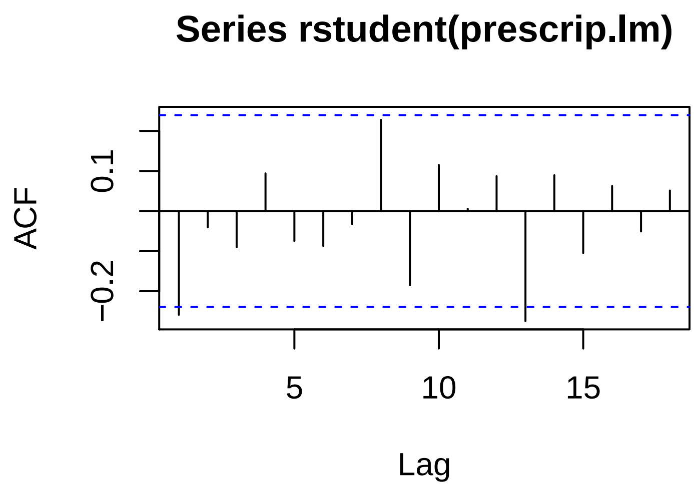

Figure 17: ACF plot for the precription model.

Part D

Figure 18 suggests that the distribution of the standardized residuals of the combined retail model is relatively normal. Figure 19 suggests the same.

```r
hist(rstudent(prescrip.lm),
     xlab = "Standardized Residuals",
     main = "Historgram of Standardized Residuals")
```

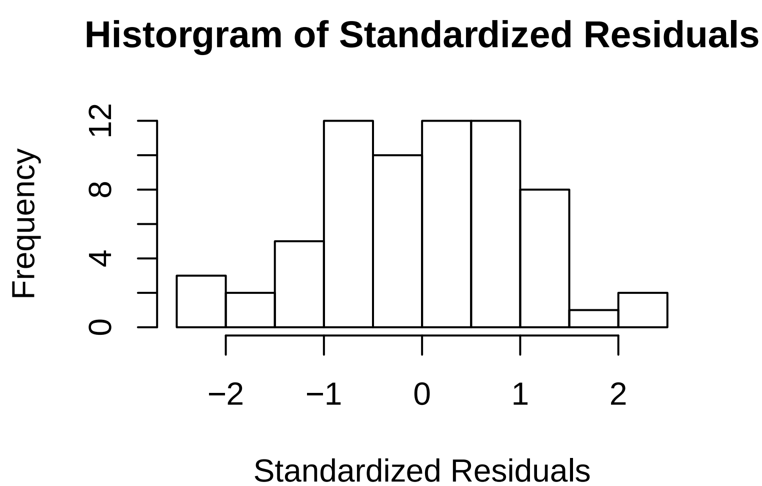

Figure 18: Histogram of standardized residuals of combined retail model.

```r
qqnorm(rstudent(prescrip.lm))
qqline(rstudent(prescrip.lm))
```

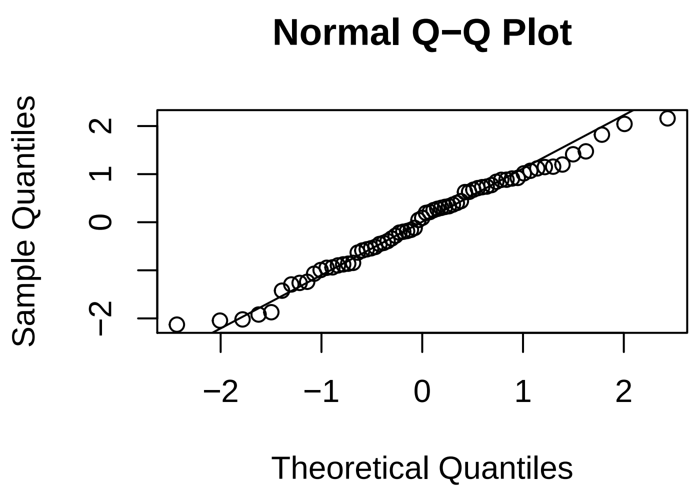

Figure 19: QQ plot of standardized residuals of combined retail model.