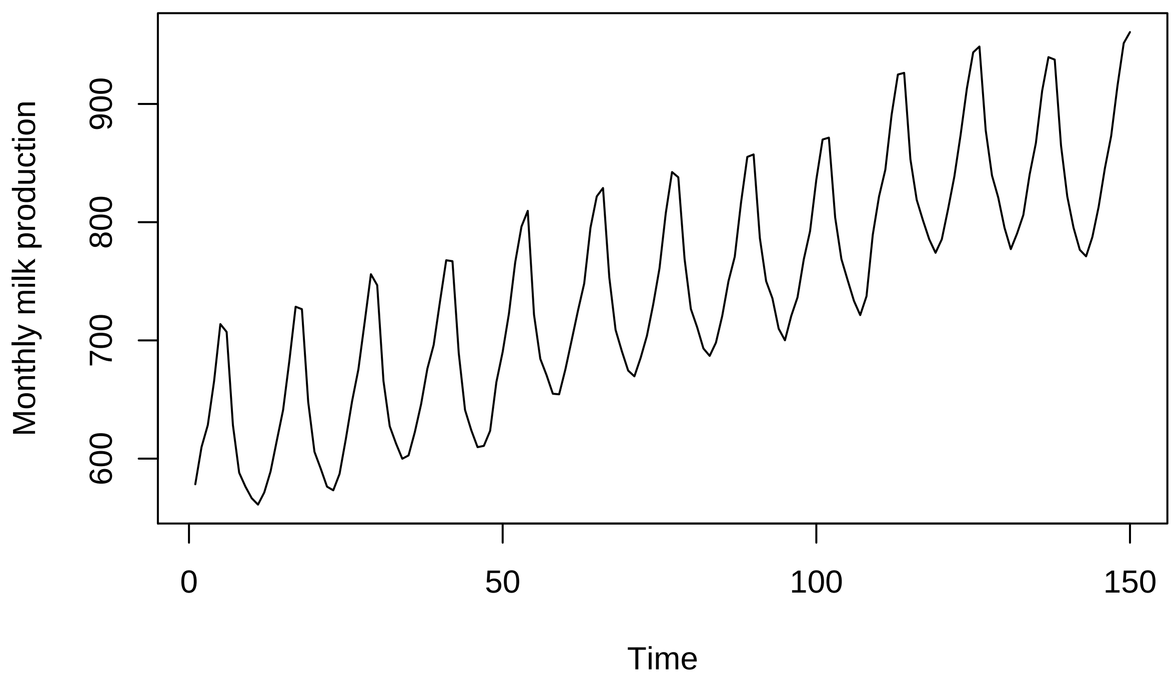
PSTAT174_Lab07

Celeste Herrera

5/14/2021

1. We will again analyze adjusted monthly milk production measured in pounds per from Jan. 1962 to Dec. 1975. And we can import the dataset from tsdl package as milk in R, and denote the milk time series as Xt. For comparison, we split the dataset into training set train and testing set test. The training set is used for model building, and the testing set is used for prediction verification and comparision.

```
library(tsdl)
milk <- subset(tsdl, 12, "Agriculture")[[3]]
train <- milk[1:150]
test <- milk[151:156]
ts.plot(train, ylab = "Monthly milk production")
```



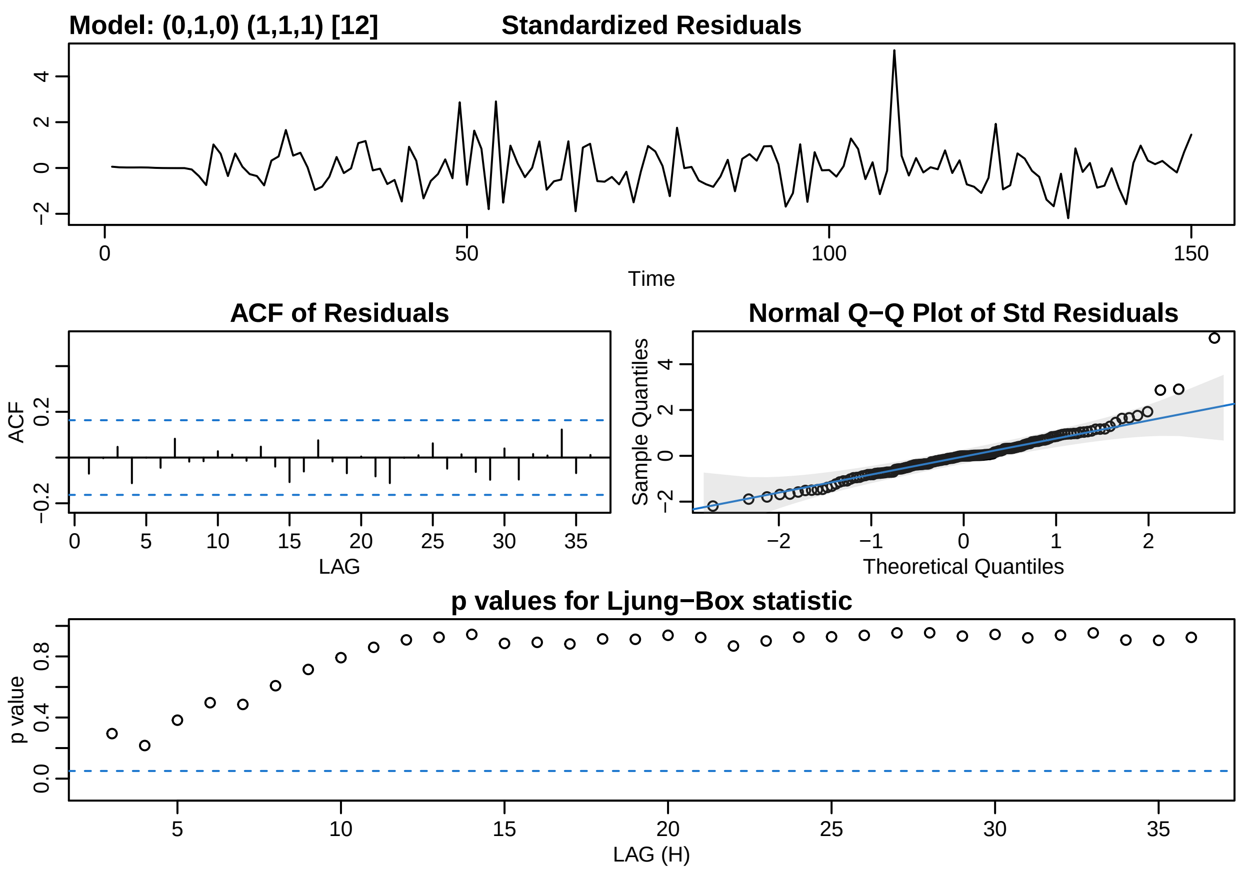
From the above graph, we can conclude that X_t is non-stationary because of the upward trend and seasonality. (You can think about whether we need to transform the series or not.) To make it more stationary, we need to remove trend and seasonality with the following code:

```
dmilk <- diff(train, 12)
ddmilk <- diff(dmilk, 1)
```

Let Y_t denote the series ddmilk. Then, Y_t = (1 - B)(1 - B^{12})X_t. As Lab Assignment 5, we can use SARIMA(0, 1, 0) $\times$ (1, 1, 1)12. Now, we will conduct model diagnostic analysis.

```r
library(astsa)
fit.i <- sarima(xdata=train, p=0, d=1, q=0, P=1, D=1, Q=1, S=12)
```

```
## initial  value 1.989465
## iter   2 value 1.850408
## iter   3 value 1.824156
## iter   4 value 1.800049
## iter   5 value 1.791131
## iter   6 value 1.789958
## iter   7 value 1.789636
## iter   8 value 1.789235
## iter   9 value 1.789186
## iter  10 value 1.789182
## iter  10 value 1.789182
## iter  10 value 1.789182
## final  value 1.789182
## converged
## initial  value 1.803940
## iter   2 value 1.803675
## iter   3 value 1.803165
## iter   4 value 1.803164
## iter   5 value 1.803163
## iter   6 value 1.803163
## iter   6 value 1.803163
## final  value 1.803163
## converged
```



(a) Perform diagnostics on the chosen model fit. Do the residuals appear to be white noise? Are they normally distributed? You should conduct hypothesis testing and plot some graphs to answer this questions. (You can think about why we want to check normality of the residuals.)

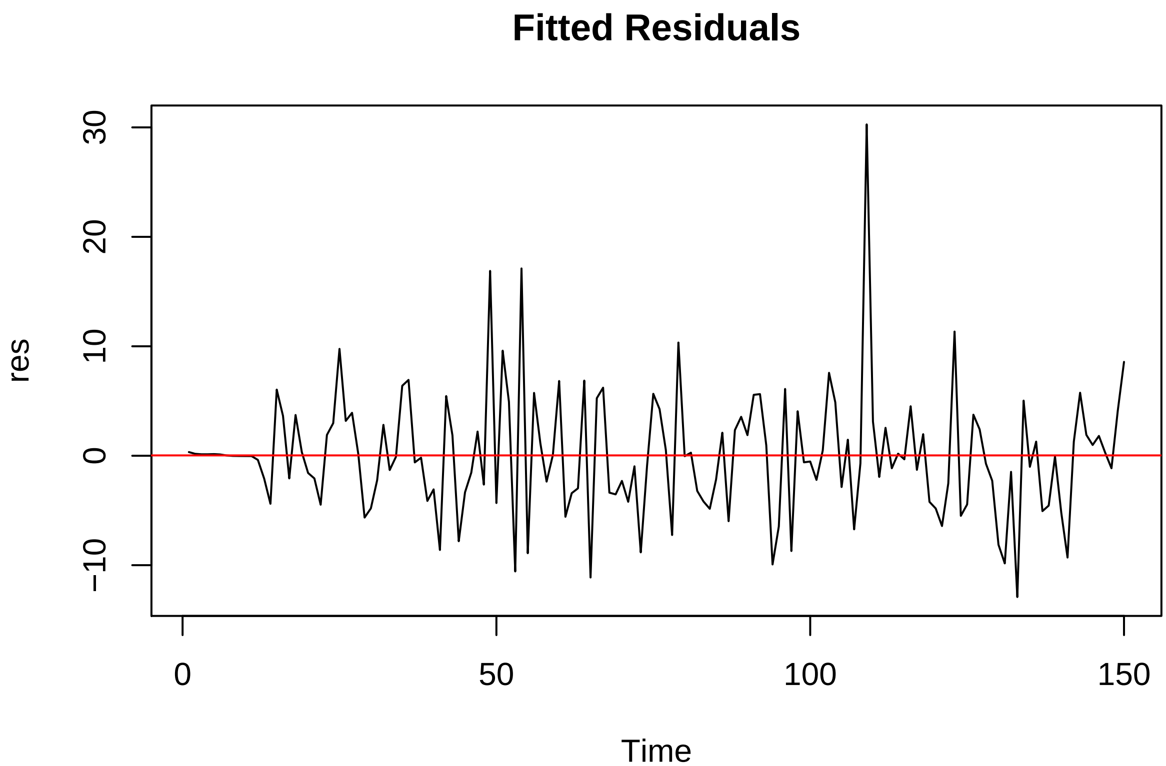
```
# Diagnostics on model
res <- residuals(fit.i$fit)
mean(res)
```

```
## [1] 0.03021635
```

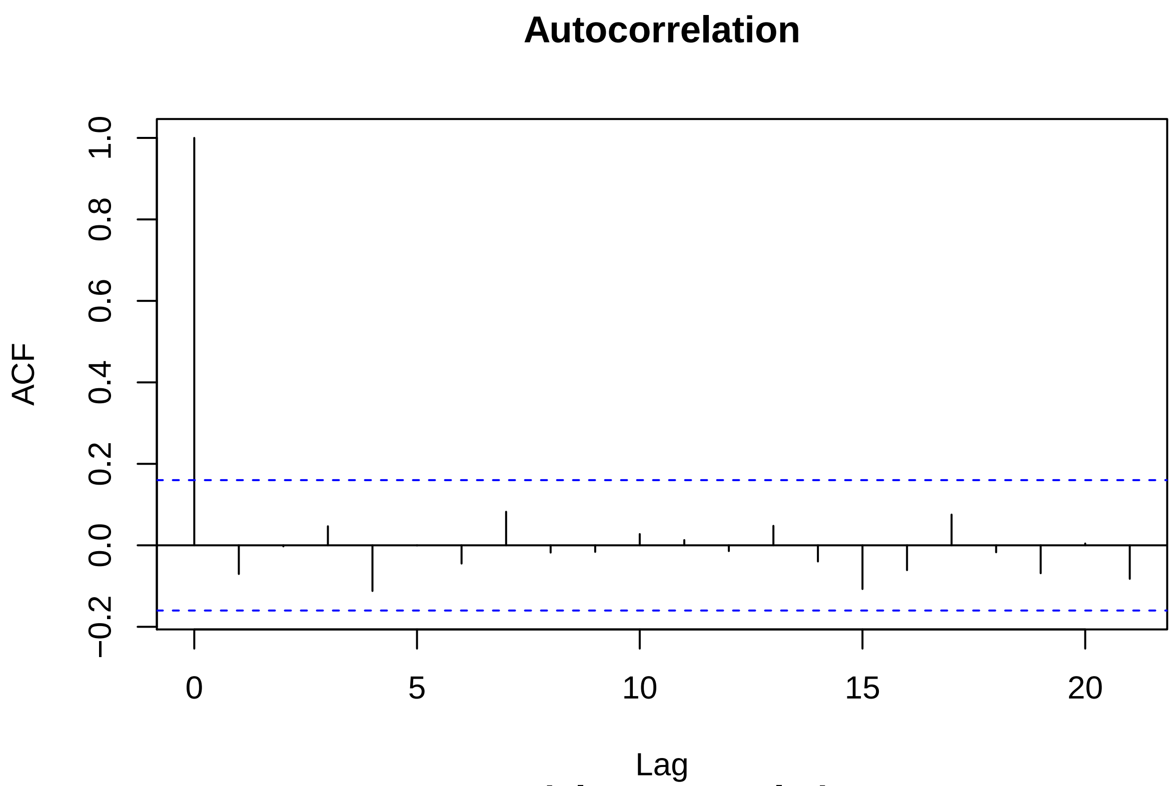
```
var(res)
```

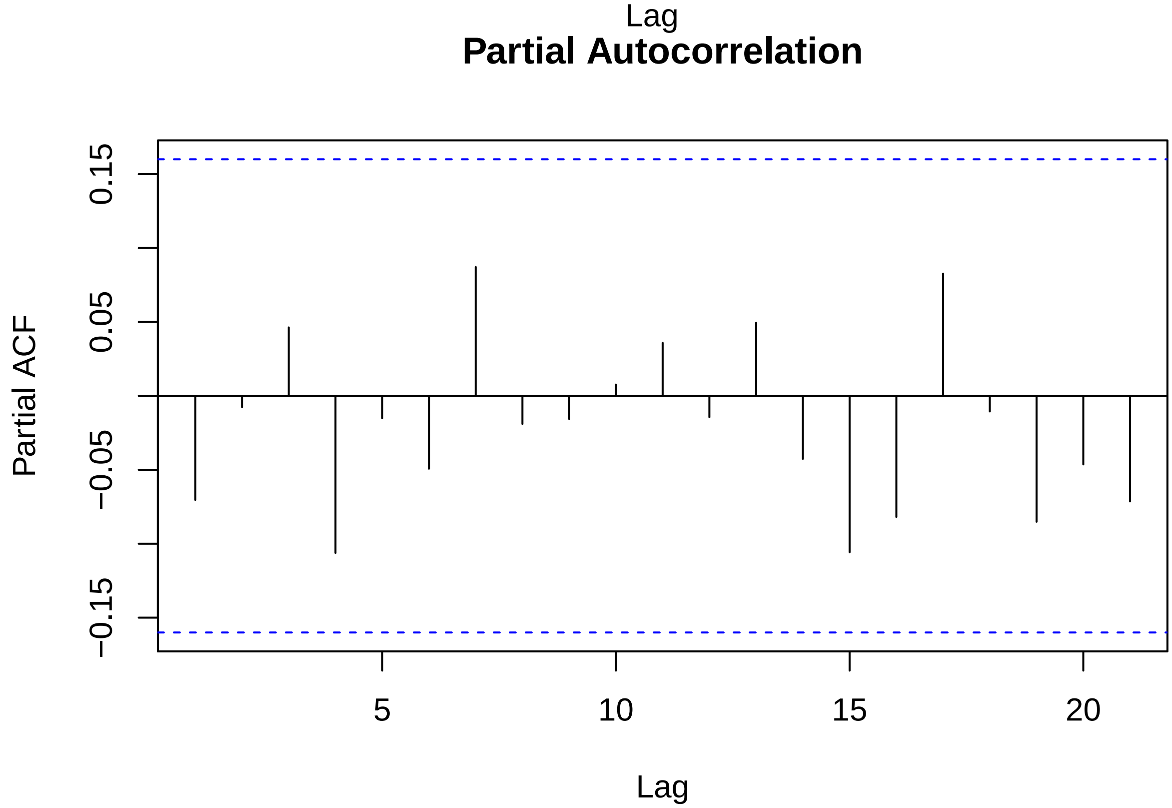
```
## [1] 31.89208
```

```
# Plots on model
ts.plot(res, main = "Fitted Residuals")
abline(h = mean(res), col = "red")
```




```
acf(res, main = "Autocorrelation"); pacf(res, main = "Partial Autocorrelation")
```






```r
# Independence test
Box.test(res, lag = 12, type = c("Box-Pierce"), fitdf = 2)
```

```
##
##  Box-Pierce test
```

```
##
## data:  res
## X-squared = 4.5287, df = 10, p-value = 0.9204
```

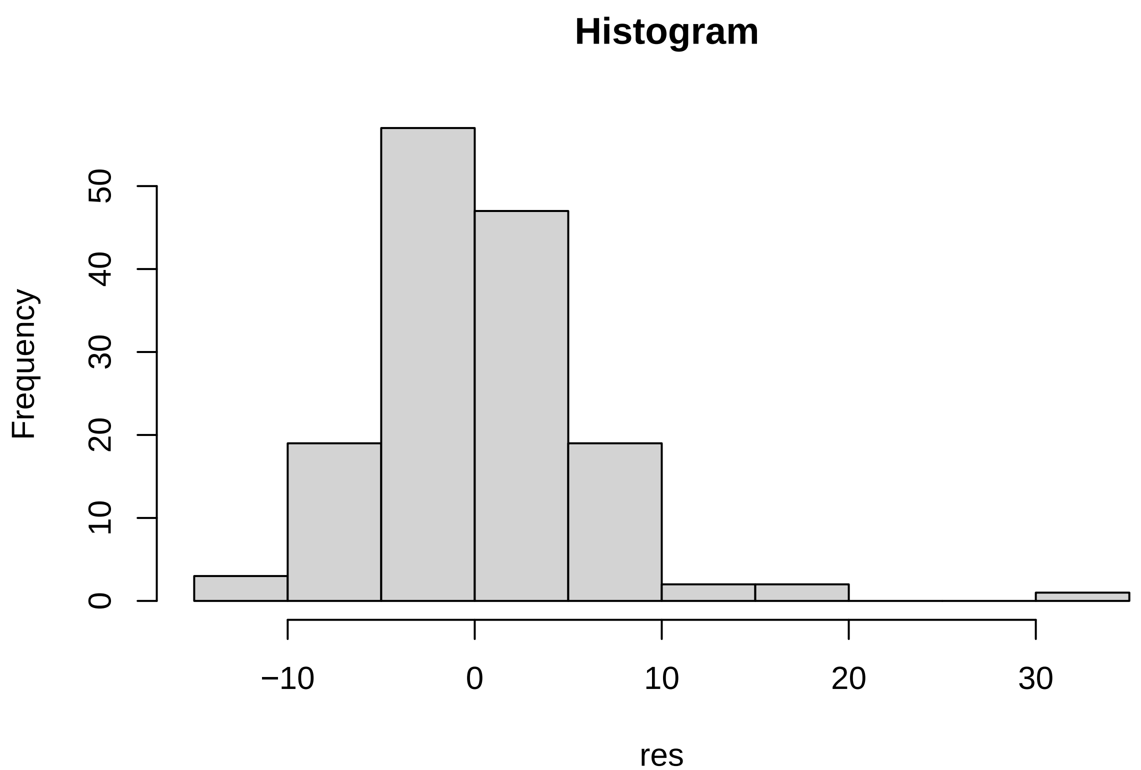
```
Box.test(res, lag = 12, type = c("Ljung-Box"), fitdf = 2)
```

```
##
##  Box-Ljung test
##
## data:  res
## X-squared = 4.7345, df = 10, p-value = 0.9082
```

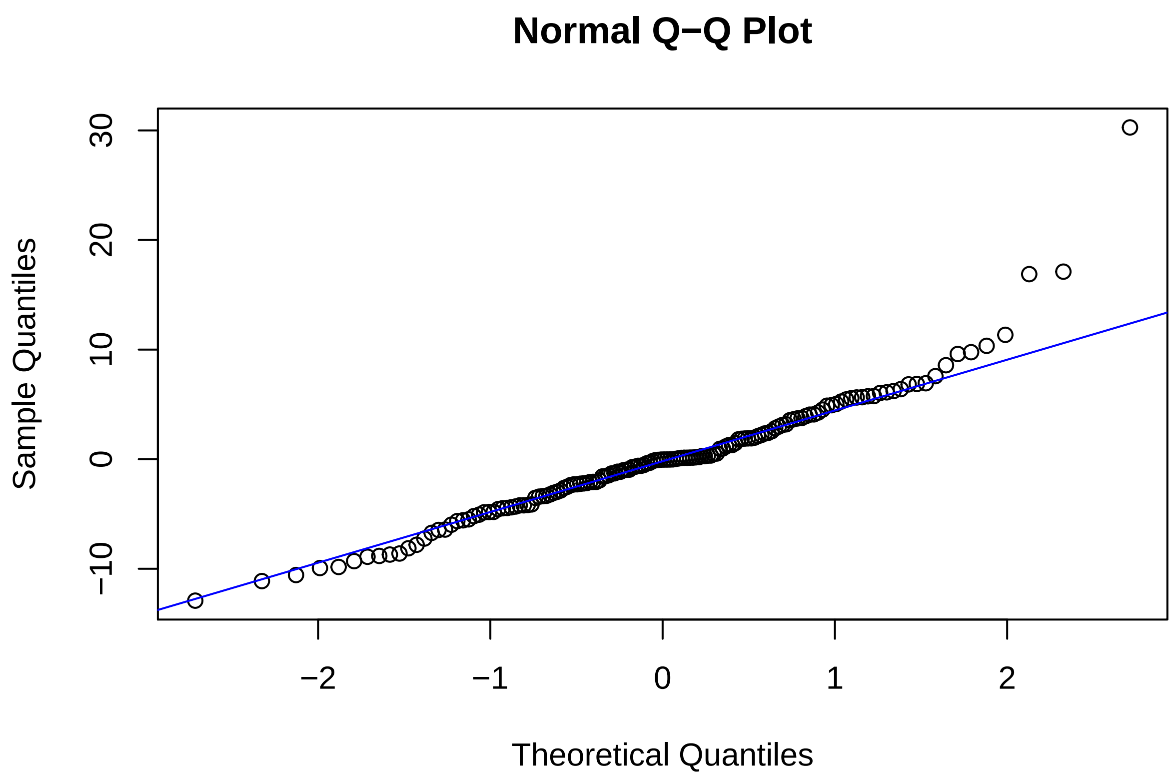
```
Box.test(res^2, lag = 12, type = c("Ljung-Box"), fitdf = 0)
```

```
##
##  Box-Ljung test
##
## data:  res^2
## X-squared = 1.9858, df = 12, p-value = 0.9994
```

```
# Checking the normality asssumption
hist(res, main = " Histogram")
```




```
qqnorm(res)
qqline(res, col = "blue")
```



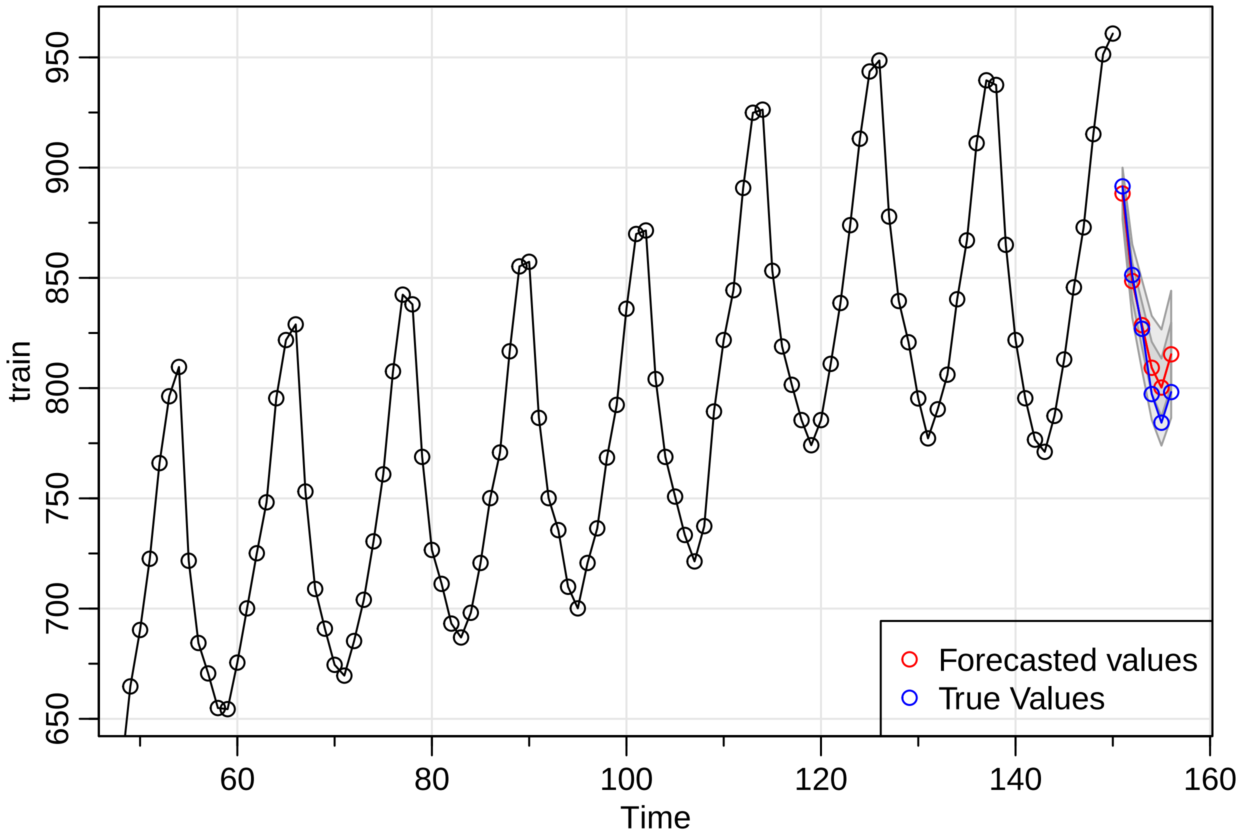

```r
# Test normality of residual
shapiro.test(res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.93109, p-value = 1.183e-06
```

(b) Forecast the next 6 observations using sarima.for(), and plot your predictions. And you should also add true milk prodcution points in test.

```r
pred.milk <- sarima.for(train, n.ahead = 6, plot.all = F,
                        p=0, d=1, q=0, P=1, D=1, Q=1, S=12)

lines(151:156, pred.milk$pred, col = "red")
lines(151:156, test, col = "blue")
points(151:156, test, col = "blue")
legend("bottomright", pch = 1, col = c("red", "blue"),
       legend = c("Forecasted values", "True Values"))
```



train
Time
950
900
850
800
750
700
650
60
80
100
120
140
160
Forecasted values
True Values