Group: Zining Fan, Mutian Wang, Siyuan Wang
UNI: zf2234, mw3386, sw3418

Graded Homework 3- Exercise 1

1. ```
   options(scipen=999)
   library("readxl")
   milk <- read_excel("milk.xls")
   ```

   We import the data after transforming the month into time point variable. Each number represent a month in a sequential order.

   ```
   model <- lm(formula = milk_production ~ time, data = milk)
   print(model)
   ```

   ```
   Call:
   lm(formula = milk_production ~ time, data = milk)

   Coefficients:
   (Intercept)         time
       611.682        1.693
   ```

   We can observe the trend from the coefficient that when month lied before 1962, the average monthly milk production is 611.682, and after that, when one month passed, the average milk production will increase by 1.693 pounds per cow on average.

   ```
   summary(model)
   ```

   ```
   Call:
   lm(formula = milk_production ~ time, data = milk)

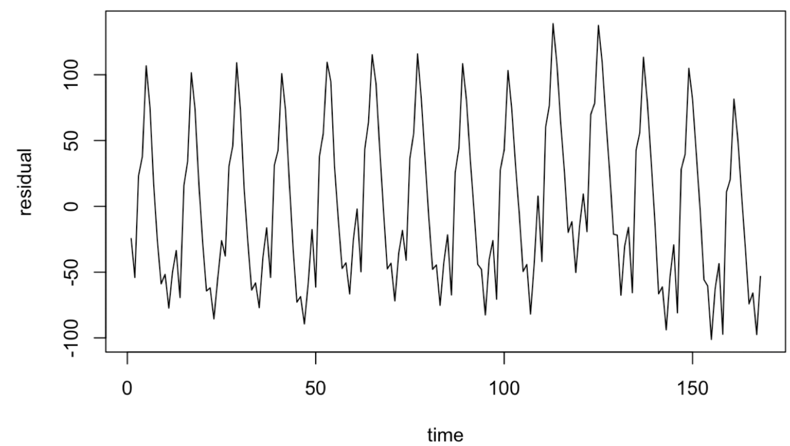
   Residuals:
       Min      1Q  Median      3Q     Max
   -101.04  -50.02  -15.30   42.88  139.05

   Coefficients:
                Estimate Std. Error t value            Pr(>|t|)
   (Intercept) 611.68235    9.41444   64.97 <0.0000000000000002 ***
   time          1.69262    0.09663   17.52 <0.0000000000000002 ***
   ---
   Signif. codes:  0   ***   0.001   **   0.01   *   0.05   .   0.1

   Residual standard error: 60.74 on 166 degrees of freedom
   Multiple R-squared:  0.6489,    Adjusted R-squared:  0.6468
   F-statistic: 306.8 on 1 and 166 DF,  p-value: < 0.00000000000000022
   ```
   ```

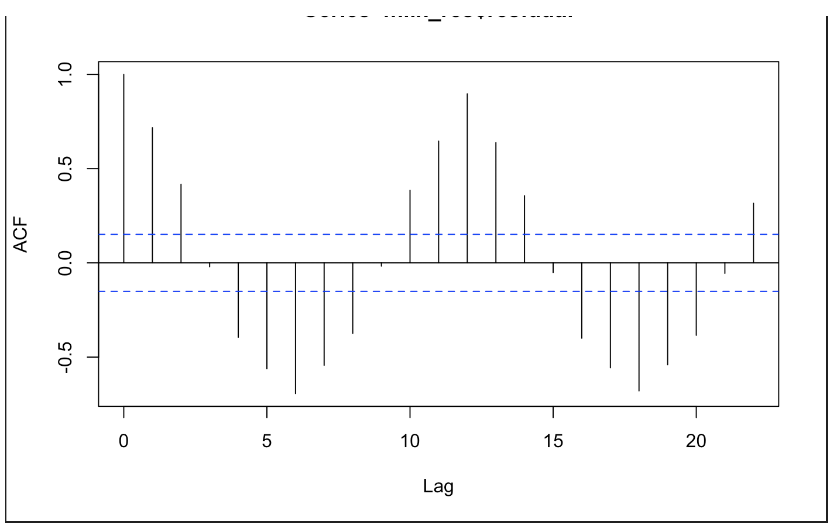
It can be deduced from the summary of the model that the p value of the coefficient is small enough so we can reject the hypothesis that the coefficients are zeros.
We calculate the residuals and plot the curve on time.

```
milk_res <- read_excel("milk_res.xls")
plot(1:168,(milk_res$residual),type = "l", xlab = "time",ylab = "residual" )
```

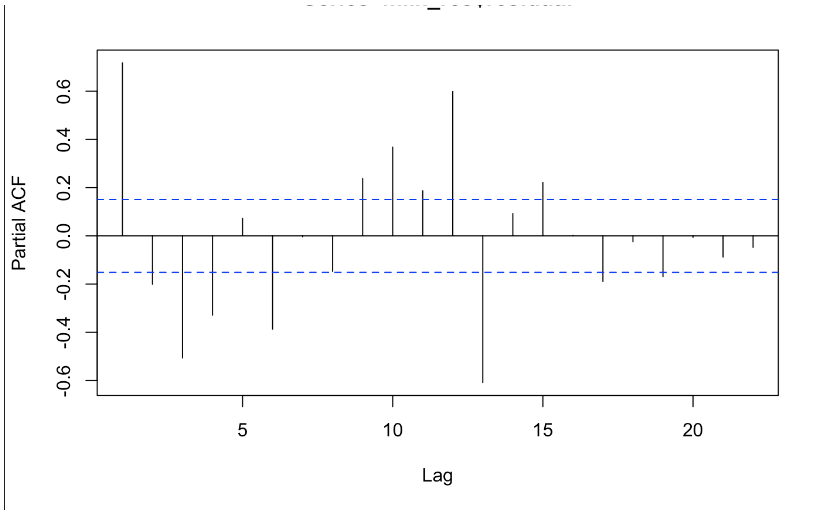


We can clearly see a seasonal structure in the residual of the model.

2. `acf(milk_res$residual)`




```
pacf(milk_res$residual)
```



We can observe systematic departure in the correlogram,it's definitely not white noise. We must try different model to remove the seasonality and try fitting them with different autoregressive models.

3. Firstly, we try with AR(1) model.

```
AR_1 <- arima(milk_res$residual,order = c(1,0,0))
print(AR_1)

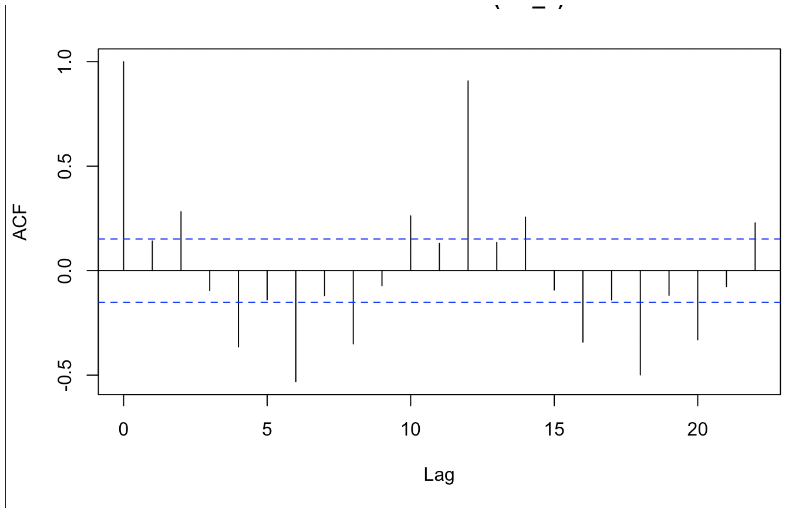
Call:
arima(x = milk_res$residual, order = c(1, 0, 0))

Coefficients:
         ar1   intercept
      0.7172     -1.2379
s.e.  0.0532     11.2744

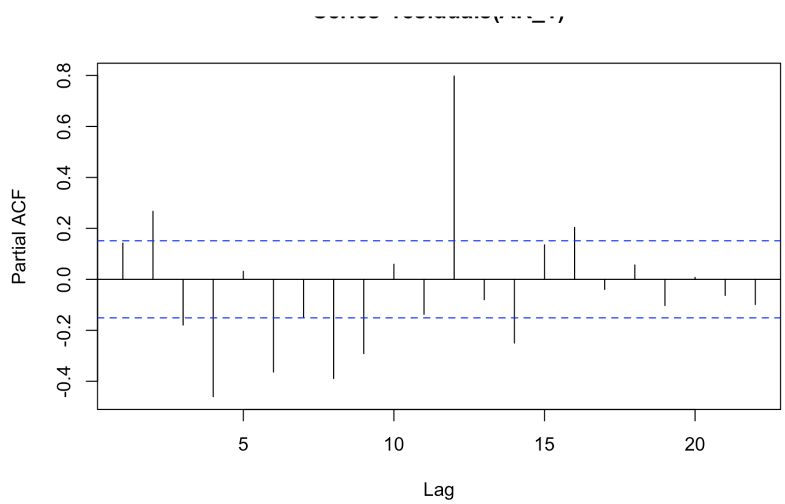
sigma^2 estimated as 1758:  log likelihood = -866.4,  aic = 1738.8
```

Here we finish fitting the AR(1) model. We wil try to plot the noew residual and compare with the results in Question 2.

```
acf(residuals(AR_1))
```




```
pacf(residuals(AR_1))
```



Compared to the results in the linear regression model, the autocorrelation seems to be less intensive. So we believe it's more close to the stationary status and does provide a better model.

```
AR_2 <- arima(milk_res$residual,order = c(2,0,0))
print(AR_2)

Call:
arima(x = milk_res$residual, order = c(2, 0, 0))

Coefficients:
         ar1      ar2  intercept
      0.8561  -0.1942    -0.4785
s.e.  0.0754   0.0761     9.3152

sigma^2 estimated as 1692:  log likelihood = -863.21,  aic = 1734.42
```

```
MA_1 <- arima(milk_res$residual, order = c(0,0,1))
print(MA_1)

Call:
arima(x = milk_res$residual, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.6158    -0.0548
s.e.  0.0579     5.8557

sigma^2 estimated as 2216:  log likelihood = -885.72,  aic = 1777.44
```

```
MA_2 <- arima(milk_res$residual, order = c(0,0,2))
print(MA_2)

Call:
arima(x = milk_res$residual, order = c(0, 0, 2))

Coefficients:
         ma1     ma2  intercept
      0.5114  0.8676    -0.7637
s.e.  0.0434  0.0410     6.9481

sigma^2 estimated as 1449:  log likelihood = -851.21,  aic = 1710.42
```

After comparing the AIC score, we can see that MA(2) model give a low value among these models. It can be considered relatively lower.

4. We can try different ARMA model. It require us to try different p and q value in the autoregressive moving average modeling process.
   We can try p=1,2 and q= 1,2, which requre us to try 4 models.

```
ARMA_1 = arima(milk_res$residual, order = c(1,0,1))
print(ARMA_1)

Call:
arima(x = milk_res$residual, order = c(1, 0, 1))

Coefficients:
         ar1     ma1  intercept
      0.6641  0.1170    -0.9145
s.e.  0.0684  0.0724    10.5463
```

```
sigma^2 estimated as 1730:  log likelihood = -865.06,  aic = 1738.12

ARMA_2 = arima(milk_res$residual, order = c(1,0,2))
print(ARMA_2)

Call:
arima(x = milk_res$residual, order = c(1, 0, 2))

Coefficients:
         ar1     ma1     ma2  intercept
      0.3352  0.4234  0.8213    -0.8263
s.e.  0.0826  0.0424  0.0790     9.3929

sigma^2 estimated as 1322:  log likelihood = -843.51,  aic = 1697.01

ARMA_3 = arima(milk_res$residual, order = c(2,0,1))
print(ARMA_3)

Call:
arima(x = milk_res$residual, order = c(2, 0, 1))

Coefficients:
          ar1     ar2     ma1  intercept
      -0.0469  0.5292  0.7952    -1.1077
s.e.   0.1589  0.1276  0.1378    11.0439

sigma^2 estimated as 1751:  log likelihood = -866.07,  aic = 1742.15

ARMA_4 = arima(milk_res$residual, order = c(2,0,2))
print(ARMA_4)

Call:
arima(x = milk_res$residual, order = c(2, 0, 2))

Coefficients:
         ar1      ar2     ma1     ma2  intercept
      0.3851  -0.1420  0.3991  0.8750    -0.5799
s.e.  0.0848   0.0971  0.0390  0.0665     8.3128

sigma^2 estimated as 1302:  log likelihood = -842.45,  aic = 1696.89
```

Just judging by the AIC score, the ARMA(2,2) model has the best performace. Also the ARMA(1,2) model has a relatively low ARMA score whose perfomance is fairly close to that of ARMA(2,2).