Ex6

June 2, 2020

0.1 Ex 6: Time Series Problem Sheet 3

In [35]:
```
lag <- 1:20
```

In [36]:
```
y.acf <- c(0.965,0.923,0.901,0.879,0.859,0.843,
        0.831,0.816,0.796,0.778,0.763,0.748,0.730,0.707,0.680,
        0.653,0.630,0.609,0.590,0.568)

w.acf <- c(0.145, -0.362, 0.004, -0.003, -0.053, -0.059, 0.047,
           0.096, -0.029, -0.057, 0.009, -0.001, 0.057, 0.084, -0.031,
           -0.031, -0.035, -0.036, 0.093, 0.038)
```

0.2 ARGH! There are 400 observations, not 20!

- They're just showing the first 20
- Note that if you have 400 observations, you're going to get 399 differenced observations, as each differenced observation requires 2 observations.

In [38]:
```
T.w <- 400
```

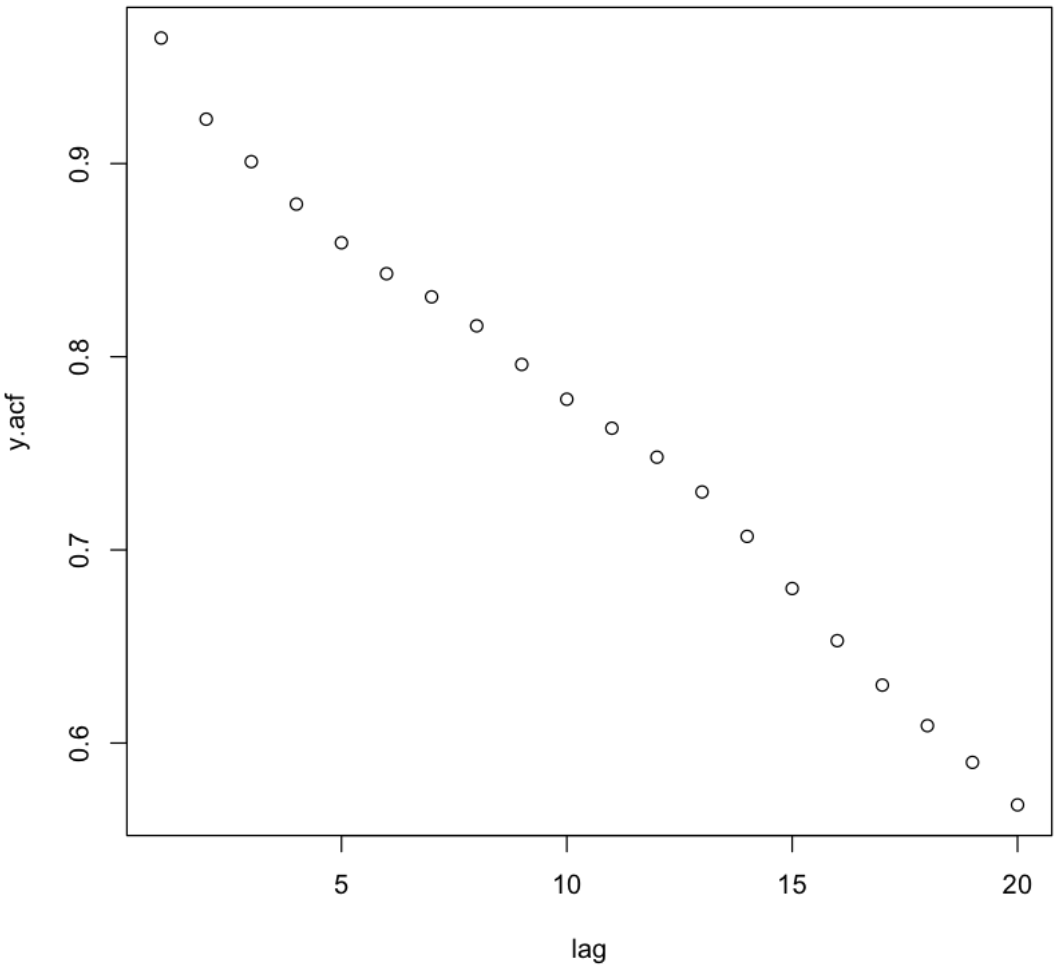
In [39]:
```
se <- 1 / sqrt(T.w)

num.se <- qnorm(0.025, lower.tail = FALSE)

CI.95 <- 0 + c(-num.se*se, num.se*se)

CI.95
```

1. -0.0979981992270027 2. 0.0979981992270027

0.2.1 ACF of $\{y_t\}$

- Linear rather than geometric decrease suggests that y_t is a non-stationary process

In [6]:
```
plot(lag, y.acf)
```



0.3 ACF of $\{W_t\} = \{\Delta Y_t\}$

- This is now a stationary process as we no longer see a linear trend
- We now see scattered points around the zero mean
- Plotting the 95% confidence interval for a white noise process, we can see that all points lie within that confidence interval (we would expect one of the twenty points to lie outside it, and we see none).
- Therefore we identify $\{\Delta Y_t\}$ as a white noise process
- $\{Y_t\}$ is a non-stationary ARIMA(1,1,0) model:

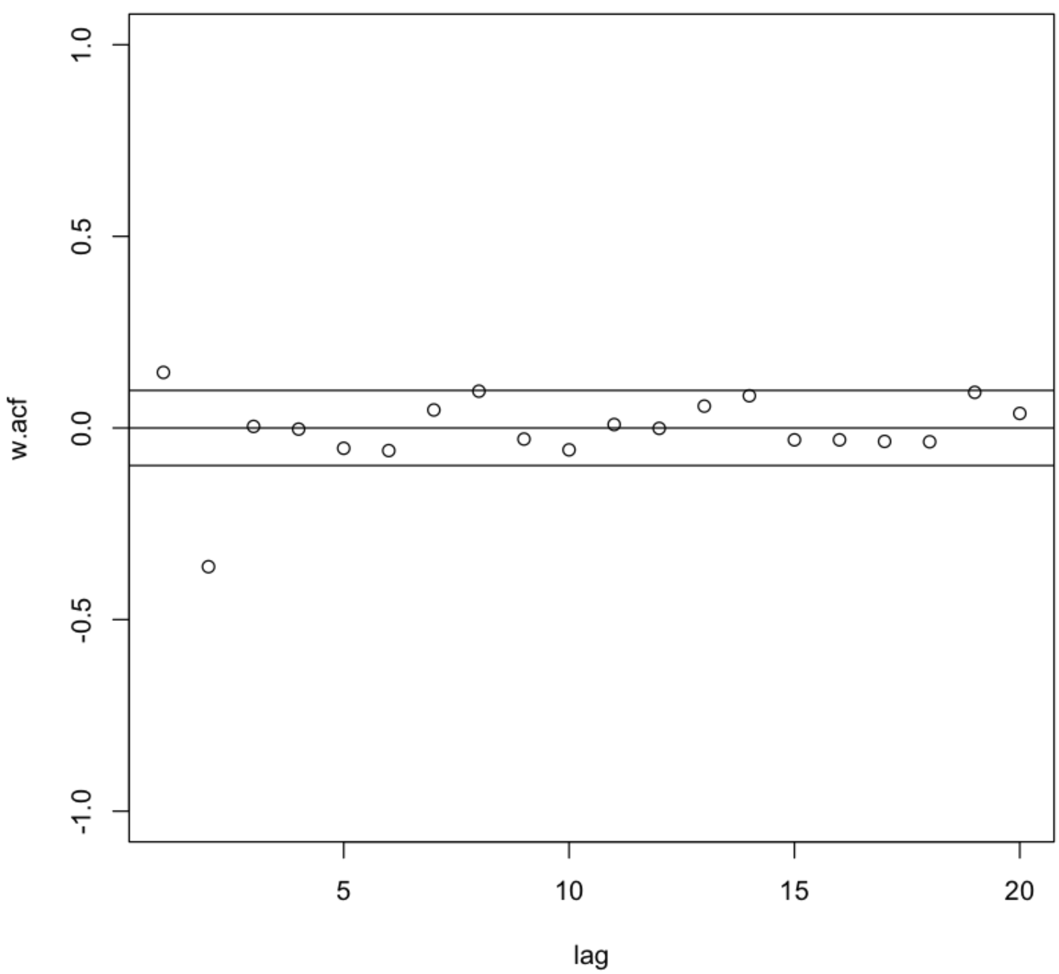
  $W_t = \epsilon_t$

  $(1 - L)Y_t = \epsilon_t$

$Y_t = \phi Y_{t-1} + \epsilon_t$

where $\phi = 1$, and is this **non-stationary**.

```
In [40]: plot(lag, w.acf, ylim = c(-1,1))
         abline(h=0)
         abline(h=CI.95[1])
         abline(h=CI.95[2])
```



1 INCORRECT!

Should have found that there were two points and a sharp cut off, highlighting an MA(2) process.

Then fit the confidence intervals around a suspected MA(2):

95% confidence interval around a suspected MA(q) process:

$$\pm 2\sqrt{\frac{1+2(r_1^2+r_2^2+...+r_q^2)}{T}}$$

We suspect MA(2), therefore let's fit a CI around:

$$\pm 2\sqrt{\frac{1+2(r_1^2+r_2^2)}{T}}$$

```
In [41]: T <- 399
```

```
In [42]: num.se
```

1.95996398454005

```
In [43]: se <- sqrt( (1 + 2*(0.145**2 + 0.362**2)) / T)
         se
```

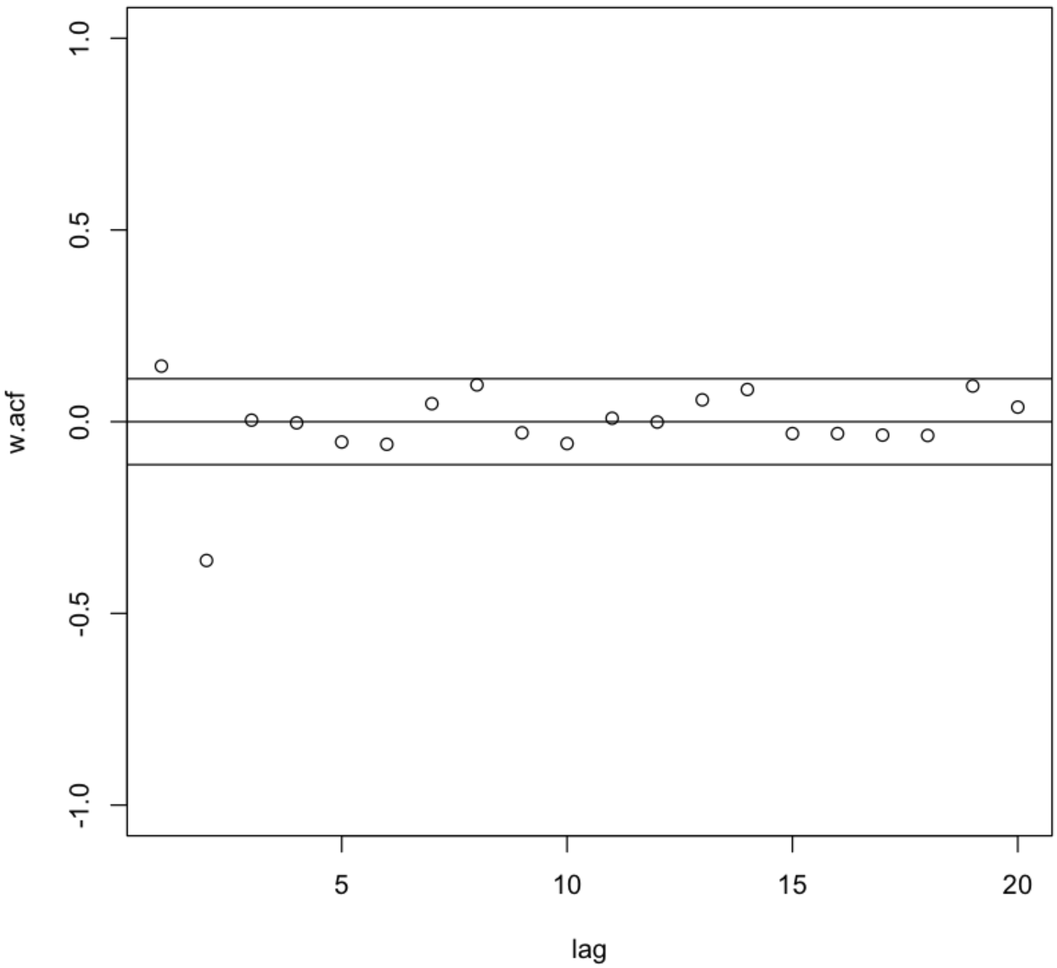
0.0571709392150139

```
In [44]: CI.95 <- 0 + c(-num.se*se, num.se*se)
```

```
In [45]: CI.95
```

1. -0.112052981823756 2. 0.112052981823756

```
In [46]: plot(lag, w.acf, ylim = c(-1,1))
         abline(h=0)
         abline(h=CI.95[1])
         abline(h=CI.95[2])
```



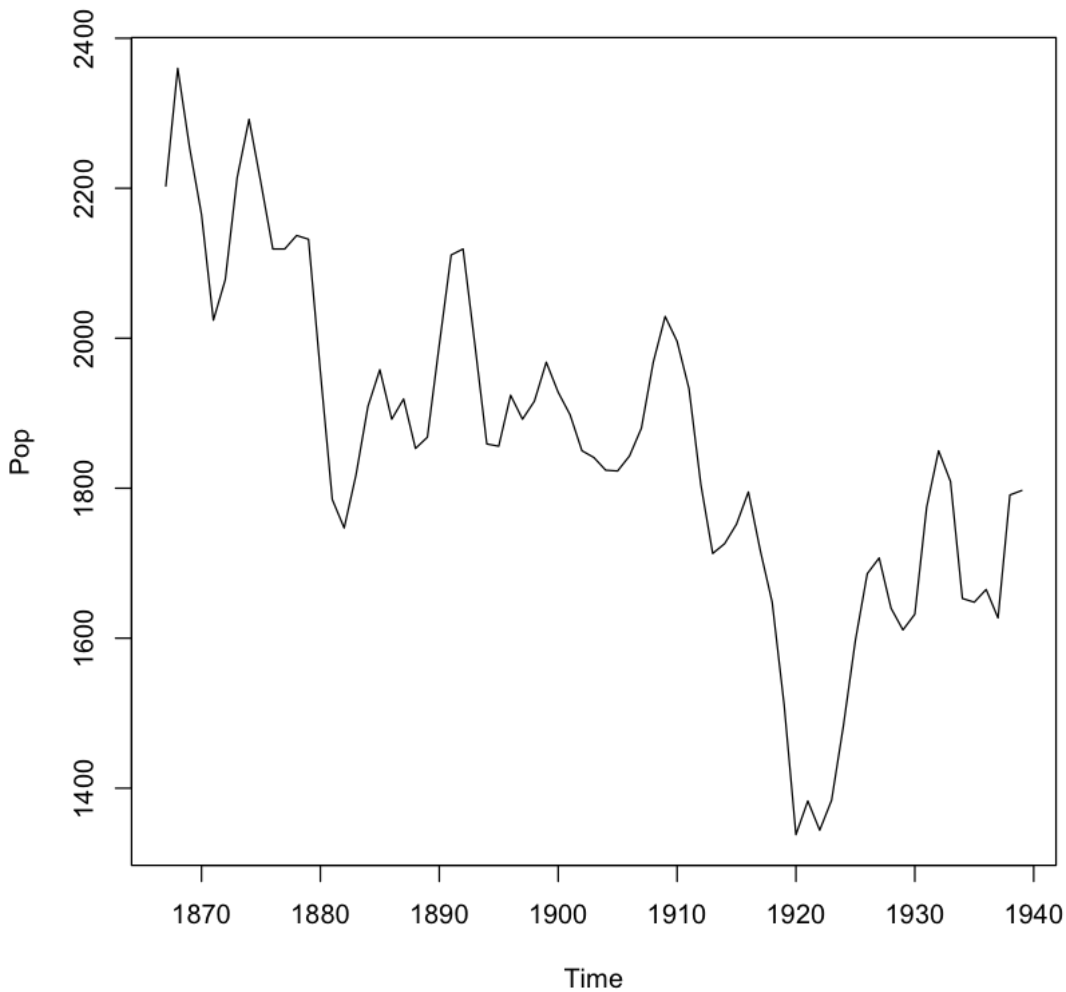
1.1 New Summary:

- Can see that the data supports our MA(2) hypothesis for W_t
- Thus there's evidence to suggest Y_t is an ARIMA(0,1,2) process.

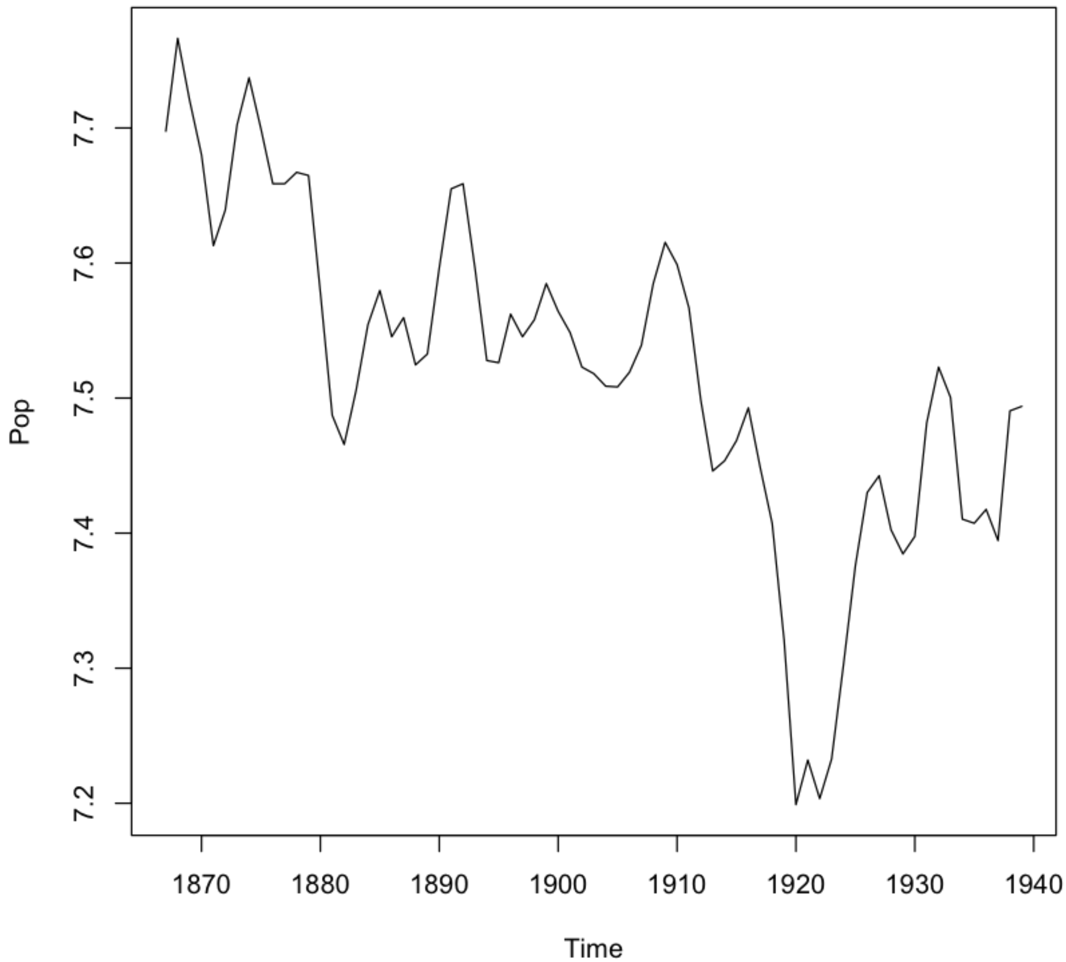
1.2 Q2) Sheep Data

In [95]:
```
sheep <- read.csv('sheepdata.txt', header = FALSE)
colnames(sheep) <- c('Pop')

sheep.ts <- ts(sheep, start = 1867, frequency = 1)
```

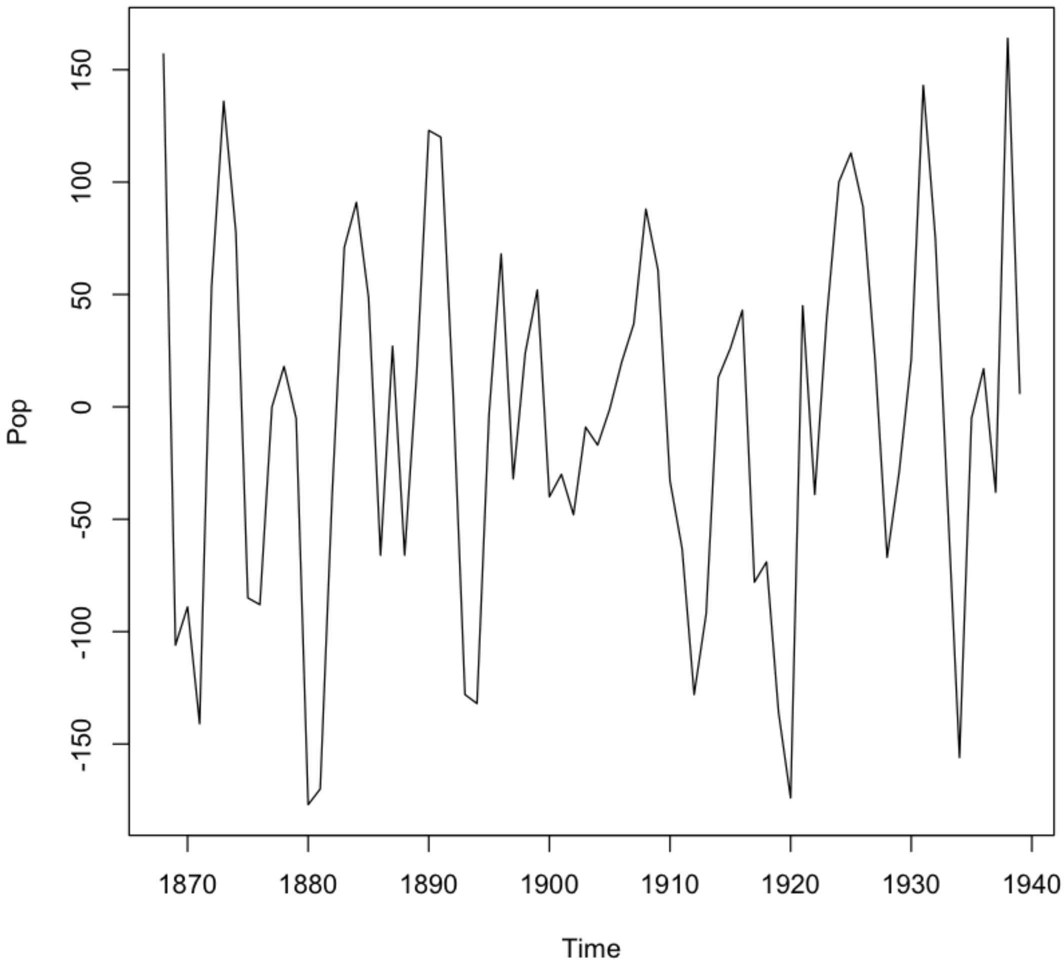
```
        sheep.log.ts <- log(sheep.ts)

        sheep.lag.1.ts <- diff(sheep.ts, lag=1)

In [96]: plot(sheep.ts)
```




```
In [97]: plot(sheep.log.ts)
```



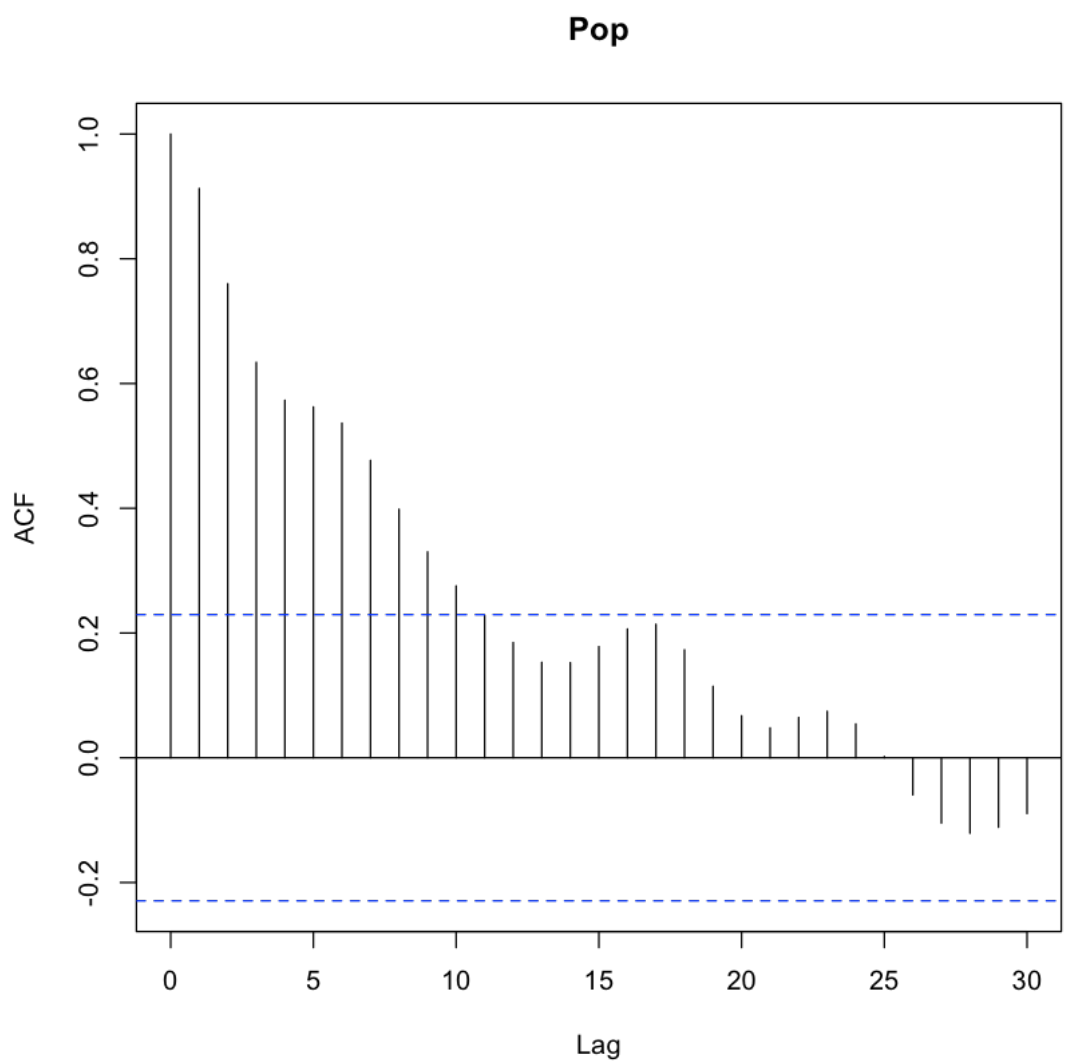

```
In [98]: plot(sheep.lag.1.ts)
```



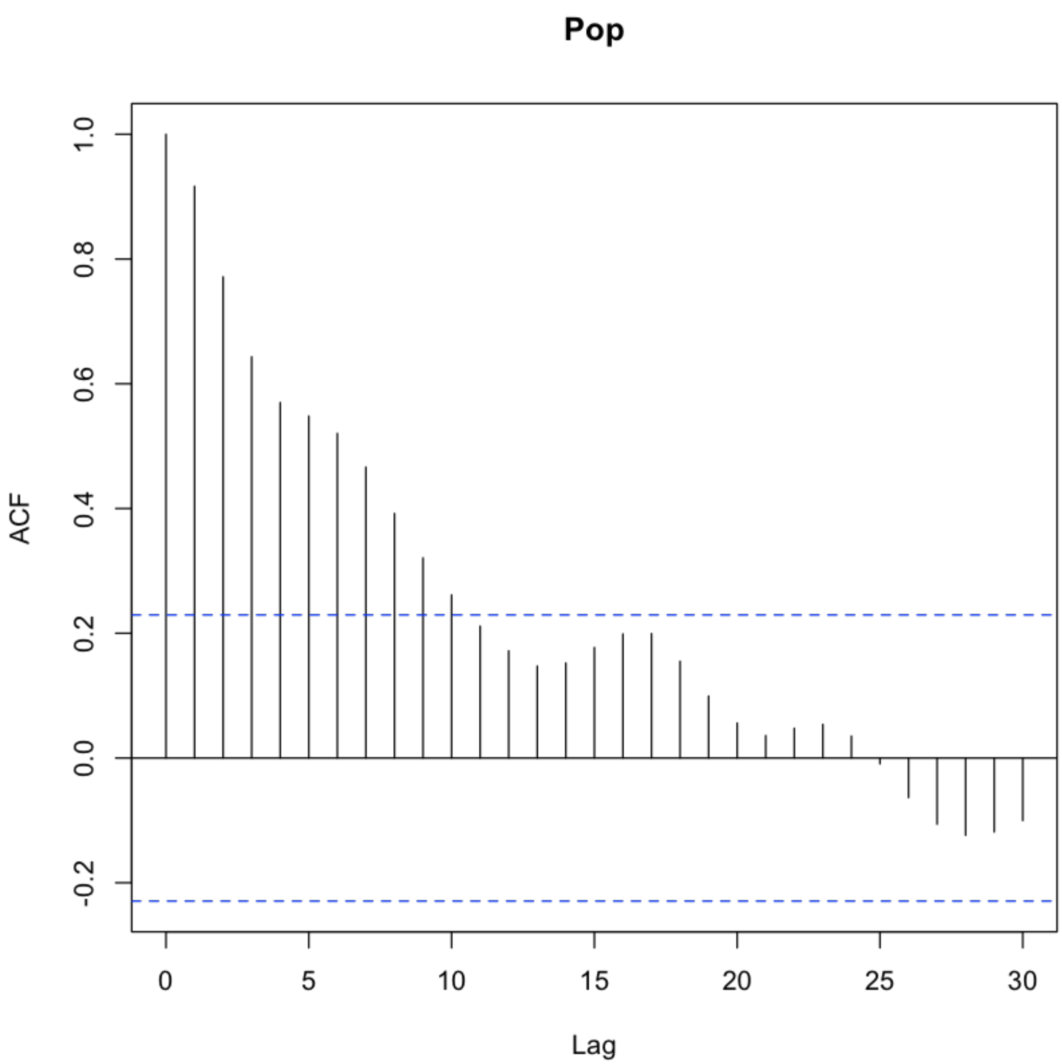
1.3 1) Is it stationary or not?

- Have looked at sample ACFs for both the process and the log of the process, and it looks like there's a linear decrease in the ACF that suggests the process is non-stationary

```
In [99]: acf(sheep.ts, lag = 30)
```



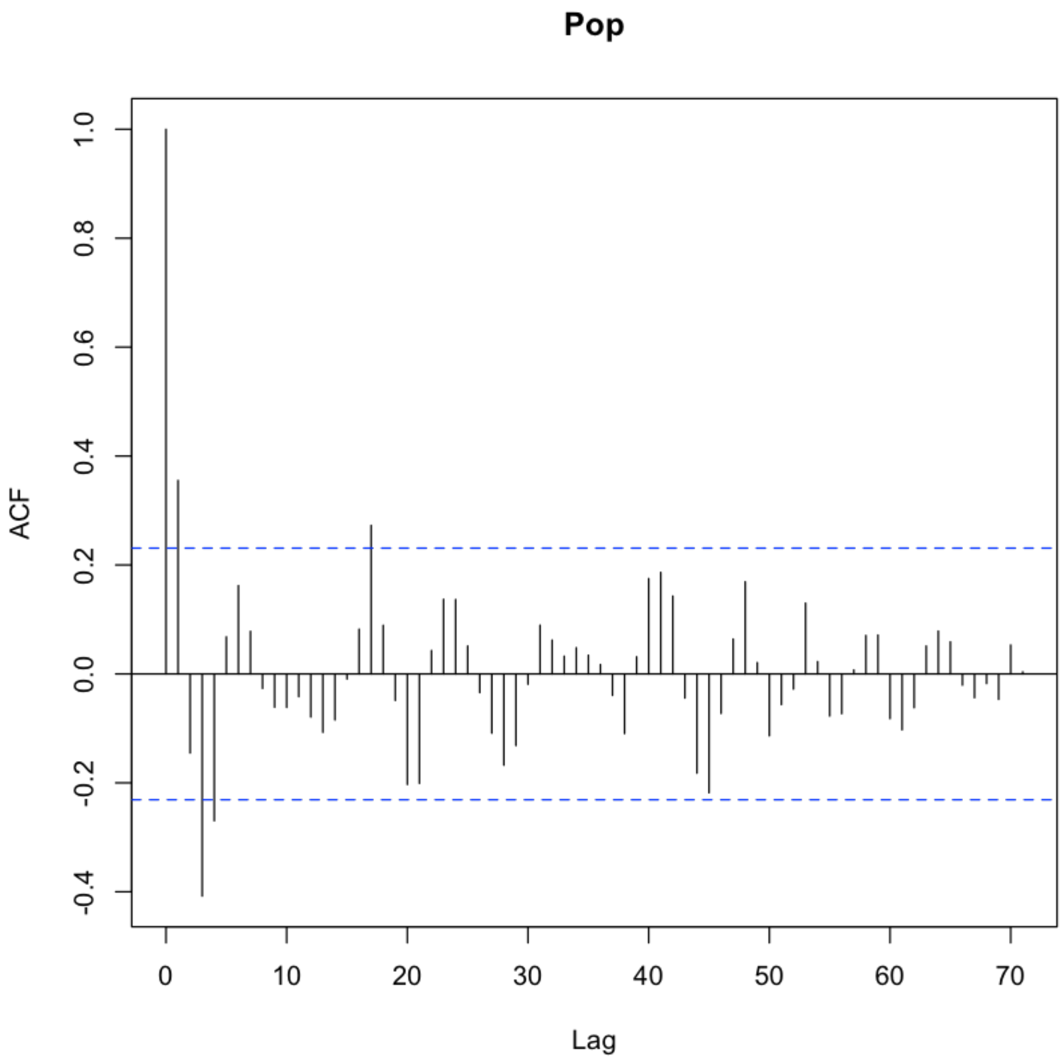

```
In [100]: acf(sheep.log.ts, lag=30)
```



1.4 Not a stationary process.

1.5 Must apply differencing to transform into a stationary process.

```
In [103]: acf(sheep.lag.1.ts, lag=100)
```



1.6 Strongly suggests an ARIMA model

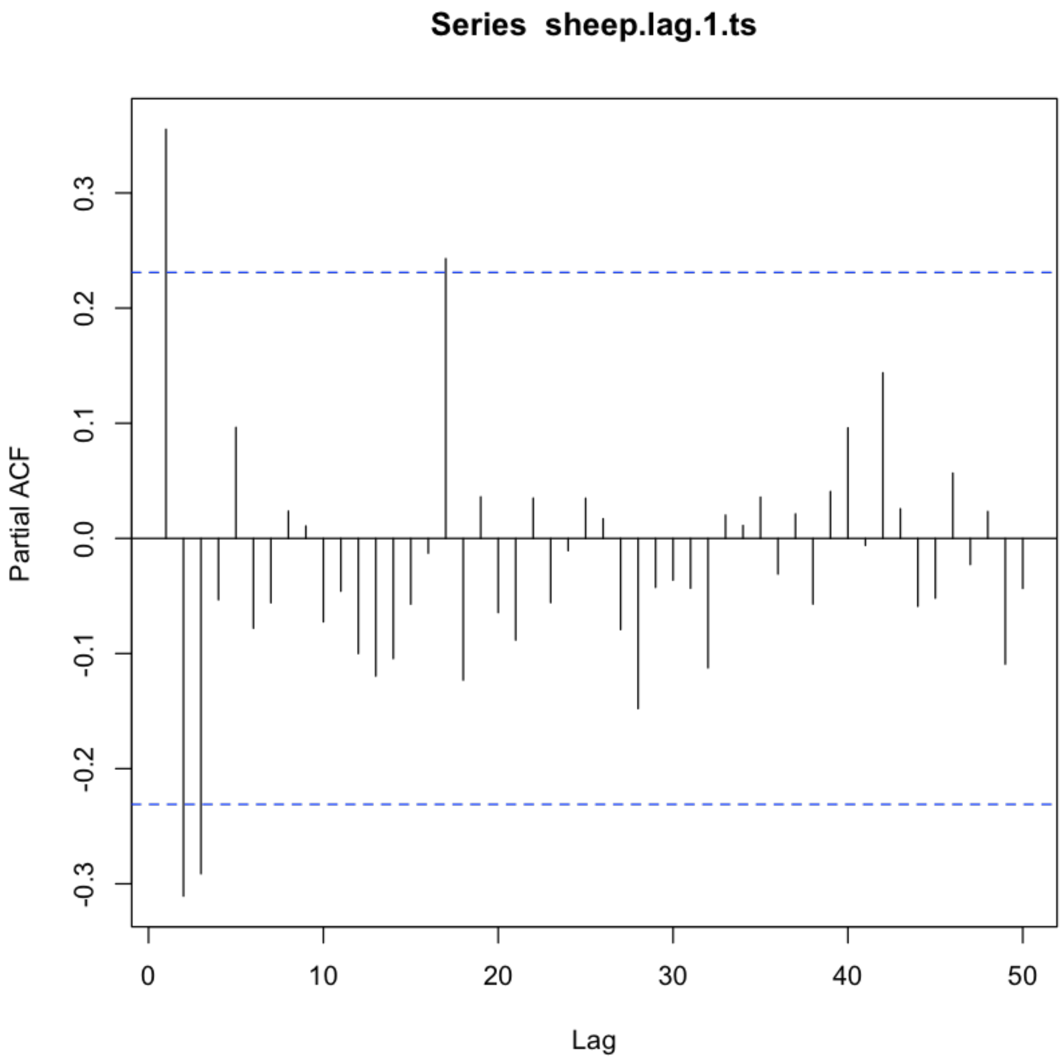
ACF shows geometric progression, so unlikely to be MA

Shows values above the 95% confidence interval for a white noise process

So the differenced process is either AR(p) or ARMA(p,q)

Suspect it's AR(3)

```
In [81]: pacf(sheep.lag.1.ts, lag=50)
```



2 Fitting ARIMA models

```
In [104]: sheep.3.1.0 <- arima(sheep.ts, order=c(3,1,0))
          sheep.3.1.0
```

```
Call:
arima(x = sheep.ts, order = c(3, 1, 0))

Coefficients:
         ar1      ar2      ar3
      0.4210  -0.2018  -0.3044
```

```
s.e.  0.1193   0.1363   0.1243

sigma^2 estimated as 4783:  log likelihood = -407.56,  aic = 823.12
```

```
In [105]: sheep.3.1.1 <- arima(sheep.ts, order=c(3,1,1))
          sheep.3.1.1
```

```
Call:
arima(x = sheep.ts, order = c(3, 1, 1))

Coefficients:
         ar1      ar2      ar3      ma1
      0.4927  -0.2411  -0.2770  -0.0766
s.e.  0.3128   0.2141   0.1724   0.3055

sigma^2 estimated as 4779:  log likelihood = -407.53,  aic = 825.05
```

```
In [106]: sheep.4.1.0 <- arima(sheep.ts, order=c(4,1,0))
          sheep.4.1.0
```

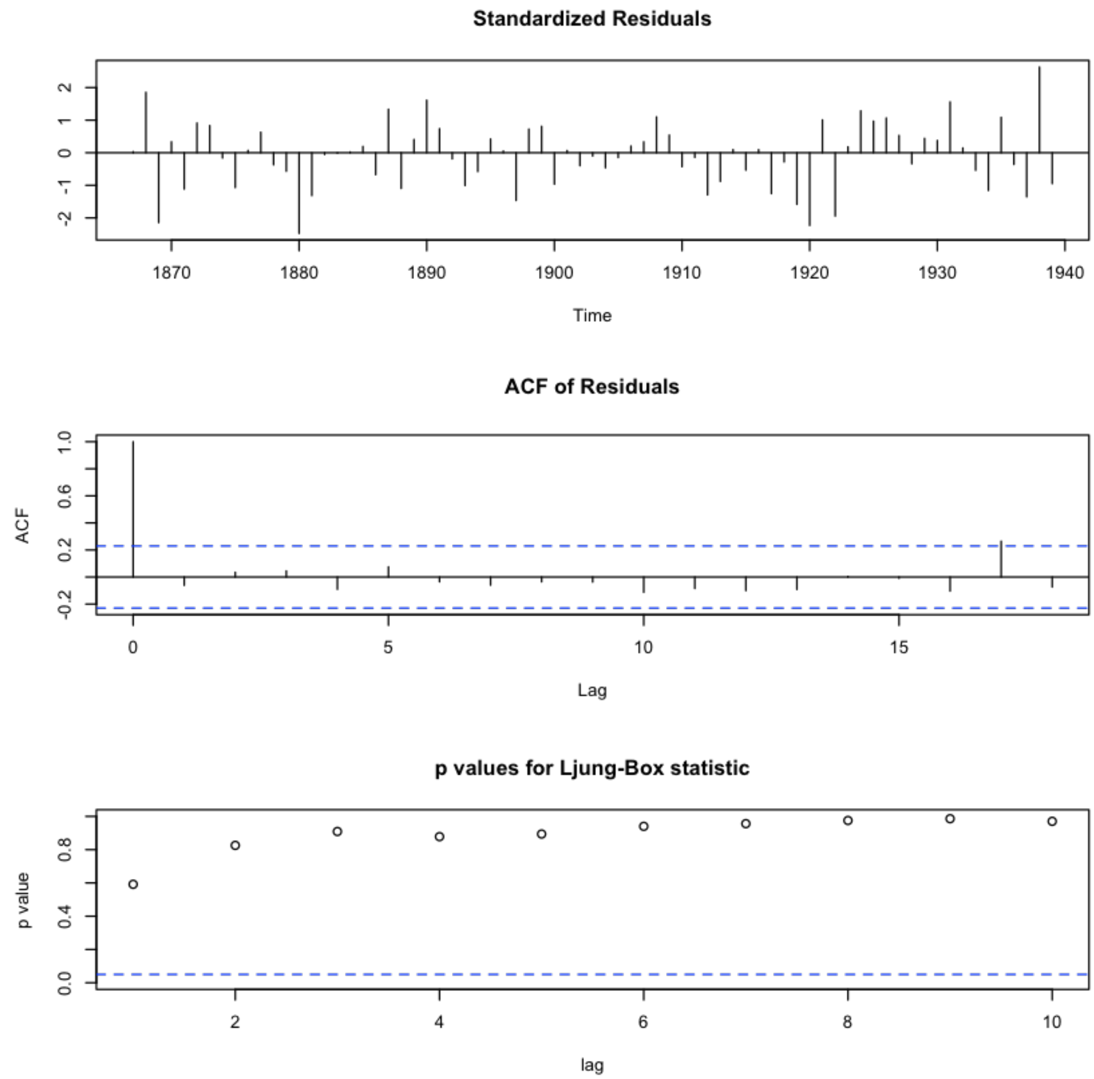
```
Call:
arima(x = sheep.ts, order = c(4, 1, 0))

Coefficients:
         ar1      ar2      ar3      ar4
      0.4117  -0.2176  -0.2779  -0.0481
s.e.  0.1219   0.1433   0.1441   0.1323

sigma^2 estimated as 4774:  log likelihood = -407.49,  aic = 824.98
```

```
In [108]: sheep.4.1.1 <- arima(sheep.ts, order=c(4,1,1))
          sheep.4.1.1
```

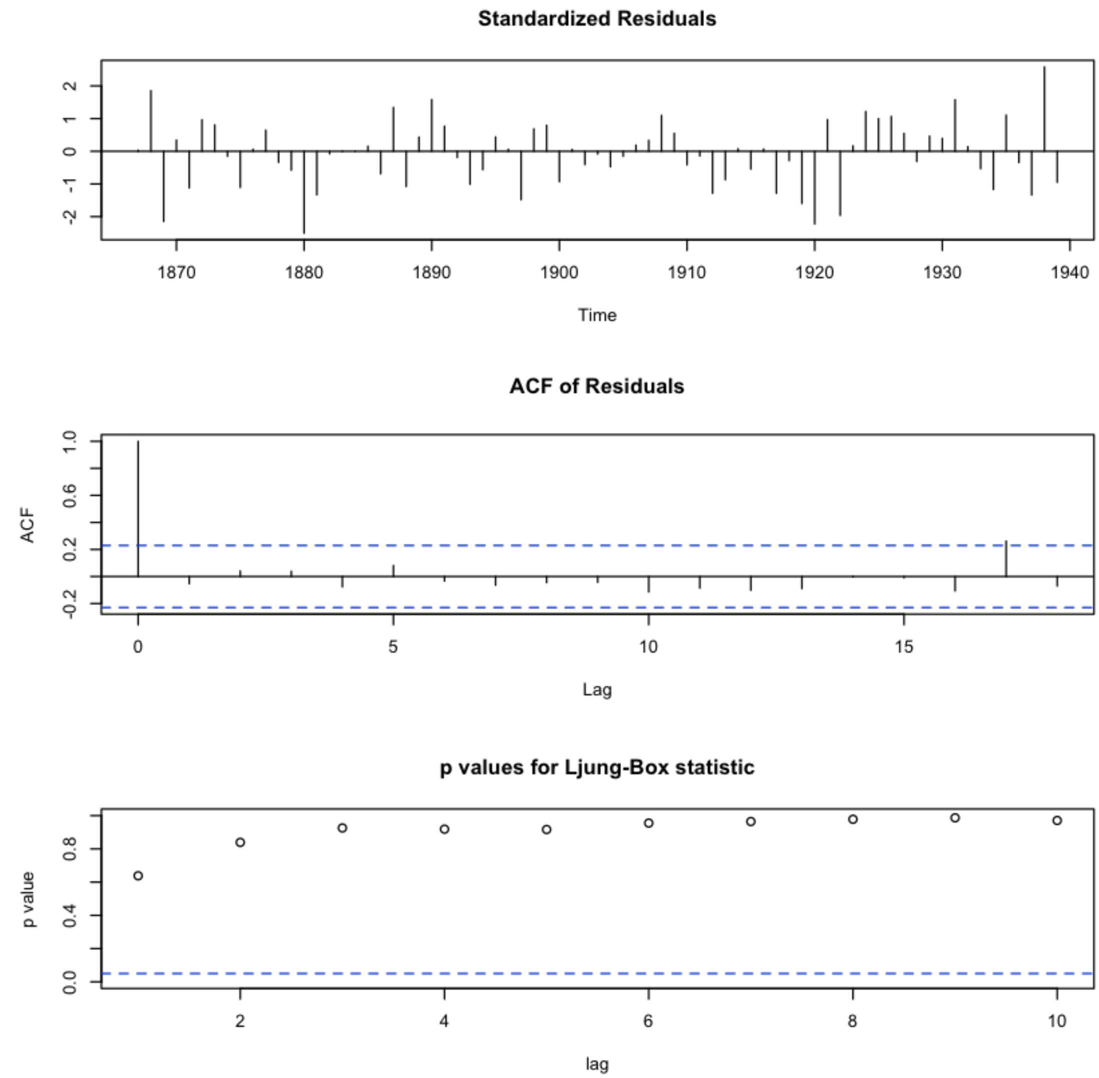
```
Call:
arima(x = sheep.ts, order = c(4, 1, 1))

Coefficients:
          ar1     ar2      ar3      ar4     ma1
      -0.5117  0.1725  -0.4578  -0.3496  0.9448
s.e.   0.1905  0.1574   0.1389   0.1240  0.1602

sigma^2 estimated as 4621:  log likelihood = -406.51,  aic = 825.03
```

2.1 Firstly let's take a look at some diagnostics for the four potential models:

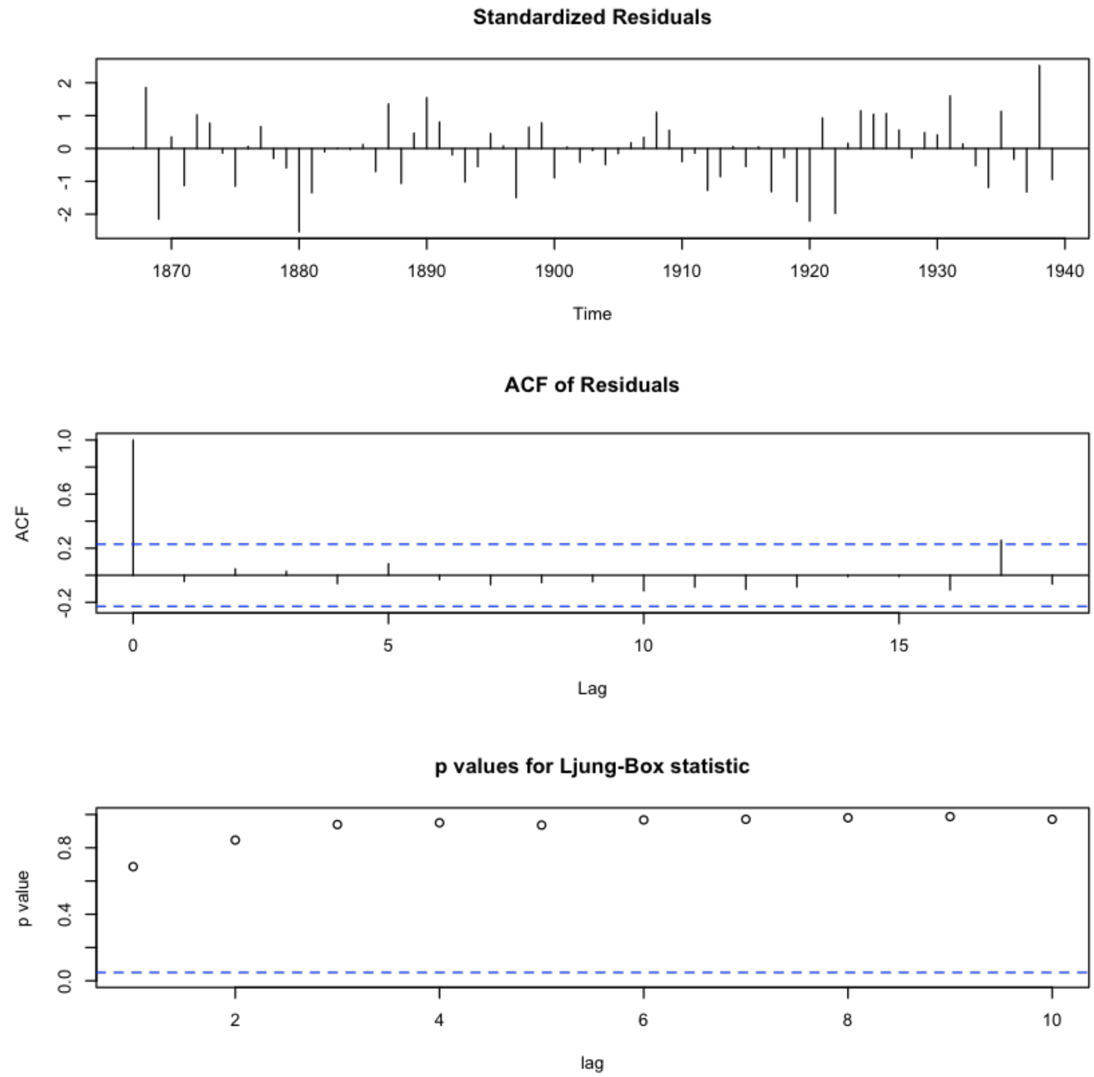
```
In [109]: tsdiag(sheep.3.1.0)
```



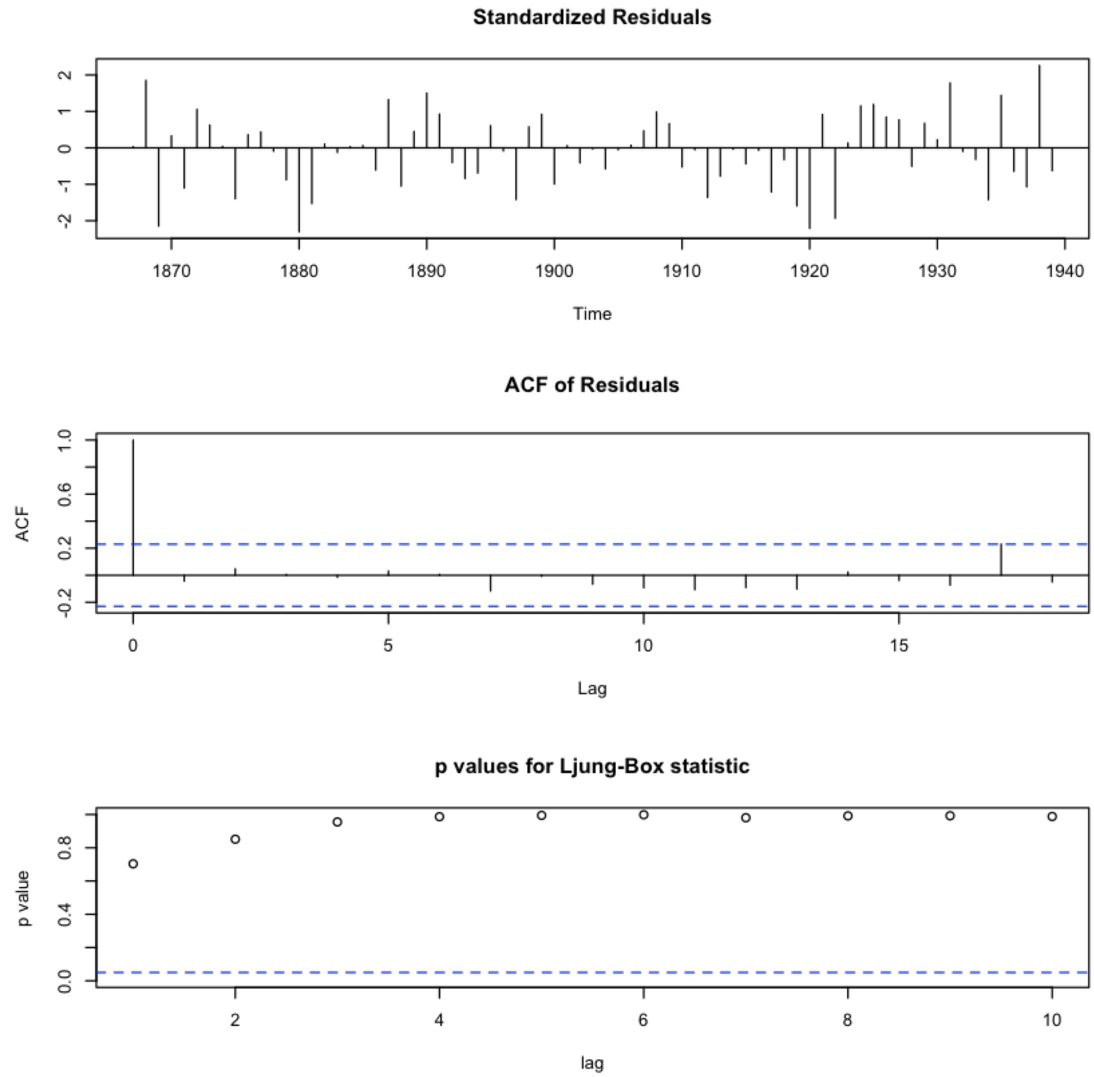

```
In [111]: tsdiag(sheep.3.1.1)
```




```
In [112]: tsdiag(sheep.4.1.0)
```




```
In [113]: tsdiag(sheep.4.1.1)
```



3 Summary:

- All four models pass the plot adequecy tests.
- Lowest AIC is ARIMA(3,1,0)
- Will look at the significance of the 4th parameter in ARIMA(4,1,0) to see if significantly differs from 0 in an overfitting test.

3.1 Looking at ARIMA(4,1,0)

```
In [115]: sheep.4.1.0
```

```
Call:
arima(x = sheep.ts, order = c(4, 1, 0))

Coefficients:
         ar1      ar2      ar3      ar4
      0.4117  -0.2176  -0.2779  -0.0481
s.e.  0.1219   0.1433   0.1441   0.1323

sigma^2 estimated as 4774:  log likelihood = -407.49,  aic = 824.98
```

```
In [117]: t <- -0.0481 / 0.1323
```

```
In [119]: t
```

-0.363567649281935

```
In [120]: T <- length(sheep.ts)
          T
```

73

```
In [121]: # degrees of freedom:
          ## T - 4 (parameters)
          ## - 1 (differencing so really our modified T = T-1)
          ## - 1 (mean) - 1 (extra)
          df <-  T - 4 - 1 - 1 - 1
          df
```

66

```
In [122]: 2*pt(t, df, lower.tail=TRUE)
```

0.717342885357635

Cannot reject the null hypothesis that the fourth AR parameter is equal to zero, thus we accept the alternative that $\phi_4 = 0$

3.2 Model equation and fitted equation:

ARIMA(3,1,0)

$$W_t = (1 - L)Y_t$$

ARMA(3,0) for $\{W_t\}$

$$\phi(L)(W_t - \mu) = \theta(L)\epsilon_t$$

For ARMA(3,0):

$$\theta(L) = 1$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3$$

Full ARMA(3,0):

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - L)Y_t = \epsilon_t$$

$$Y_t = (1 + \phi_1)Y_{t-1} - (\phi_1 - \phi_2)Y_{t-2} - (\phi_2 - \phi_3)Y_{t-3} - \phi_3 Y_{t-4} + \epsilon_t$$