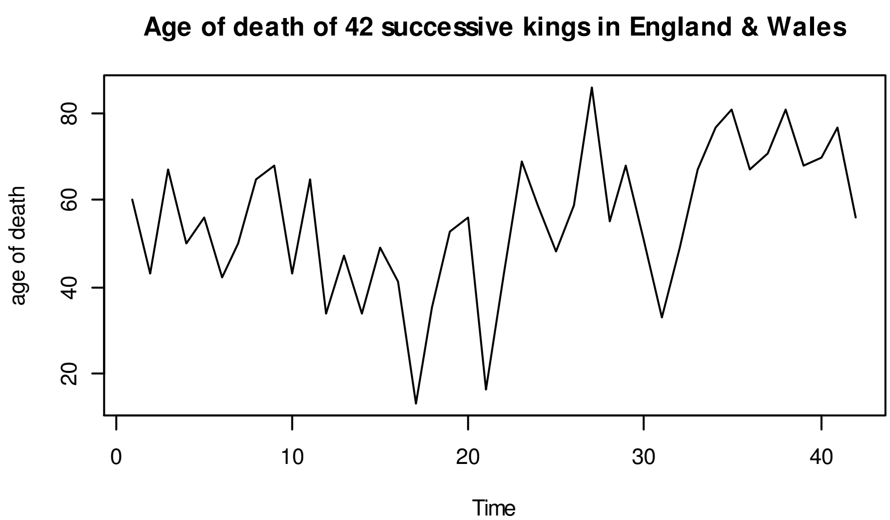
First load forecast package

```
x=c(60,43,67,50,56,42,50,65,68,43,65,34,47,34,49,41,13,35,
+ 53,56,16,43,69,59,48,59,86,55,68,51,33,49,67,77,81,67,71,81,
+ 68,70,77,56)

x( age of death of 42 successive kings in England)
60 43 67 50 56 42 50 65 68 43 65 34 47 34 49 41 13 35 53 56 16 43 69 59 48
59 86 55 68 51 33 49 67 77 81 67 71 81 68 70 77 56

y=ts(x)# create a time series object
```



- **From the diagram, we can see that the time series is not stationary in mean.**

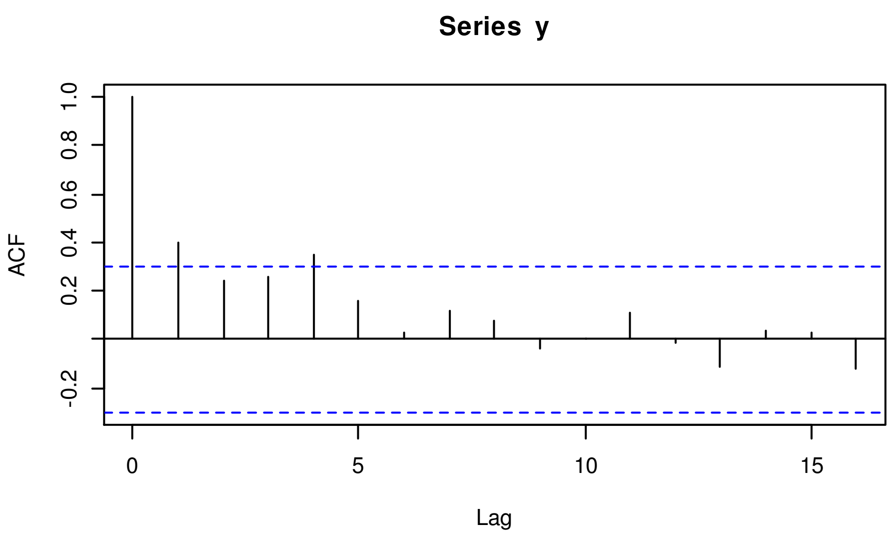
Let us now look at the acf and pacf values and the correlograms

```
acf(x,lag.max=15,plot=F)#give the autocorrelation values

Autocorrelations of series 'y', by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.401  0.238  0.260  0.348  0.161  0.031  0.115  0.078 -0.036 -0.001
    11     12     13     14     15
 0.111 -0.010 -0.117  0.035  0.025
```

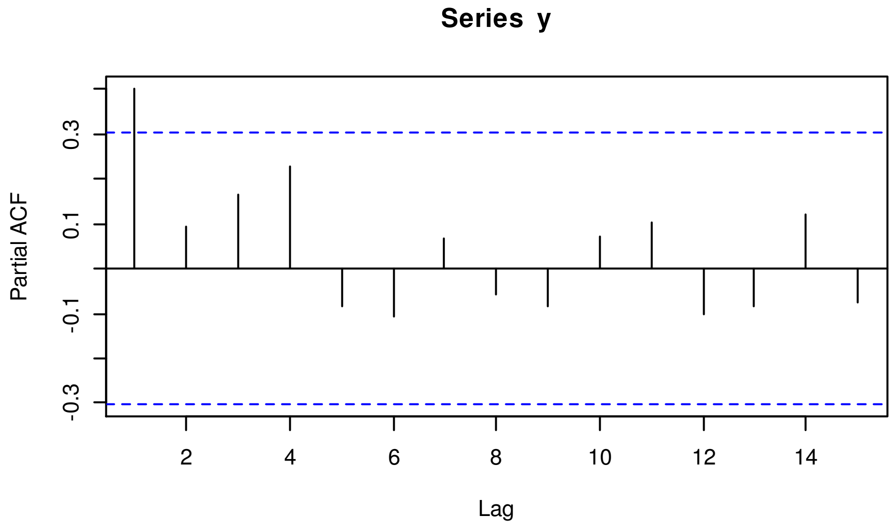
```
pacf(x,lag.max=15,plot=F)# gives the partial autocorrelation values
Partial autocorrelations of series 'y', by lag

     1      2      3      4      5      6      7      8      9     10     11 
 0.401  0.092  0.163  0.227 -0.085 -0.108  0.067 -0.058 -0.082  0.071  0.104 
    12     13     14     15 
-0.103 -0.084  0.119 -0.077 
```

- The top rows are lags and the bottom rows are correlation values.
- We can see from the observed values of the correlations(acf/pacf) that the series is not stationary

```
acf(y,lag.max=15)# gives the correlogram
```

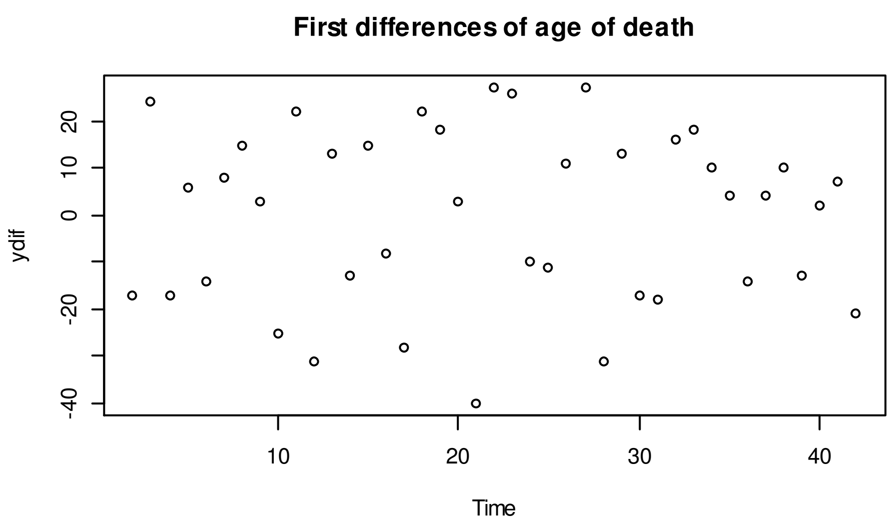


```
pacf(x,lag.max=15,plot=F)# gives the partial autocorrelation values
```



Take the first difference

```
ydif=diff(y,differences=1)
plot(ydif,type="p",main="First differences of age of death")
```



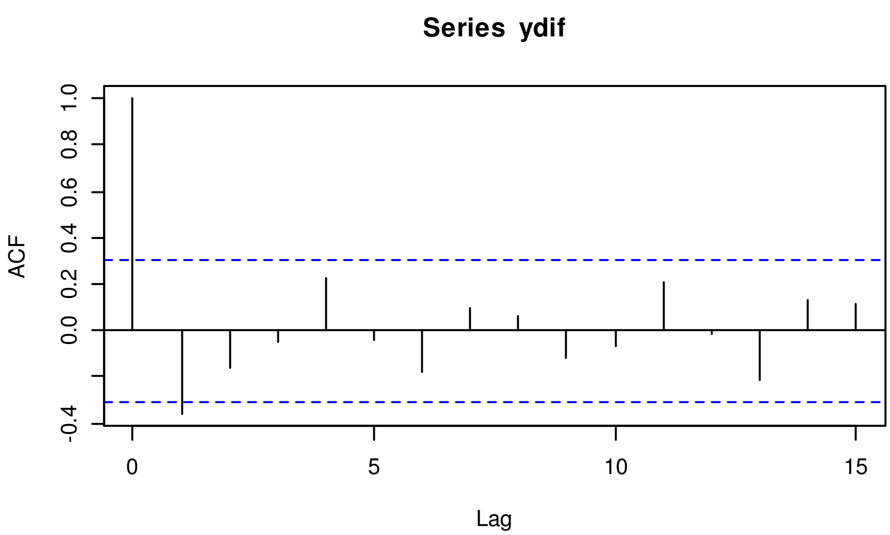
- **The series of the first differences appears to be stationary in mean and variance.**
- ***The first differences have removed the trend component in the series.***

- ***We can now examine whether there are correlations between successive terms of the irregular component; if so, would aid us to make predictive model for the ages at death of the kings.***

```
acf(ydif,lag.max=15,plot=F)

Autocorrelations of series 'ydif', by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000 -0.360 -0.162 -0.050  0.227 -0.042 -0.181  0.095  0.064 -0.116 -0.071
    11     12     13     14     15
 0.206 -0.017 -0.212  0.130  0.114
```

- From the first differences, correlation values after lag 1 is relatively small as lag increases.

```
acf(ydif,lag.max=15)
```



- From the correlogram the autocorrelation at lag 1 (-0.360) exceeds the significance bounds, whereas all others are within. This suggests that the series is stationary.

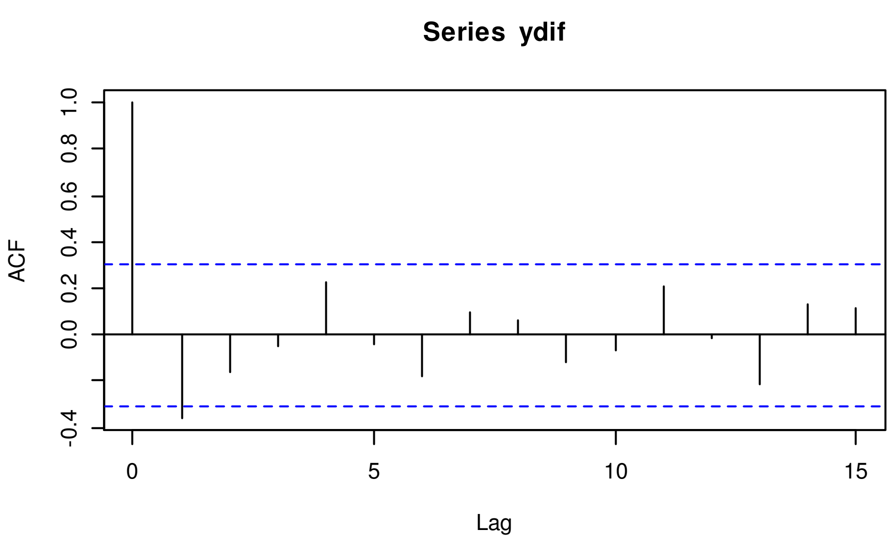
- To do a comparison of the acfs of the original series and the first differences. Then we should ensure that the both y-axes and x-axes of both series are similar.

```
To do this:-
acf(y,lag.max=15,ylim=c(-0.4,0.4))#ylim adjust the y-axes, while xlim
acf(ydif,lag.max=15,ylim=c(-0.4,0.4))# will adjust the x-axes
```

Selecting candidate ARIMA model

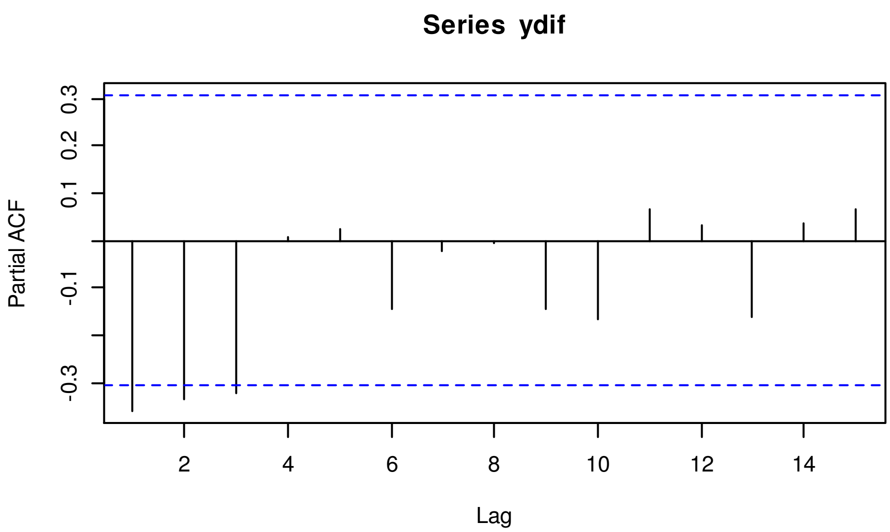
- If your time series is stationary, or you have transformed it to be a stationary time series.
- The next step is to select the appropriate ARIMA model, which means finding the most appropriate values of p and q for an ARIMA (p,d,q) model.
- To do this, we examine the acf and pacf of the stationary time series.

```
acf(ydif,lag.max=15,plot=F)
acf(ydif,lag.max=15)

Autocorrelations of series 'ydif', by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000 -0.360 -0.162 -0.050  0.227 -0.042 -0.181  0.095  0.064 -0.116 -0.071 
    11     12     13     14     15 
 0.206 -0.017 -0.212  0.130  0.114 
```



```
pacf(ydif,lag.max=15,plot=F)
pacf(ydif,lag.max=15)
```

```
Partial autocorrelations of series 'ydif', by lag

     1      2      3      4      5      6      7      8      9     10     11
-0.360 -0.335 -0.321  0.005  0.025 -0.144 -0.022 -0.007 -0.143 -0.167  0.065
    12     13     14     15
 0.034 -0.161  0.036  0.066
```



- Notice that pacf of the first differences, shows that the first three lags exceed the significance bounds and are negative, and are slowly decreasing in magnitude as the lag increases. Thus, the pacf tails off to zero after lag 3

- This is evident from the partial correlogram

How do we choose a model to fit our data:-

(i) We will use the acf and pacf of our **stationary series**

Since the correlogram is zero after lag 1, and the partial correlogram tails off to zero after lag 3. This means that the following ARMA models are possible for the data:-

- ARMA(3,0) that is, AR(3) since the pacf tails to zero after lag 3, and the acf tails to zero

- ARMA(0,1) that is MA(1) since the acf is zero after lag 1 and the pacf tails off to zero

(ii) We use the principle of parsimony to decide which model is best; that is, we assume that the model with the fewest parameters is best.

(iii) In addition, our selection criteria; i.e, the lowest AICc and/or BICc values

(iv) Using the principle of parisomony, then ARMA(0,1) will be selected since it has the 1 parameter compared to ARMA(3,0) which has 3 parameters. Or one could choose from a family of ARIMA

General Behaviour of the ACF and PACF for ARMA Models			
	AR(p)	**MA(q)**	**ARMA(p,q), p>0 & q>0**
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

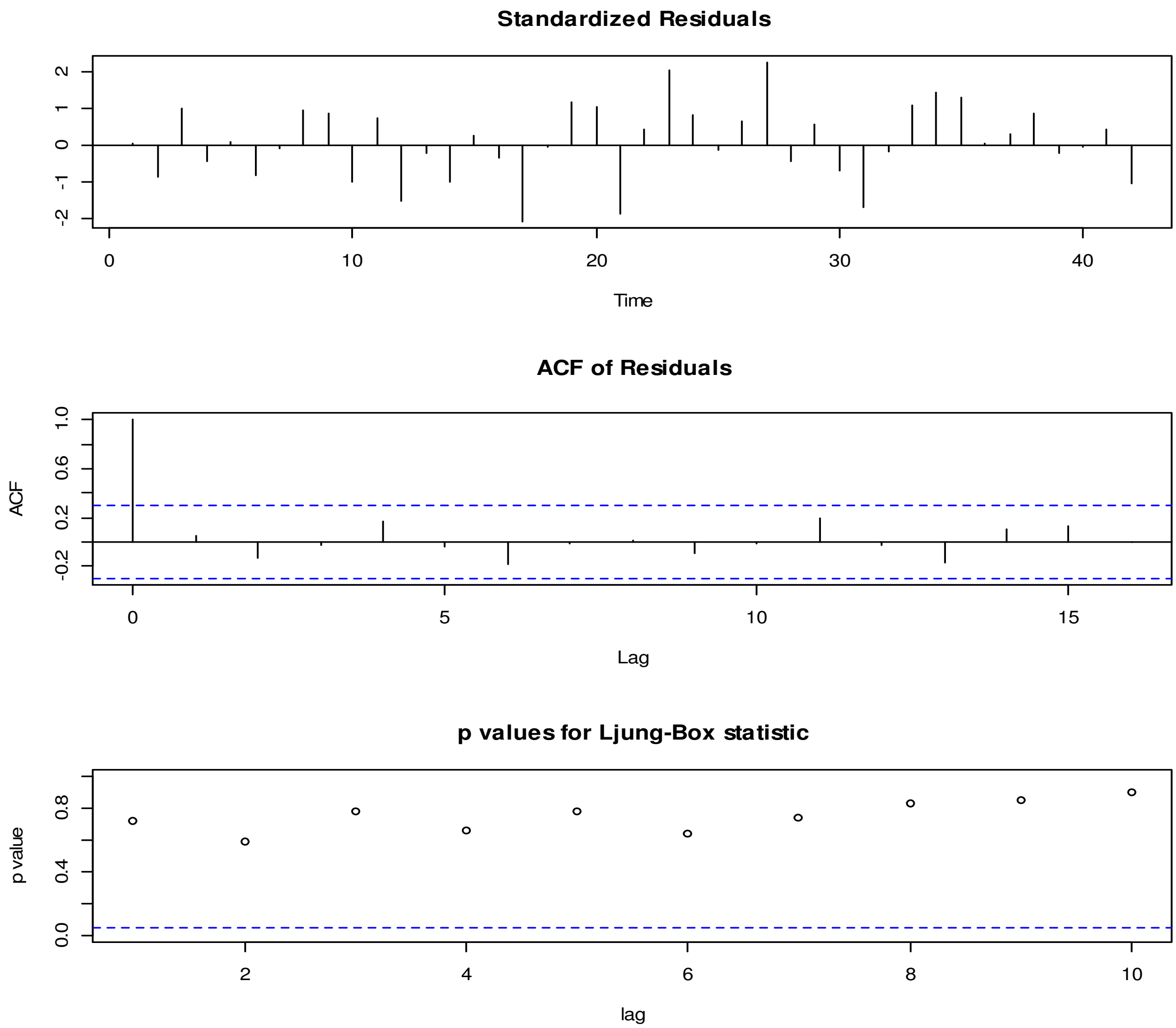
Now ARIMA(p,d,q):- autoregressive integrated moving average where 'p' is for AR, 'q' for MA and 'd' for the differencing

To fit both models in R, we use the original series of the ages of death and model the time series as an ARIMA (3,1,0) or ARIMA(0,1,1).

```
mod=Arima(y,order=c(0,1,1))
summary(mod)
Series: y
ARIMA(0,1,1)

Coefficients:
          ma1
      -0.7218
s.e.   0.1208

sigma^2 estimated as 230.4:  log likelihood=-170.06
AIC=344.13   AICc=344.44   BIC=347.56

Training set error measures:
                    ME     RMSE       MAE       MPE    MAPE      MASE
Training set 0.9712931 14.99836 11.92162 -10.40664 29.5176 0.9047383
```

```
Alternative way:-
m1=Arima(ydif,order=c(0,0,1),include.mean=0)
summary(m1)
Series: ydif
ARIMA(0,0,1) with zero mean

Coefficients:
          ma1
      -0.7218
s.e.   0.1208

sigma^2 estimated as 230.4:  log likelihood=-170.06
AIC=344.13   AICc=344.44   BIC=347.56

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 0.9935174 15.18016 12.21093 89.97649 100.7174 0.4953723 0.05200044
```

```
tsdiag(mod)
```



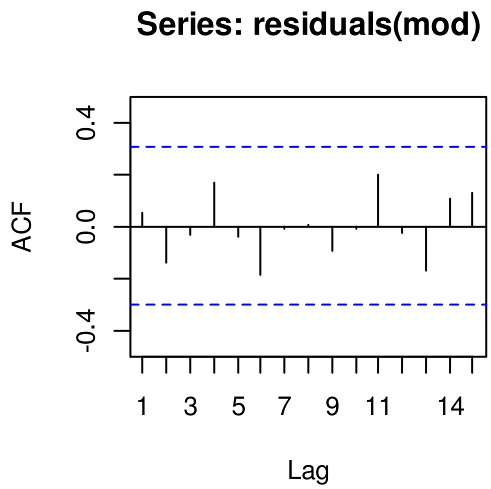
- Inspection of the time plot of the standardized residuals shows no obvious patterns.
- The ACF of the standardized residuals shows no apparent departure from the model assumptions.
- The Q-statistic is insignificant at the lags shown.

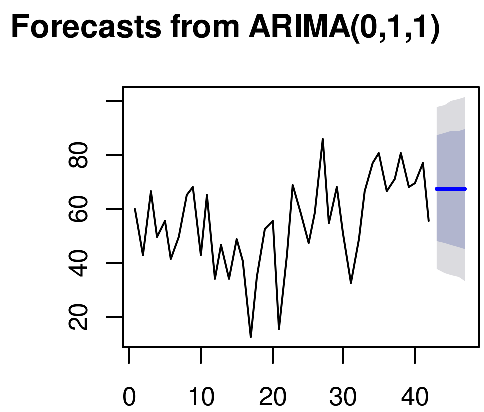
```
Acf(residuals(mod))

Box.test(mod$residuals,lag=20,type="Ljung-Box")

        Box-Ljung test

data:  mod$residuals
X-squared = 13.5844, df = 20, p-value = 0.8509
```

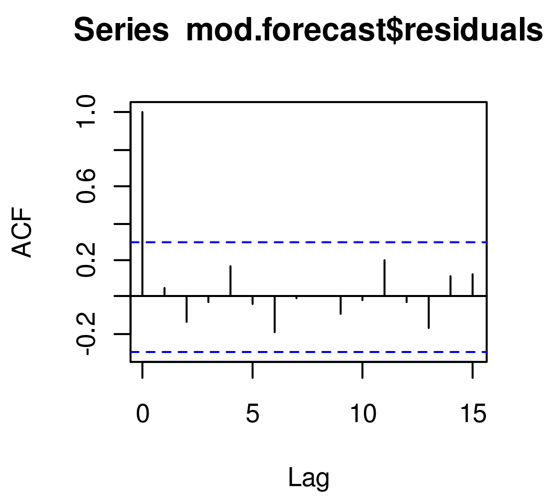
- **The test suggests that the model is specified correctly**

```
mod.forecast=forecast.Arima(mod,h=5)
mod.forecast
   Point Forecast    Lo 80    Hi 80    Lo 95     Hi 95
43       67.75063 48.29647 87.20479 37.99806  97.50319
44       67.75063 47.55748 87.94377 36.86788  98.63338
45       67.75063 46.84460 88.65665 35.77762  99.72363
46       67.75063 46.15524 89.34601 34.72333 100.77792
47       67.75063 45.48722 90.01404 33.70168 101.79958
```

```
plot(mod.forecast)
acf(mod.forecast$residuals,lag.max=15)
Box.test(mod.forecast$residuals,lag=20,type="Ljung-Box")# same results as above
```







```
mod1=Arima(x,order=c(3,1,0))
summary(mod1)
Series: x
ARIMA(3,1,0)

Coefficients:
          ar1      ar2      ar3
      -0.6063  -0.4904  -0.3284
s.e.   0.1489   0.1551   0.1477

sigma^2 estimated as 222:  log likelihood=-169.3
AIC=346.59   AICc=347.7   BIC=353.45

Training set error measures:
                    ME    RMSE      MAE     MPE    MAPE      MASE
Training set 0.4680342 14.7227 11.90333 -9.9101 29.0211 0.9033502
```

```
mod2=Arima(x,order=c(3,1,1))
summary(mod2)
Series: x
ARIMA(3,1,1)

Coefficients:
          ar1      ar2      ar3      ma1
      -0.5805  -0.4778  -0.3196  -0.0293
s.e.   0.4646   0.2669   0.2140   0.4985

sigma^2 estimated as 222:  log likelihood=-169.29
AIC=348.59   AICc=350.3   BIC=357.16

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE
Training set 0.4792116 14.72198 11.90275 -9.899733 29.00887 0.7484857
                    ACF1
Training set -0.01216043
```