Homework 6

Brian Detweiler

March 11, 2017

1. Consider three seperate AR(1) models: $\phi = 0.1, \phi = 0.5$, and $\phi = 0.8$.

(a) For each model, calculate ρ_1 and ρ_7.

$$Y_t = 0.1Y_{t-1} - e_t$$

We must find $\rho_k = \frac{\gamma_k}{\gamma_0}$. First we need γ_0 in each case.

$$\begin{aligned}
\gamma_0 &= Cov(Y_t, Y_t) = Var(Y_t) \\
&= Var(0.1Y_{t-1} - e_t) \\
&= 0.01Var(Y_{t-1}) + Var(e_t) \\
&= 0.01\gamma_0 + \sigma_e^2 \\
\gamma_0 - 0.01\gamma_0 &= \sigma_e^2 \\
0.99\gamma_0 &= \sigma_e^2 \\
\gamma_0 &= \frac{\sigma_e^2}{0.99}
\end{aligned}$$

More generally, $\gamma_0 = \frac{\sigma_e^2}{1-\phi_1^2}$.

We will expand the Yule-Walker equations,

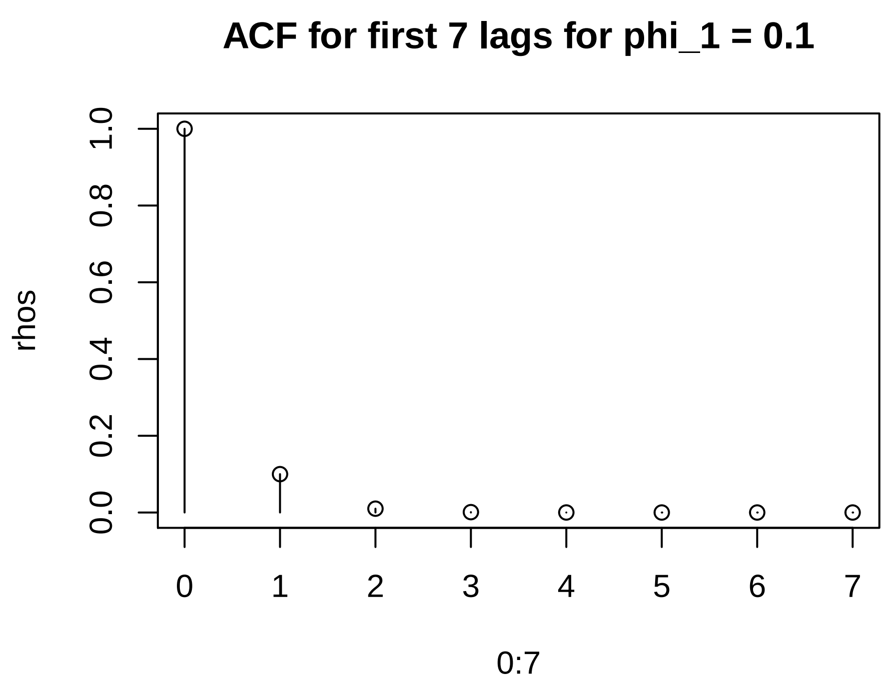
$$\begin{aligned}
\gamma_k &= \phi_1\gamma_{k-1} \\
\rho_k &= \phi_1\rho_{k-1}
\end{aligned}$$

So we now have

$$
\begin{aligned}
\gamma_1 &= \phi_1\gamma_0 \\
&= 0.1\gamma_0 \\
&= \frac{0.1\sigma_e^2}{0.99} \\
&= 0.1010101\sigma_e^2 \\
\rho_1 &= \frac{\gamma_1}{\gamma_0} \\
&= \frac{0.1010101\sigma_e^2}{\frac{\sigma_e^2}{0.99}} \\
&= 0.1010101(0.99) \\
&= 0.1 \\
&= \phi_1 \\
\rho_2 &= \phi_1\rho_1 \\
\rho_2 &= 0.1(0.1) \\
\rho_2 &= \phi_1^2
\end{aligned}
$$

Now we can see the pattern emerging. $\rho_k = \phi_1^k$, and hence $\rho_7 = \phi_1^7 = (0.1)^7$.

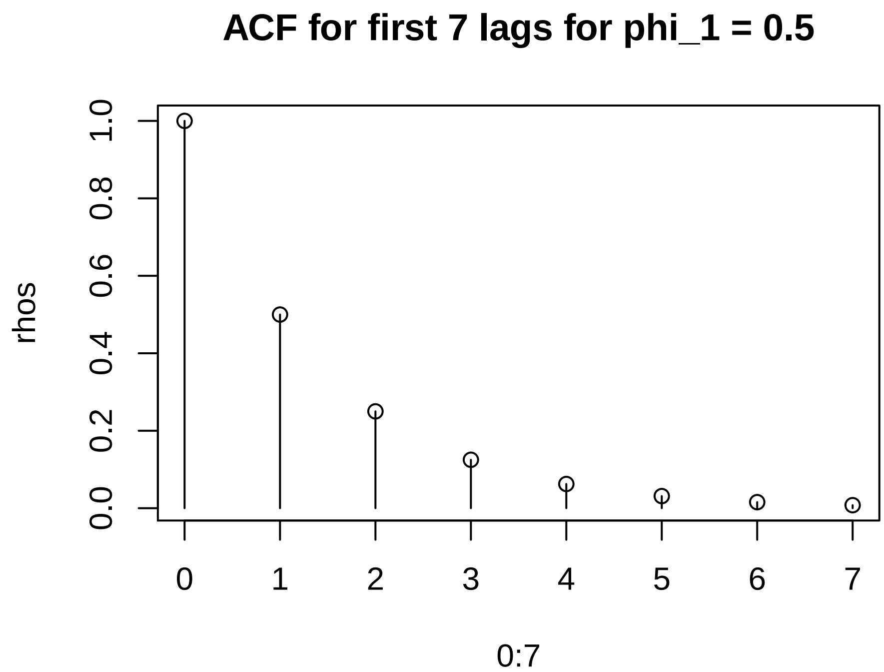
The autocorrelation function shows a decaying sequence.



Now, since we've generalized it, we can easily apply this to the other models.

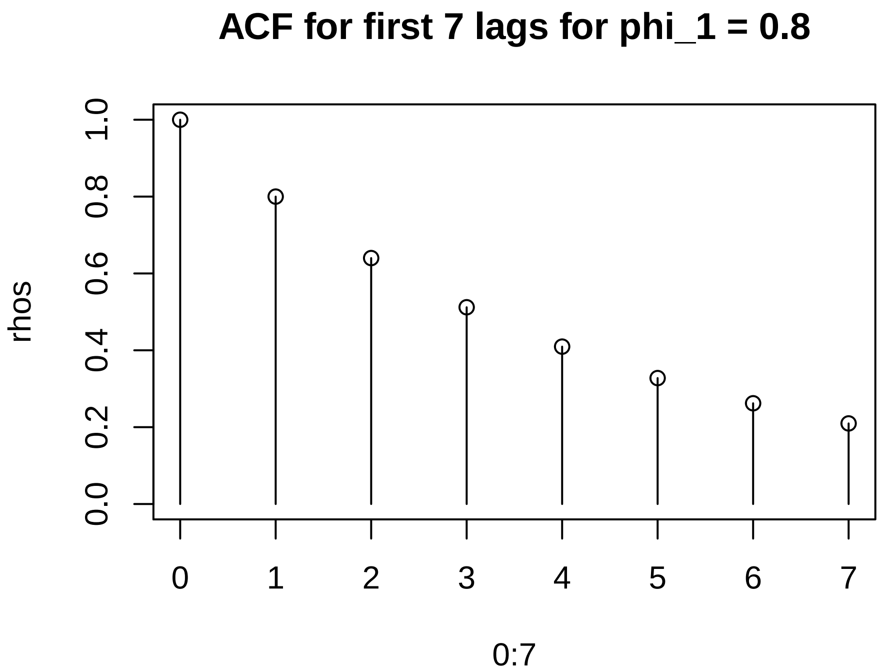
$$Y_t = 0.5Y_{t-1} - e_t$$

$$\begin{aligned}
\gamma_0 &= \frac{\sigma_e^2}{1-\phi_1^2} \\
&= \frac{\sigma_e^2}{1-0.5} \\
&= \frac{\sigma_e^2}{0.5} \\
\gamma_1 &= \phi_1\gamma_0 \\
&= 0.5\gamma_0 \\
&= \frac{0.5\sigma_e^2}{0.5} \\
&= \sigma_e^2 \\
\rho_1 &= \frac{\gamma_1}{\gamma_0} \\
&= \frac{\sigma_e^2}{\frac{\sigma_e^2}{0.5}} \\
&= 0.5 \\
&= \phi_1 \\
\rho_2 &= \phi_1\rho_1 \\
\rho_2 &= 0.5(0.5) \\
\rho_2 &= \phi_1^2 \\
\rho_7 &= \phi_1^7 = 0.5^7
\end{aligned}$$


ACF for first 7 lags for phi_1 = 0.5
rhos
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
0:7

$$Y_t = 0.8Y_{t-1} - e_t$$

$$\rho_1 = \phi_1 = 0.8$$
$$\rho_7 = \phi_1^7 = 0.2097152$$



(b) For each model, calculate $Var(r_1)$ and $Var(r_7)$.

We define the sample ACF as

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

For an AR(1) process with $\rho_k = \phi^k$ for $k > 0$,

$$Var(r_k) \approx \frac{1}{n}\left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}\right]$$

We'll create a helper function to calculate the numerical parts of the variance in R:

```r
partial.var.rk <- function(phi, k) {
  return(((1 + phi^2) * (1 - phi^(2 * k)) / (1 - phi^2) - 2 * k * phi^(2 * k)))
}
```

So for the given models and $k = 1, 7$, we have

Model 1, $\phi_1 = 0.1$

$$\begin{aligned}
Var(r_1) &\approx \frac{1}{n}\left[\frac{(1+(0.1)^2)(1-(0.1)^2)}{1-(0.1)^2} - 2(0.1)^2\right] \\
&\approx \frac{0.99}{n} \\
Var(r_7) &\approx \frac{1}{n}\left[\frac{(1+(0.1)^2)(1-(0.1)^{14})}{1-(0.1)^2} - 14(0.1)^{14}\right] \\
&\approx \frac{1.020202}{n}
\end{aligned}$$

Model 2, $\phi_1 = 0.5$

$$\begin{aligned}
Var(r_1) &\approx \frac{1}{n}\left[\frac{(1+(0.5)^2)(1-(0.5)^2)}{1-(0.5)^2} - 2(0.5)^2\right] \\
&\approx \frac{0.75}{n} \\
Var(r_7) &\approx \frac{1}{n}\left[\frac{(1+(0.5)^2)(1-(0.5)^{14})}{1-(0.5)^2} - 14(0.5)^{14}\right] \\
&\approx \frac{1.6657104}{n}
\end{aligned}$$

Model 3, $\phi_1 = 0.8$

$$
\begin{aligned}
Var(r_1) &\approx \frac{1}{n}\left[\frac{(1+(0.8)^2)(1-(0.8)^2)}{1-(0.8)^2} - 2(0.8)^2\right] \\
&\approx \frac{0.36}{n} \\
Var(r_7) &\approx \frac{1}{n}\left[\frac{(1+(0.8)^2)(1-(0.8)^{14})}{1-(0.8)^2} - 14(0.8)^{14}\right] \\
&\approx \frac{3.7394736}{n}
\end{aligned}
$$

(c) For each model, use the `arima.sim` function to simulate a time series of length $n = 60$. Then use the `acf` function to calculate r_1 and r_7. Remember to set up a random seed for your simulation.

```
set.seed(0)

phi1 <- 0.1

r1 <- c()
r7 <- c()
for (i in 1:1000) {
  sim <- arima.sim(n = 60, model = list(ar=(phi1)))
  r <- acf(sim, plot = FALSE)
  r1 <- c(r1, r[[1]][1])
  r7 <- c(r7, r[[1]][7])
}

sd(r1)
```

```
## [1] 0.1274207
```

```
sd(r7)
```

```
## [1] 0.1169221
```

```
phi1 <- 0.5

r1 <- c()
r7 <- c()
for (i in 1:1000) {
  sim <- arima.sim(n = 60, model = list(ar=(phi1)))
  r <- acf(sim, plot = FALSE)
  r1 <- c(r1, r[[1]][1])
  r7 <- c(r7, r[[1]][7])
}

sd(r1)
```

```
## [1] 0.1160971
```

```
sd(r7)
```

```
## [1] 0.1350831
```

```r
phi1 <- 0.8

r1 <- c()
r7 <- c()
for (i in 1:1000) {
  sim <- arima.sim(n = 60, model = list(ar=(phi1)))
  r <- acf(sim, plot = FALSE)
  r1 <- c(r1, r[[1]][1])
  r7 <- c(r7, r[[1]][7])
}

sd(r1)
```

```
## [1] 0.09641568
```

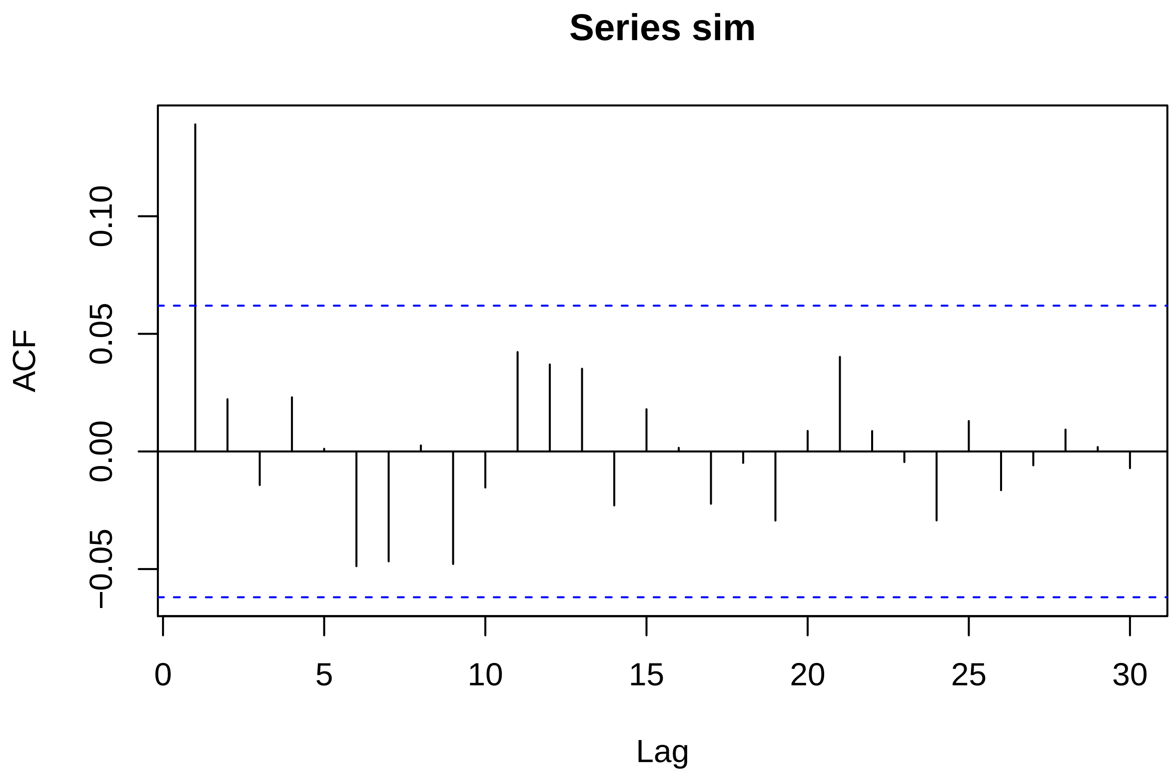
```r
sd(r7)
```

```
## [1] 0.1894419
```

(d) Based on your results in parts (a) and (b), are r_1 and r_7 from part (c) within 2 standard deviations of ρ_1 and ρ_7 respectively?

(e) Repeat part (c) for 1000 times. Draw histograms for r_1's and r_7's for each model. What proportion of r_1's and r_7's are within 2 standard deviations of ρ_1 and ρ_7?

```r
phi1 <- 0.1

sim <- arima.sim(n = 1000, model = list(ar=(phi1)))

r <- acf(sim)
```



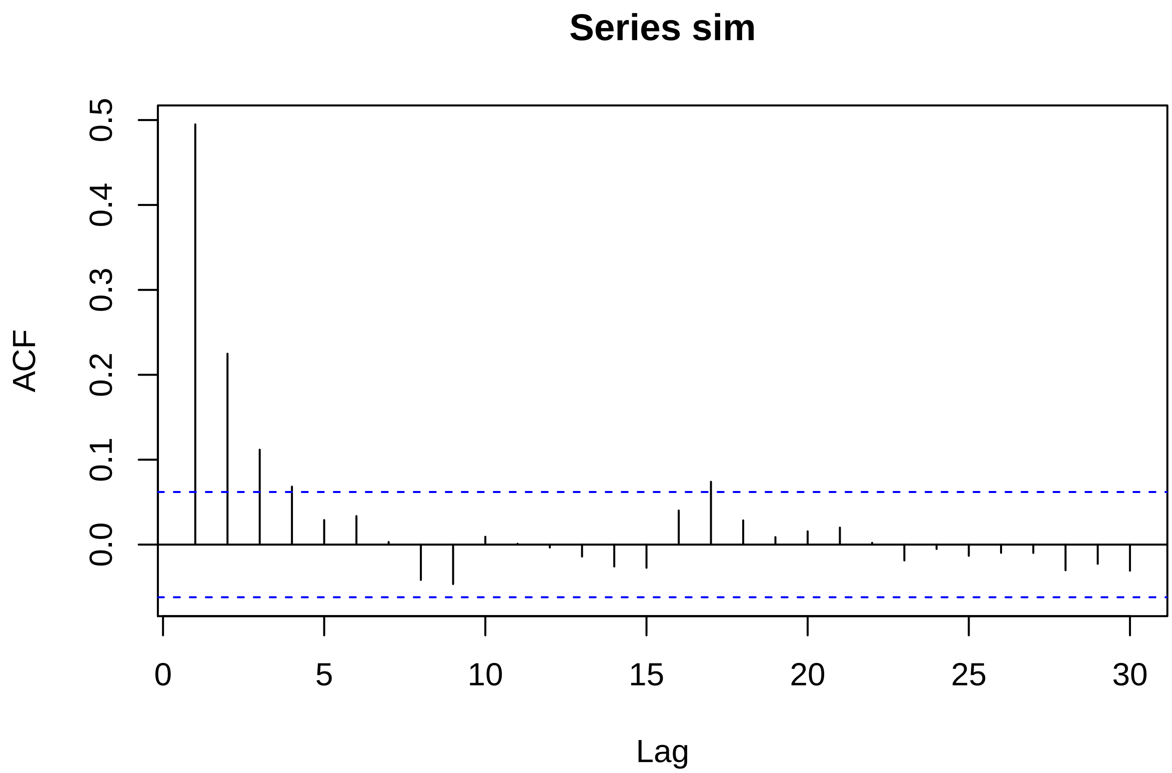

```
r[1]
```

```
##
## Autocorrelations of series 'sim', by lag
##
##     1
## 0.139
```

```
r[7]
```

```
##
## Autocorrelations of series 'sim', by lag
##
##      7
## -0.047
```

```
phi1 <- 0.5

sim <- arima.sim(n = 1000, model = list(ar=(phi1)))

r <- acf(sim)
```



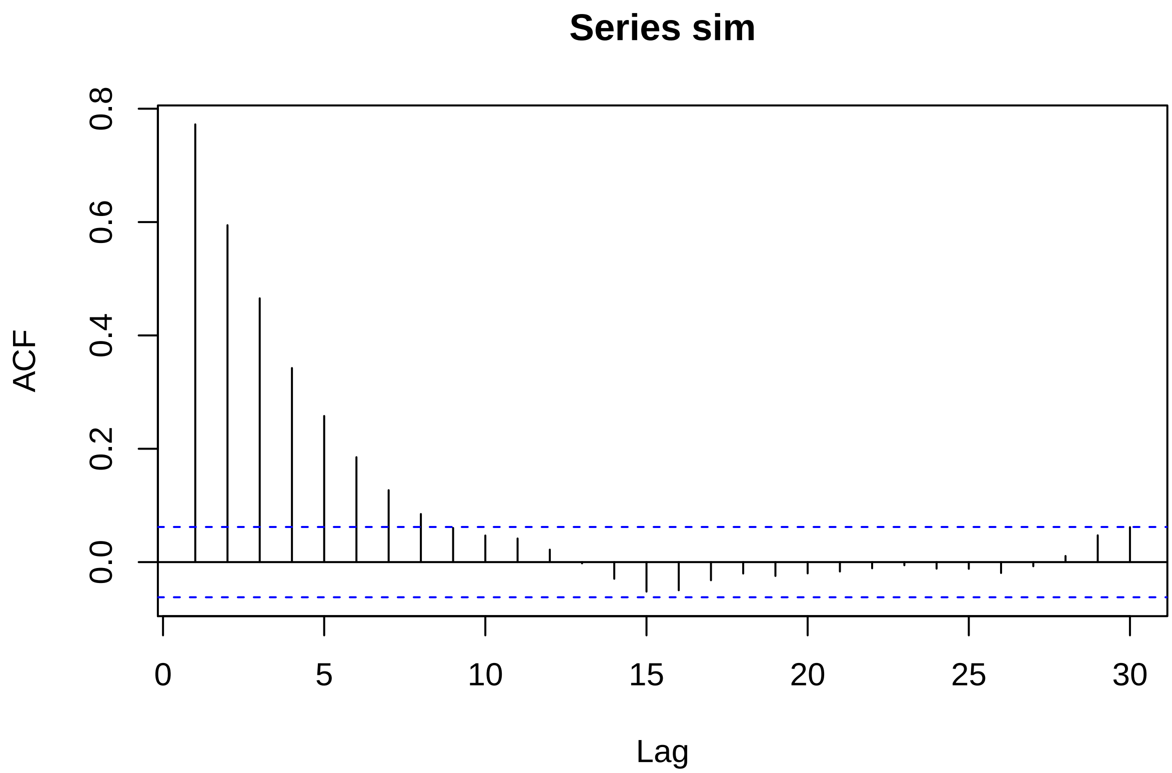

```r
r[1]
```

```
##
## Autocorrelations of series 'sim', by lag
##
##     1
## 0.495
```

```r
r[7]
```

```
##
## Autocorrelations of series 'sim', by lag
##
##     7
## 0.003
```

```r
phi1 <- 0.8

sim <- arima.sim(n = 1000, model = list(ar=(phi1)))

r <- acf(sim)
```




```
r[1]
```

```
##
## Autocorrelations of series 'sim', by lag
##
##     1
## 0.772
```

```
r[7]
```

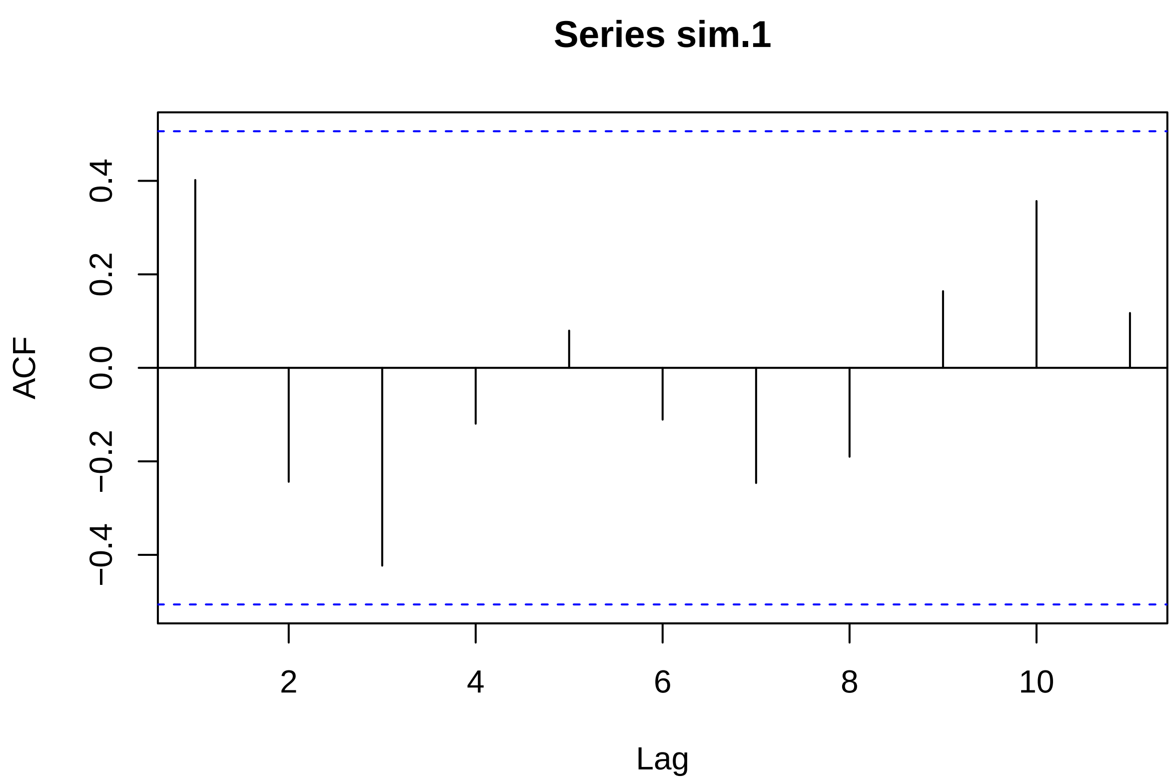
```
##
## Autocorrelations of series 'sim', by lag
##
##     7
## 0.127
```

■

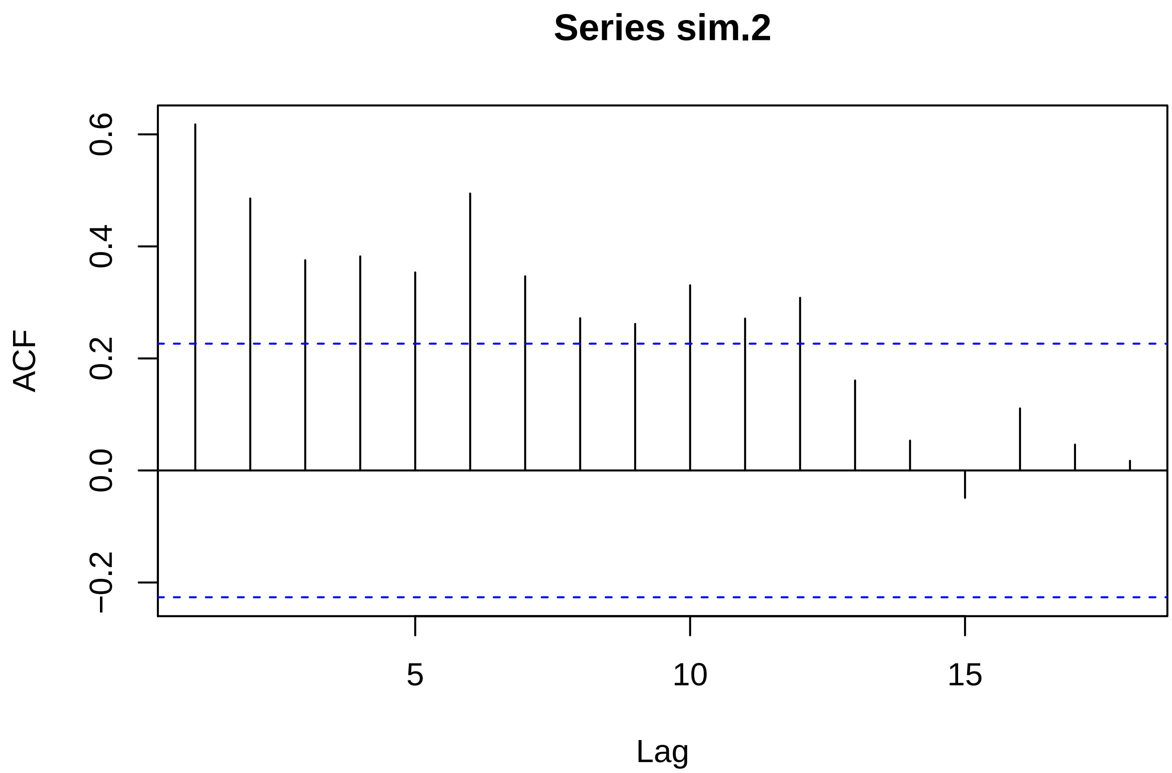
2. Consider an $AR(1)$ model with $\phi = 0.6$.

(a) Use the `arima.sim` function to simulate three time series of lengths $n = 15$, 75, and 100.

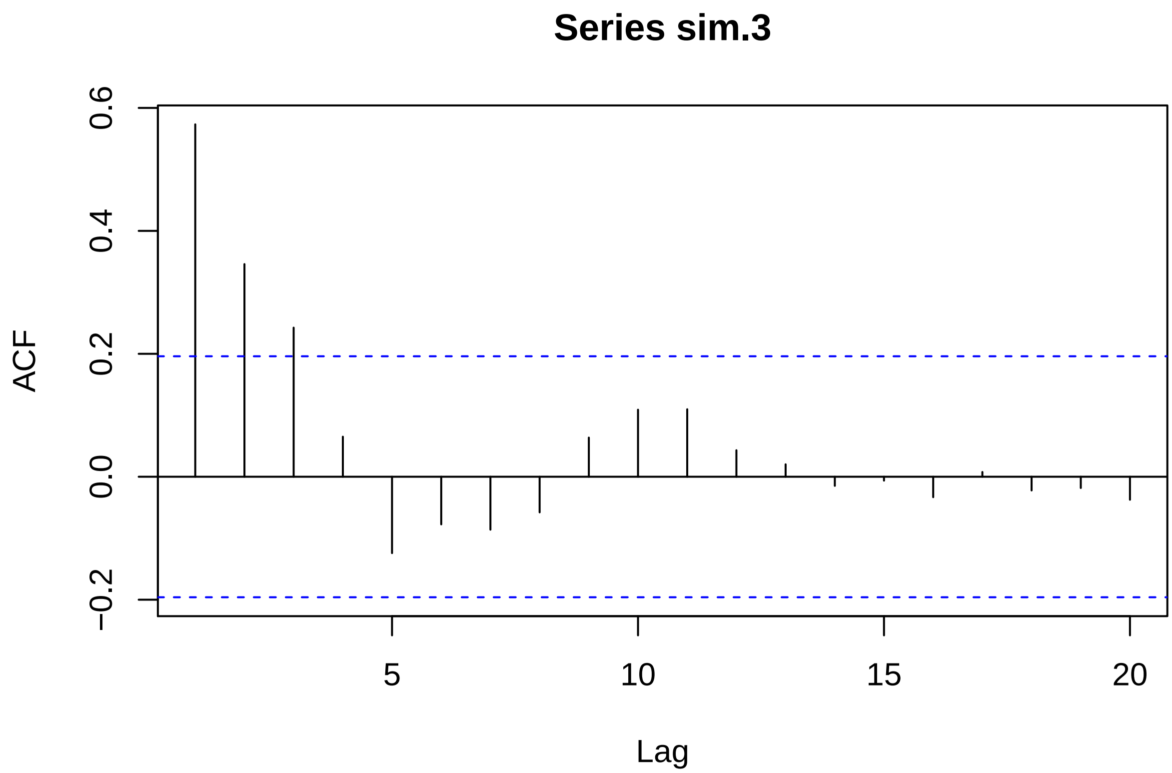
```
phi1 <- 0.6

sim.1 <- arima.sim(n = 15, model = list(ar=(phi1)))
sim.2 <- arima.sim(n = 75, model = list(ar=(phi1)))
sim.3 <- arima.sim(n = 100, model = list(ar=(phi1)))

r.1 <- acf(sim.1)
```




```
r.2 <- acf(sim.2)
```




```
r.3 <- acf(sim.3)
```



(b) For each set of simulated data, calculate r_1.

```
r.1[1]
```

```
##
## Autocorrelations of series 'sim.1', by lag
##
##     1
## 0.402
```

```
r.1[7]
```

```
##
## Autocorrelations of series 'sim.1', by lag
##
##      7
## -0.246
```

```
r.2[1]
```

```
##
## Autocorrelations of series 'sim.2', by lag
##
##     1
## 0.618
```

```r
r.2[7]
```

```
##
## Autocorrelations of series 'sim.2', by lag
##
##     7
## 0.347
```

```r
r.3[1]
```

```
##
## Autocorrelations of series 'sim.3', by lag
##
##     1
## 0.573
```

```r
r.3[7]
```

```
##
## Autocorrelations of series 'sim.3', by lag
##
##      7
## -0.086
```

(c) For each n, what is $Var(r_1)$? Is r_1 within 2 standard deviations of ρ_1 for each sample?

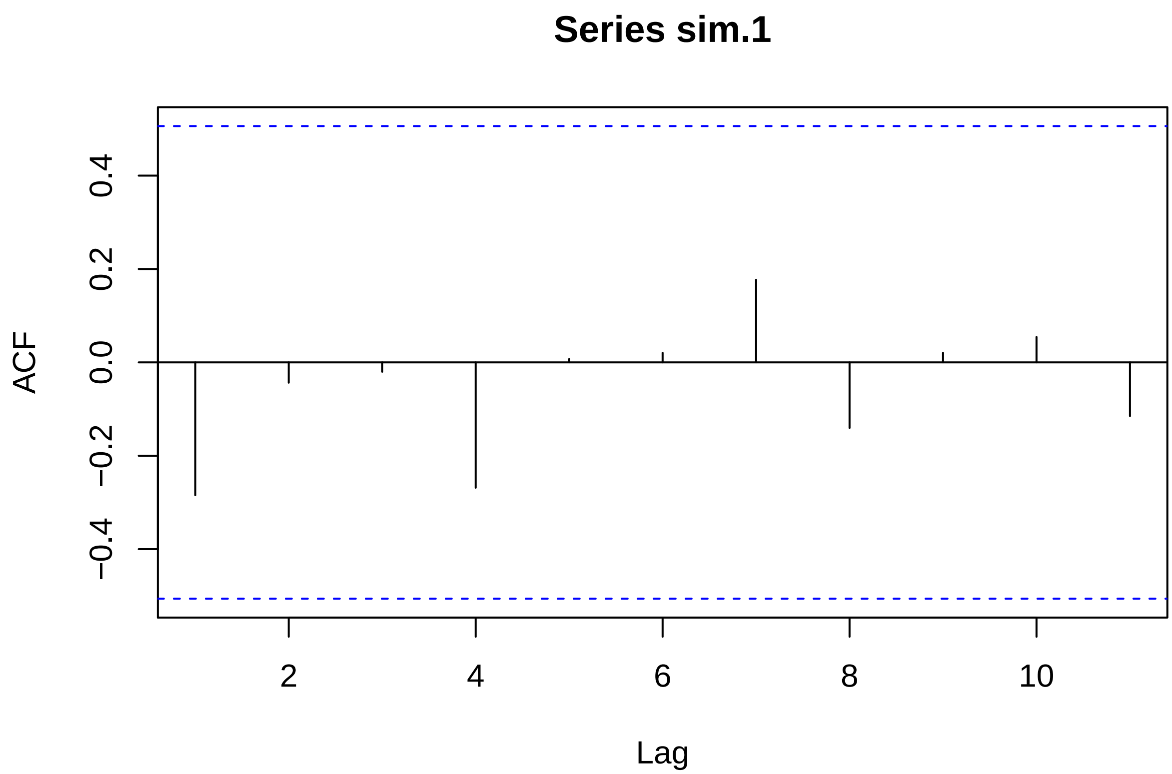
(d) Repeat part (a) for 1000 times. For each n, draw a histogram of the 1000 r_1's, and find what proportion of r_1's are within 2 standard deviations of ρ_1.

■

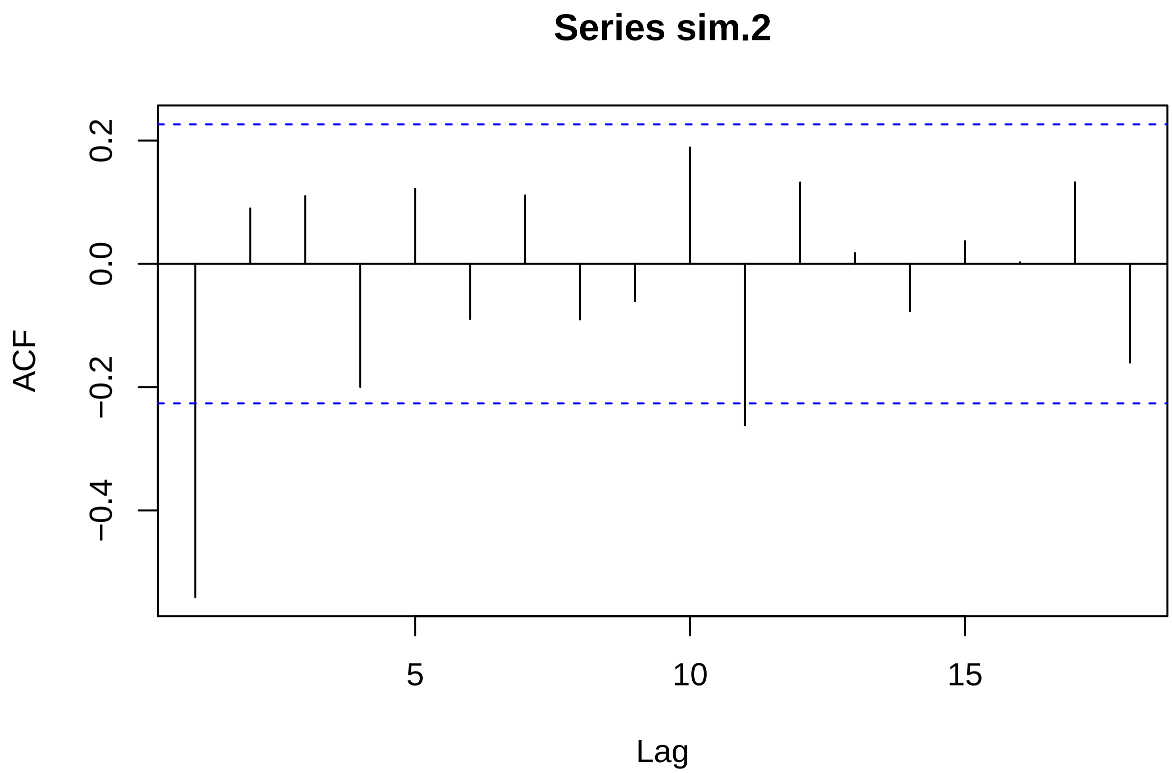
3. Consider an MA(1) model with $\theta = 0.6$

(a) Use the `arima.sim` function to simulate three time series of lengths n = 15, 75, and 150. Note that R uses the negative of the MA coefficients.

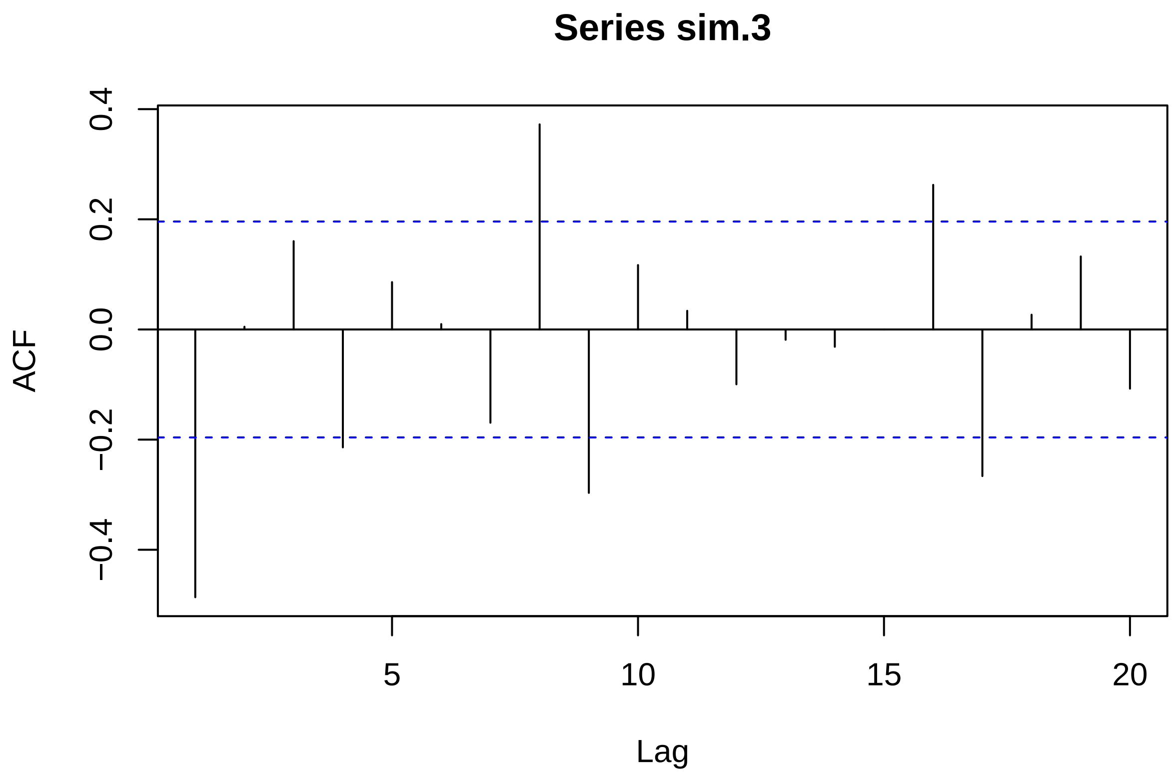
```
theta1 <- -0.6

sim.1 <- arima.sim(n = 15, model = list(ma=(theta1)))
sim.2 <- arima.sim(n = 75, model = list(ma=(theta1)))
sim.3 <- arima.sim(n = 100, model = list(ma=(theta1)))

r.1 <- acf(sim.1)
```




```
r.2 <- acf(sim.2)
```




```
r.3 <- acf(sim.3)
```



(b) For each set of simulated data, calculate r_1.

(c) For each n, what is $Var(r_1)$? Is r_1 within 2 standard deviations of ρ_1 for each sample?

(d) Repeat part (a) for 1000 times. For each n, draw a histogram of the 1000 r_1's, and find what proportion of r_1's are within 2 standard deviations of ρ_1.

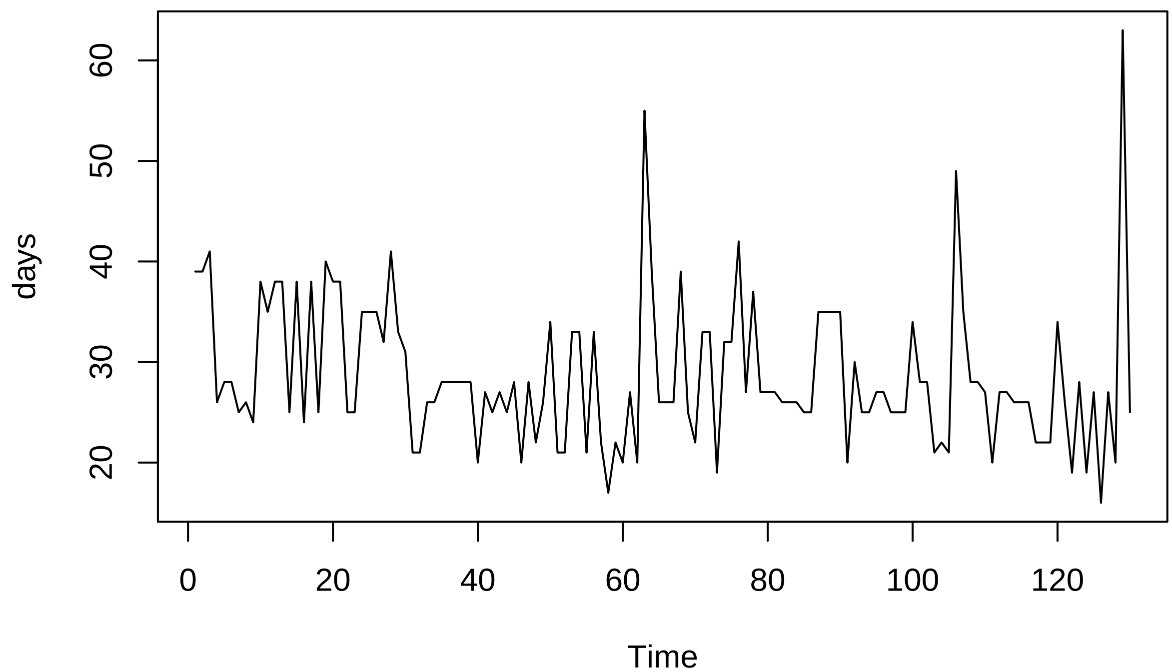
■

4. The dataset `days` contains accounting data. The data is the number of days it took to receive payment for 130 consecutive orders from a particular distributor.

```r
data(days)
```

(a) Plot the times series. Are there any unusual values?

```r
plot(days)
```

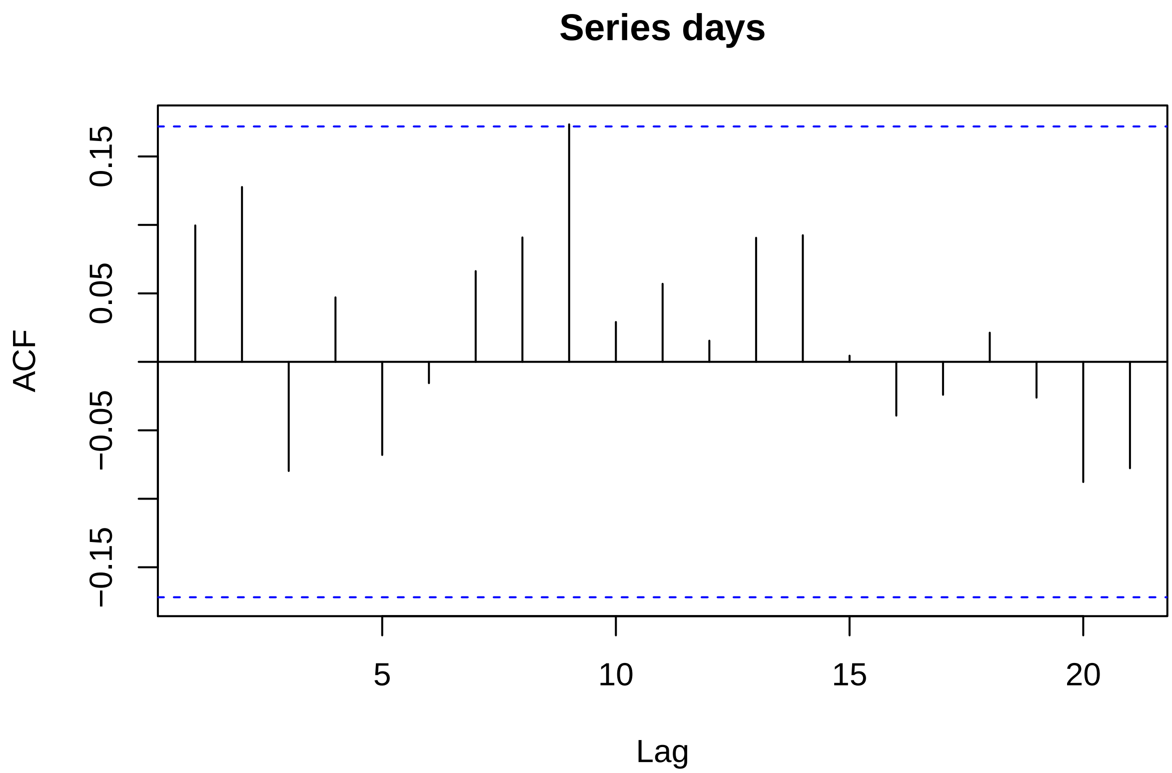



```r
day1 <- which(days == 55)
day2 <- which(days == 49)
day3 <- which(days == 63)
```

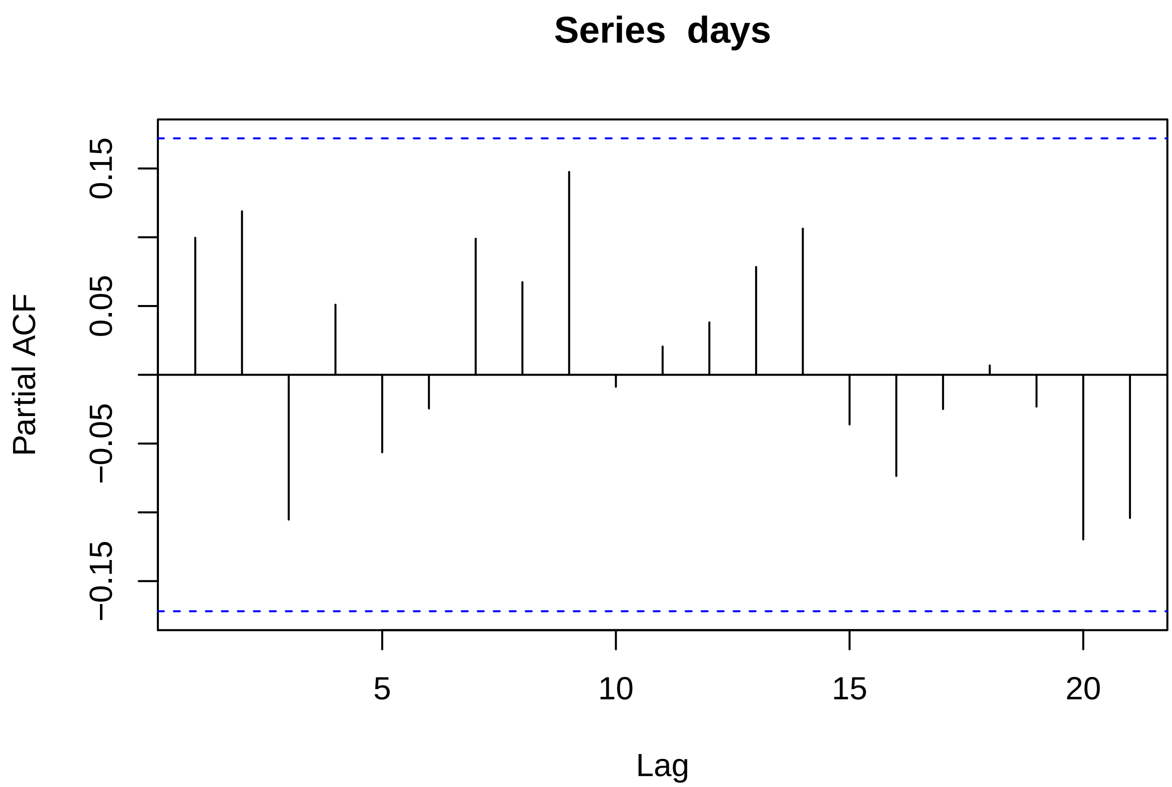
There are three highly unusual days, at days 63, 106, and 129.

(b) Draw the sample ACF and sample PACF plots. What do you find?

```r
acf(days)
```

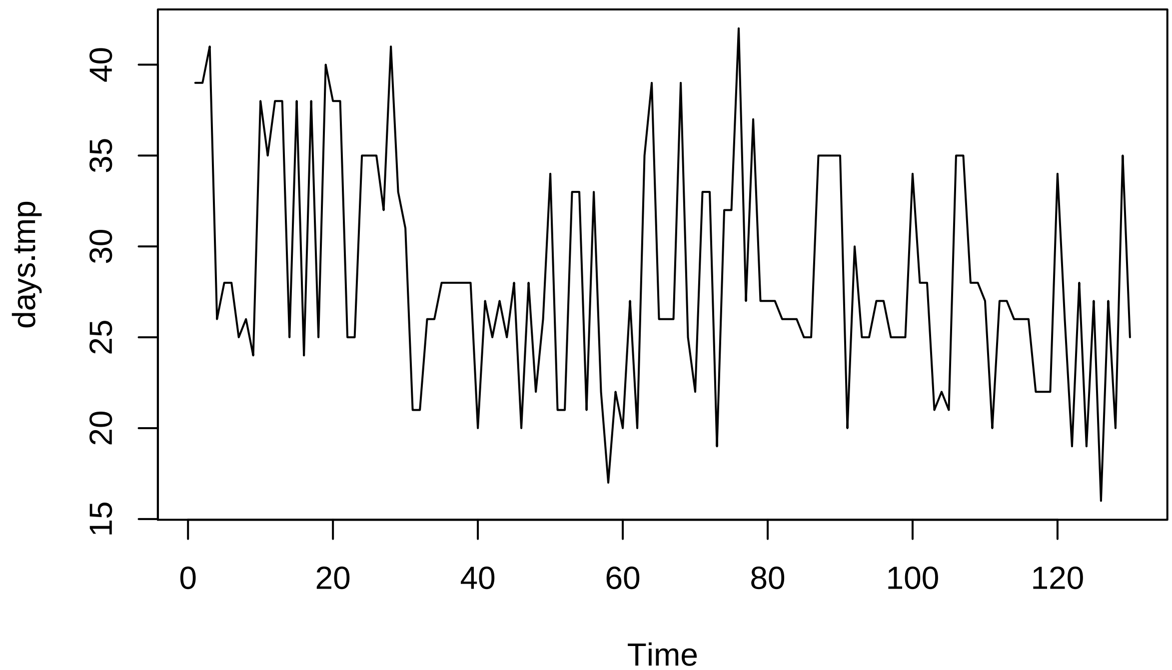



```r
pacf(days)
```

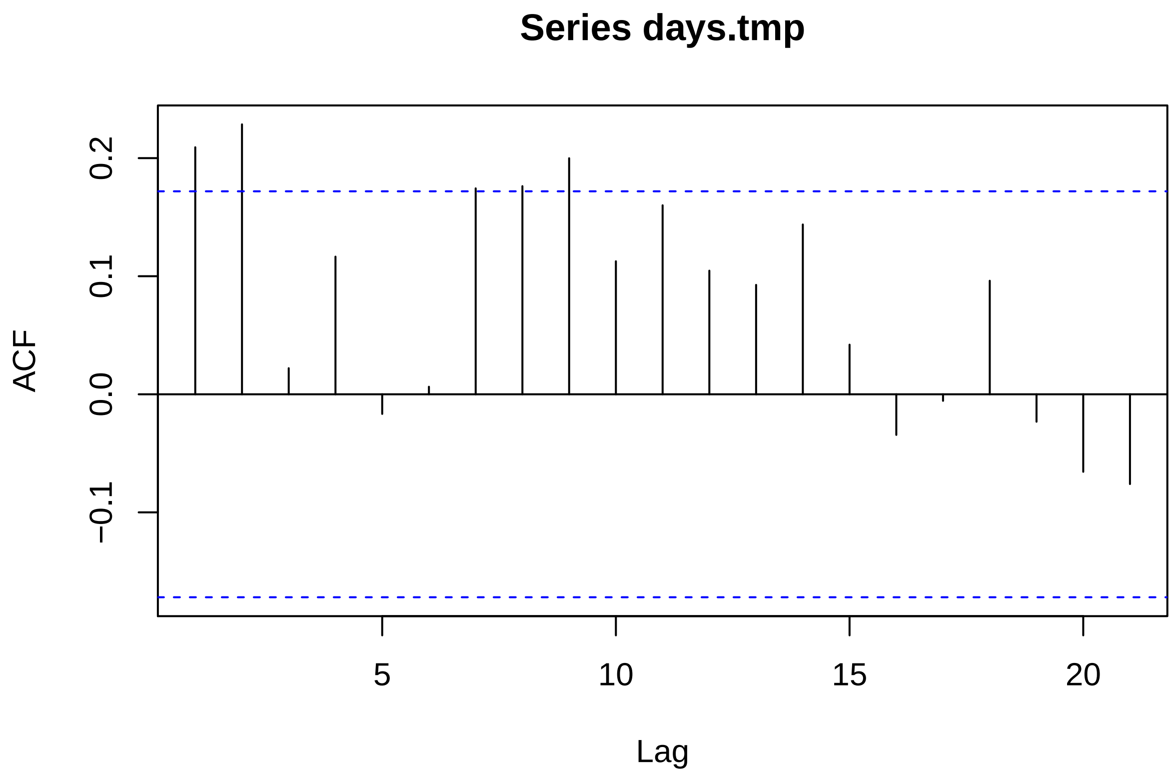


(c) Replace the unusual values with a value of 35 days. Redraw the sample ACF and sample PACF plots. Are they different from part (b)?

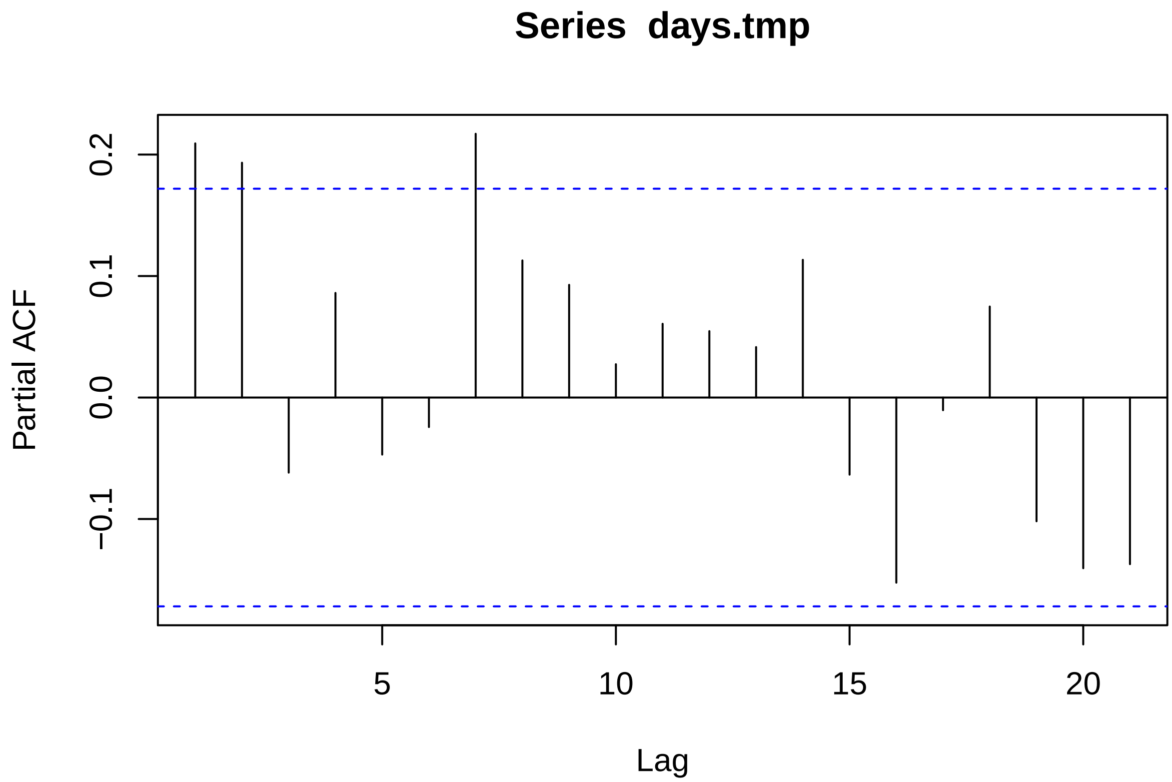
```
days.tmp <- days
days.tmp[which(days.tmp == 55)] <- 35
days.tmp[which(days.tmp == 49)] <- 35
days.tmp[which(days.tmp == 63)] <- 35
plot(days.tmp)
```




```
acf(days.tmp)
```




```
pacf(days.tmp)
```



(d) What ARMA model would you specify for this series after removing the outliers? Explain.

■