시계열 분석 및 응용

Assignment #4 (7, 10, 11)

서울대학교 통계학과 2017-11362 박건도

2022년 05월 09일

7. AIC

(1) $X_t = 0.5X_{t-1} + \epsilon_t, \epsilon_t \sim \ iid\,(0, \sigma^2)$

```
X1 <- arima.sim(list(order=c(1, 0, 0), ar=0.5), n=100)
fit.ar <- Arima(X1, order=c(1, 0, 0))
fit.arma <- Arima(X1, order=c(1, 0, 1))
fit.ar$aic
```

```
[1] 261.6923
```

```
fit.arma$aic
```

```
[1] 263.6684
```

두 모형에서 AIC 값의 차이가 크지 않으므로, ARMA(1, 1) 모형은 과적합 되었고 AR(1) 모형은 잘 적합되었다고 볼 수 있다.

(2) $X_t = 0.5\epsilon_{t-1} + \epsilon_t, \epsilon_t \sim \ iid\,(0, \sigma^2)$

```
X2 <- arima.sim(list(order=c(0, 0, 1), ma=0.5), n=100)
fit.ma <- Arima(X2, order=c(0, 0, 1))
fit.arma <- Arima(X2, order=c(1, 0, 1))
fit.ma$aic
```

```
[1] 280.1969
```

```
fit.arma$aic
```

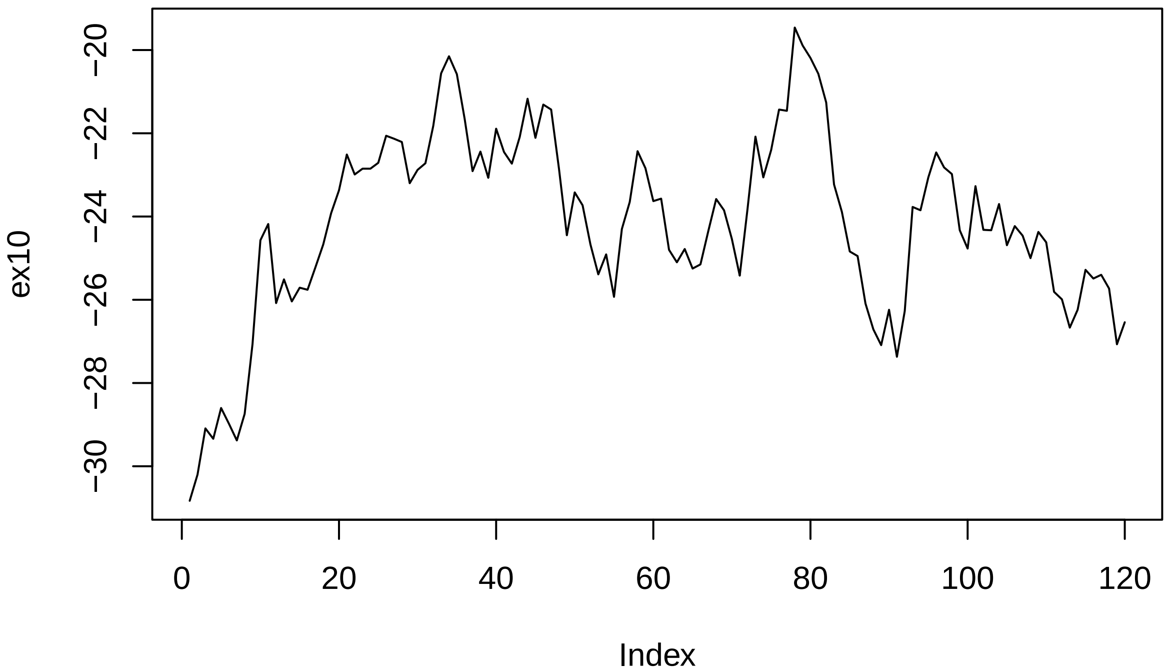
```
[1] 281.183
```

두 모형에서의 AIC 값이 크지 않으므로, ARMA(1, 1) 모형은 과적합 되었고 MA(1) 모형은 잘 적합되었다고 볼 수 있다.

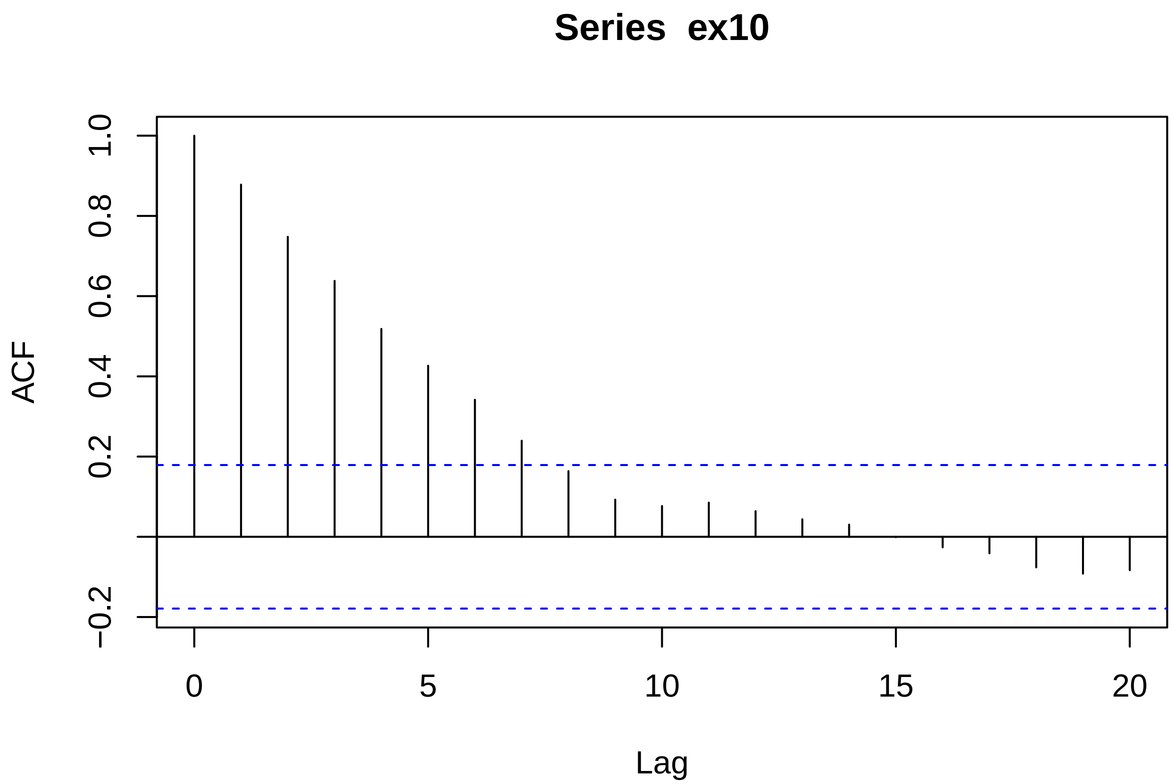
10. ex_ch4_10.txt

(1) plot

```
ex10 <- read.table("ex_ch4_10.txt", header=T)$data
plot(ex10, type="l")
```




```
acf(ex10)
```



시계열도를 그려봤을 때, 별다른 규칙이 없으므로 정상시계열이라 할 수 있다. 또한 SACF를 그려봤을 때, 점점 감소하는 것으로 보아 AR 모형이라고 생각할 수 있다.

(2) unit root test

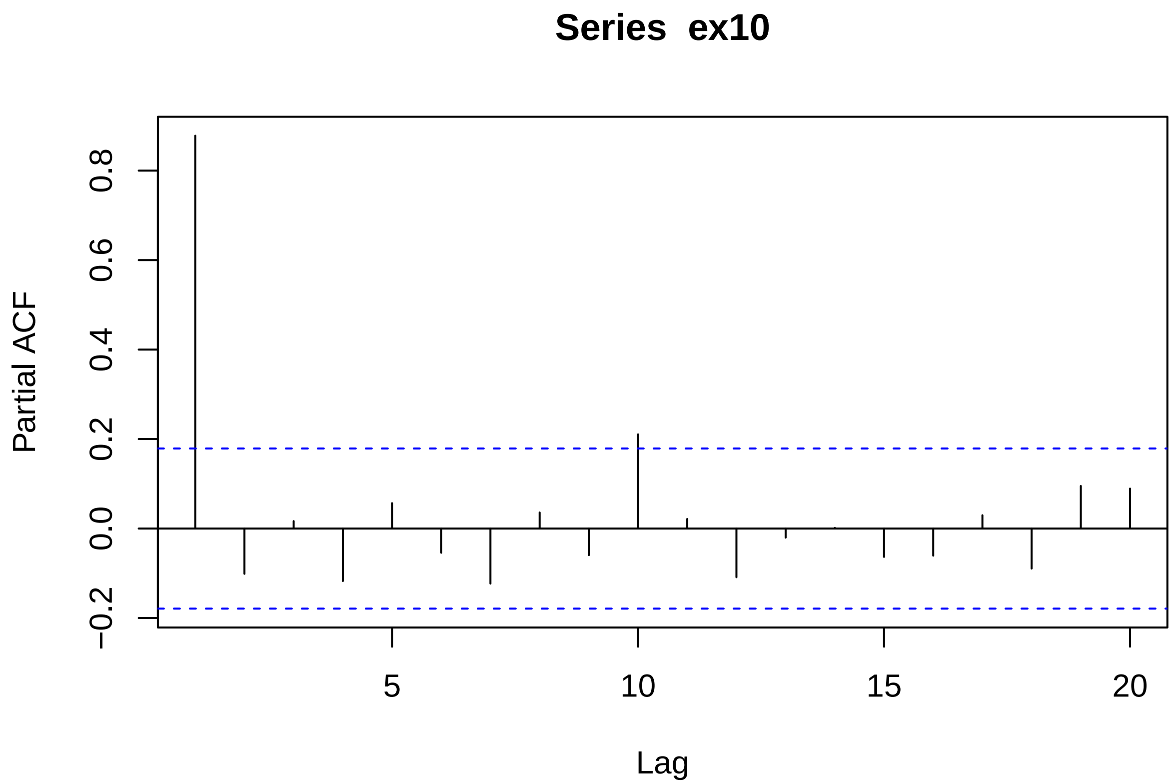
```
adf.test(ex10)
```

```
    Augmented Dickey-Fuller Test

data:  ex10
Dickey-Fuller = -3.0271, Lag order = 4, p-value = 0.15
alternative hypothesis: stationary
```

p-value가 0.15이므로 귀무가설을 기각할 수 없고, 비정상시계열이 아닌 정상시계열이라 할 수 있다.

(3) adequate model

```
pacf(ex10)
```



(1) 의 SACF와 위의 SPACF를 보면, 점점 감소하는 SACF와 2 이상에서 작은 값을 가지는 SPACF를 관찰할 수 있다. 따라서 AR(1) 모형을 따를 것으로 추측할 수 있다.

(4) fit

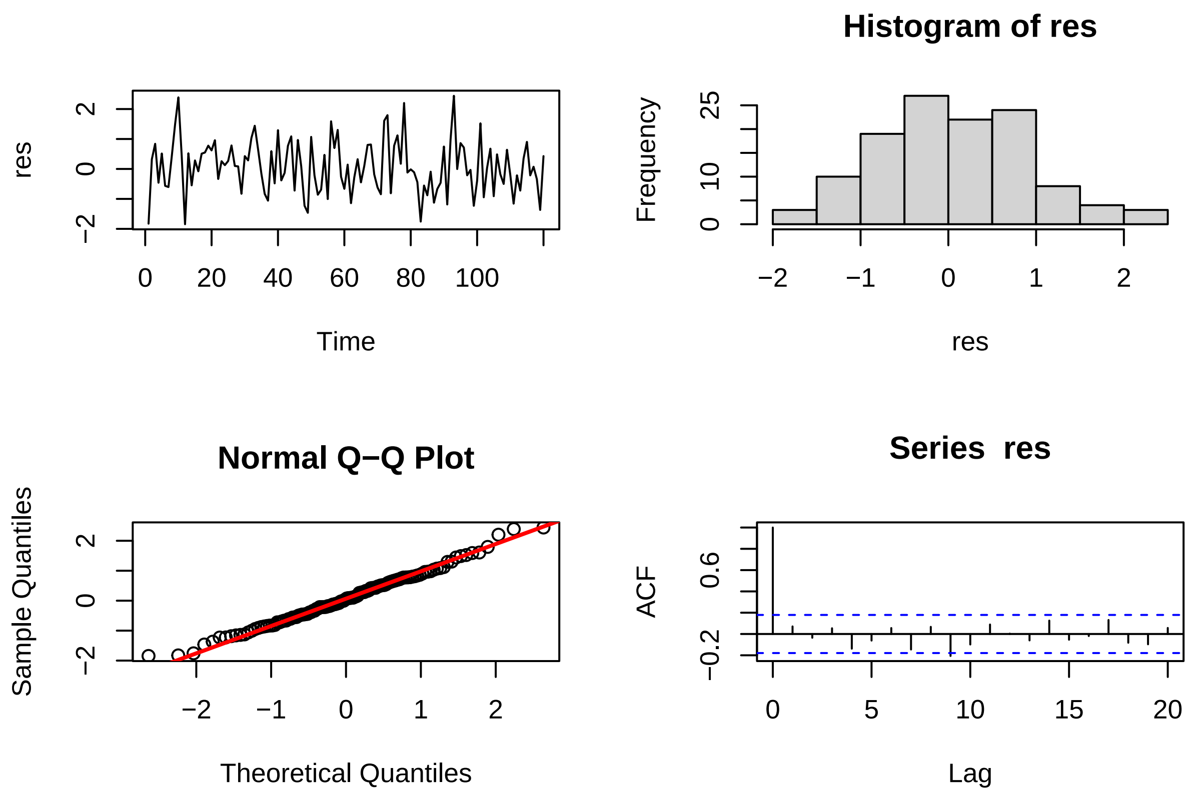
```
fit.ar <- Arima(ex10, order=c(1, 0, 0))
fit.ar
```

```
Series: ex10
ARIMA(1,0,0) with non-zero mean

Coefficients:
        ar1      mean
      0.946  -25.1935
s.e.  0.033    1.3969

sigma^2 = 0.7768:  log likelihood = -155.24
AIC=316.48   AICc=316.69   BIC=324.84
```

AR(1) 모형으로 적합한 결과, $\phi = 0.946,\ \mu = -25.1935$으로 적합되었다.

```
res <- fit.ar$residuals
par(mfrow=c(2,2))
```

```r
plot(res, type="l")
hist(res)
qqnorm(res)
qqline(res, col="red", lwd=2)
acf(res)
```



적합 시킨 결과에서 잔차를 분석해 보면, 정규성과 독립성을 만족함을 확인할 수 있다. 또한 잔차에 대해 SACF를 그려보면, 1 이상에서 작은 값을 가지므로 확실히 독립성을 만족한다고 할 수 있다.

```r
Box.test(res, type="Ljung-Box")
```

```
    Box-Ljung test

data:  res
X-squared = 0.62207, df = 1, p-value = 0.4303
```

```r
durbinWatsonTest(as.vector(res))
```

```
[1] 1.802802
```

Ljung-Box test와 Durbin-Watson test를 진행하면, 모형이 적절하고 오차항에 상관관계가 없다는 결론을 내릴 수 있다.

```r
jarque.bera.test(res)
```

```
    Jarque Bera Test

data:  res
X-squared = 1.0901, df = 2, p-value = 0.5798
```

```r
shapiro.test(res)
```

```
    Shapiro-Wilk normality test

data:  res
W = 0.99141, p-value = 0.6649
```

Jarque-Bera test와 Shapiro-Wilk test를 통해서도 정규성을 만족함을 보일 수 있다.

```r
Arima(ex10, order=c(1, 0, 1))
```

```
Series: ex10
ARIMA(1,0,1) with non-zero mean

Coefficients:
         ar1     ma1      mean
      0.9269  0.1305  -24.9638
s.e.  0.0418  0.1005    1.1812

sigma^2 = 0.7727:  log likelihood = -154.38
AIC=316.77   AICc=317.12   BIC=327.92
```

```r
Arima(ex10, order=c(2, 0, 0))
```

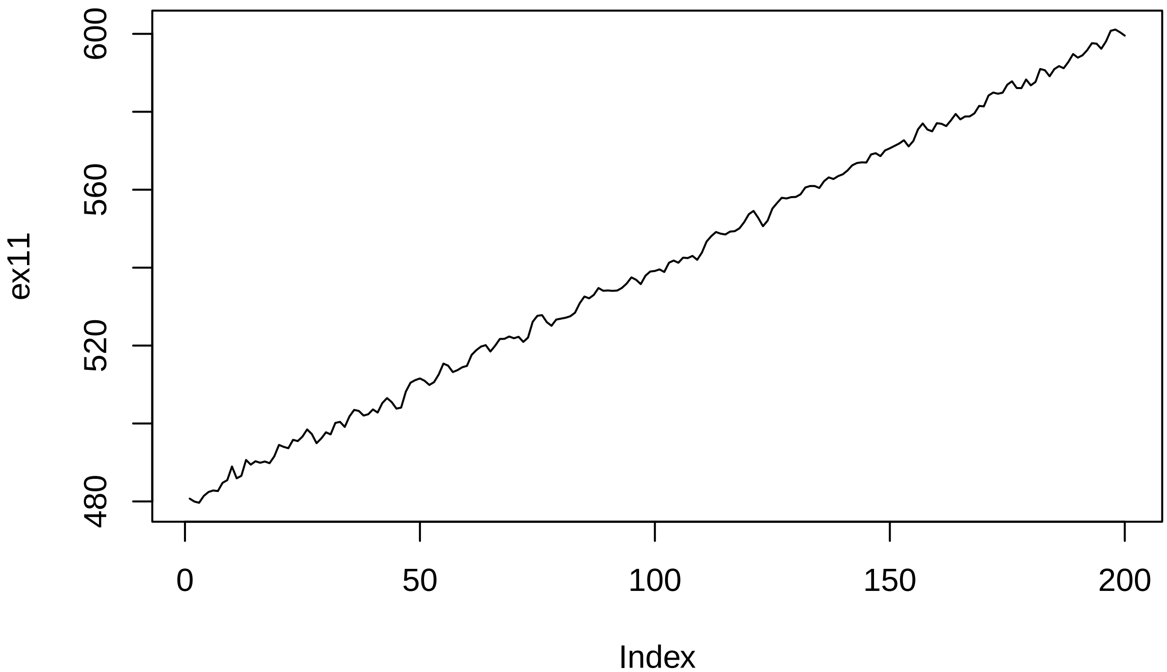
```
Series: ex10
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2      mean
      1.0575  -0.1238  -24.9383
s.e.  0.0905   0.0936    1.1592

sigma^2 = 0.7725:  log likelihood = -154.37
AIC=316.74   AICc=317.08   BIC=327.89
```

과적합 판단을 위해, AR(2) 와 ARMA(1, 1) 모형에 대해서 적합을 시도해 보았다. 그 결과 계수들이 크게 변하지 않았고 새로 추가한 계수의 값이 0에 가까웠다. 또한 AIC 값 또한 비슷하므로 AR(1) 모형이 적절한 모형임을 알 수 있다.

11. ex_ch4_11.txt

(1) plot

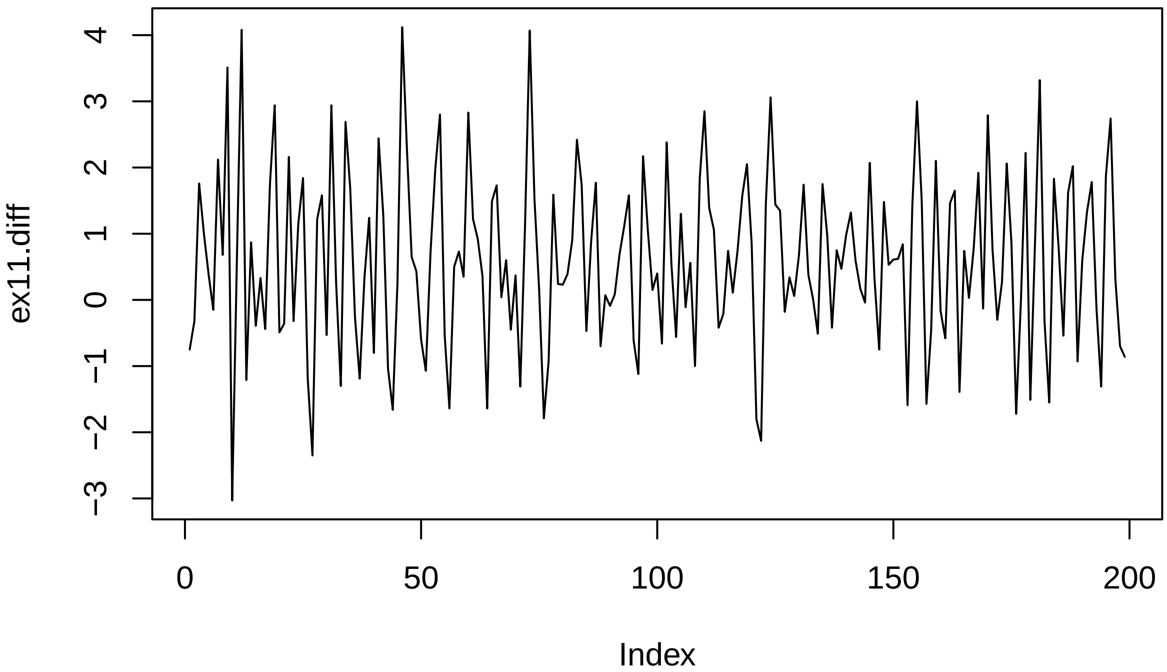
```
ex11 <- read.table("ex_ch4_11.txt", header=T)$data
plot(ex11, type="l")
```



시계열도를 그려본 결과, 증가하는 추세가 뚜렷하게 관찰되므로 정상시계열이 아니라고 생각할 수 있다.

(2) Diff

```
ex11.diff <- diff(ex11)
plot(ex11.diff, type="l")
```

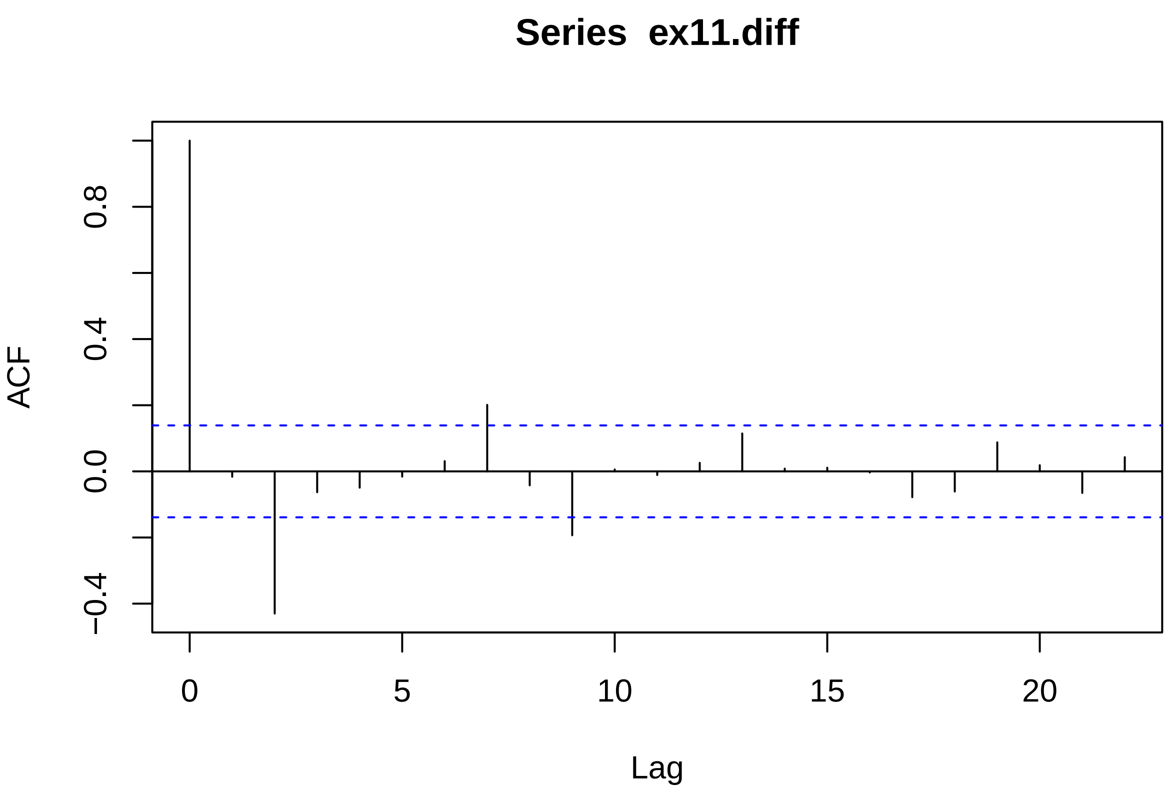


1 차

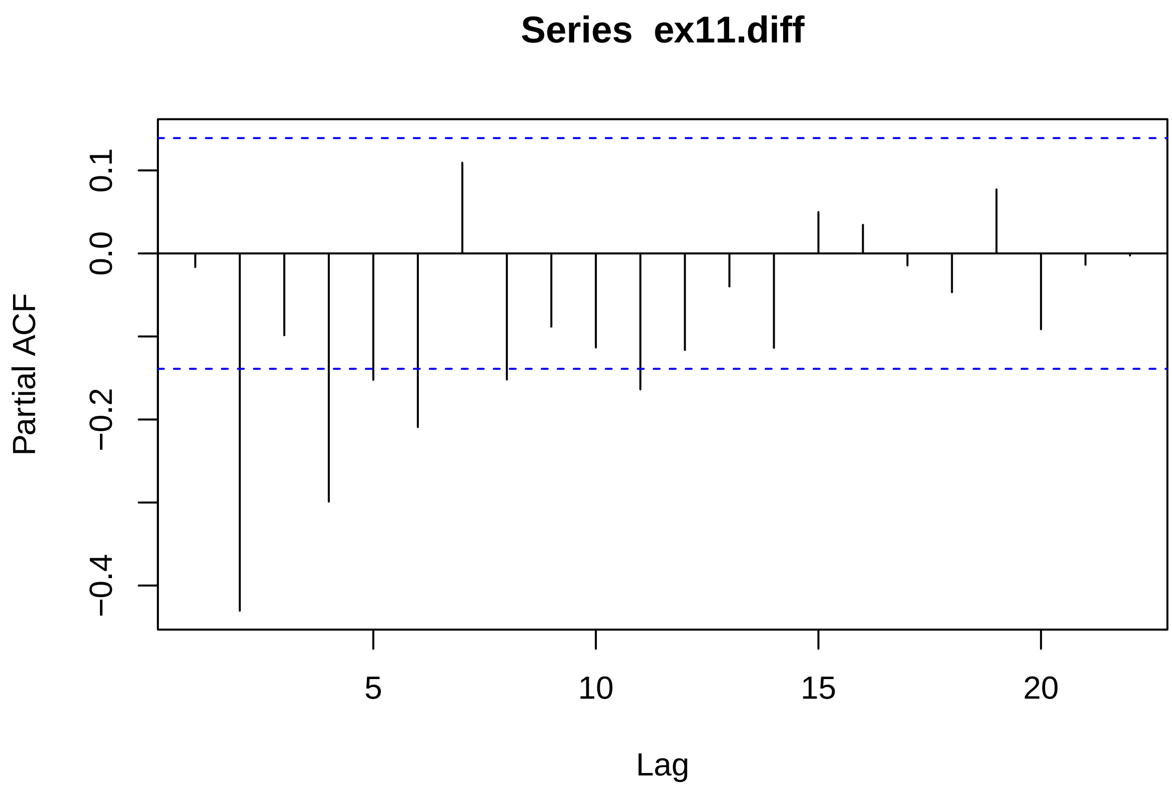
차분 후 시계열도를 그려본 결과, 정상성을 만족한다고 할 수 있다.

(3) SACF, SPACF

```
acf(ex11.diff)
```



```r
pacf(ex11.diff)
```



SACF와 SPACF 모두 시차가 1 이상에서 0에 가까운 값을 나타내고 있다. 다시말해, 시차 $(p-q)=0$ 이후부터 소멸하는 현상을 보이므로 ARIMA(p, 1, p) 모형이라고 생각할 수 있다.

(4) fit

```r
for (p in 1:4){
  print(Arima(ex11, order=c(p, 1, p), include.drift = T)$aic)
}
```

```
[1] 627.4889
[1] 596.7906
[1] 599.2809
[1] 601.7988
```

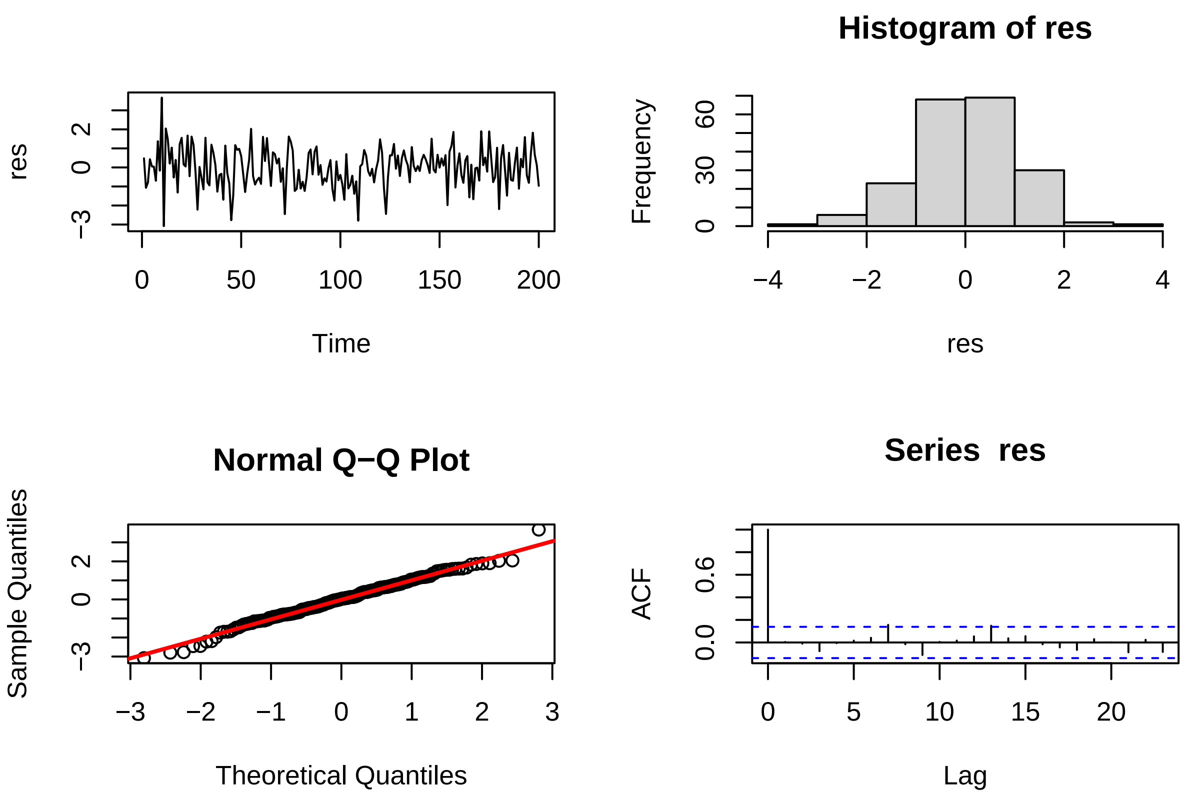
$p = 1, 2, 3, 4$에 대해 fitting 후 AIC 값을 비교해 본 결과, $p = 2$에서 가장 적게 나왔다. 따라서 ARIMA(2, 1, 2) 모형이 적절하다고 생각할 수 있다.

```r
fit.arma <- Arima(ex11, order=c(2, 1, 2), include.drift = T)
fit.arma
```

```
Series: ex11
ARIMA(2,1,2) with drift
```

```
Coefficients:
          ar1     ar2      ma1      ma2   drift
      -0.0688  0.0802  -0.1739  -0.8261  0.6019
s.e.   0.1113  0.0982   0.0858   0.0851  0.0023

sigma^2 = 1.108:  log likelihood = -292.4
AIC=596.79   AICc=597.23   BIC=616.55
```

```
res <- fit.arma$residuals
par(mfrow=c(2,2))
plot(res, type="l")
hist(res)
qqnorm(res)
qqline(res, col="red", lwd=2)
acf(res)
```




```
Box.test(res, type="Ljung-Box") # Ljung-Box test
```

```
    Box-Ljung test

data:  res
X-squared = 0.004488, df = 1, p-value = 0.9466
```

```r
durbinWatsonTest(as.vector(res)) # Durbin-Watson test
```

```
[1] 1.985145
```

```r
jarque.bera.test(res) # Jarque-Bera test
```

```

    Jarque Bera Test

data:  res
X-squared = 2.5411, df = 2, p-value = 0.2807
```

```r
shapiro.test(res) # Shapiro-Wilk test
```

```

    Shapiro-Wilk normality test

data:  res
W = 0.99153, p-value = 0.2953
```

잔차의 plot과 함께 여러 테스트를 진행한 결과, 정규성 및 독립성을 만족한다고 볼 수 있다.

```r
Arima(ex11, order=c(3, 1, 2), include.drift = T)
```

```
Series: ex11
ARIMA(3,1,2) with drift

Coefficients:
          ar1     ar2      ar3      ma1      ma2   drift
      -0.1139  0.1145  -0.1029  -0.1227  -0.8773  0.6018
s.e.   0.0960  0.0882   0.0831   0.0672   0.0663  0.0022

sigma^2 = 1.104:  log likelihood = -291.65
AIC=597.29   AICc=597.88   BIC=620.34
```

```r
Arima(ex11, order=c(2, 1, 3), include.drift = T)
```

```
Series: ex11
ARIMA(2,1,3) with drift

Coefficients:
          ar1     ar2     ma1      ma2      ma3   drift
      -0.5770  0.0769  0.3400  -0.9366  -0.4034  0.6018
s.e.   0.3668  0.0962  0.3628   0.0730   0.3007  0.0023

sigma^2 = 1.108:  log likelihood = -291.92
```

```
AIC=597.85   AICc=598.44   BIC=620.9
```

과적합 판단을 위해 ARIMA(2, 1, 3), ARIMA(3, 1, 2) 에 대해 적합을 추가로 진행한 후, 비교해 보았다. 추가로 생긴 계수들이 0에 가깝고, 기존의 계수는 ARIMA(2, 1, 2) 의 계수와 비슷하므로 주어진 데이터 `ex_ch4_11.txt` 데이터는 ARIMA(2, 1, 2) 모형으로 잘 적합되었다고 할 수 있다.