Timeseries_HW4

김민국

2020-05-12

```r
#### 7번

set.seed(200507)
data_7_ar<- arima.sim(model = list(ar = c(0.5)), n = 100)
fit_7_ar <- arima(data_7_ar, order = c(1,0,0), include.mean = T)
fit_7_ar$aic
```

```
## [1] 297.2625
```

```r
fit_7_arma <- arima(data_7_ar, order = c(1,0,1), include.mean = T)
fit_7_arma$aic
```

```
## [1] 299.128
```

-> ar로 fitting한 aic값이 arma fitting시보다 작다.

```r
data_7_ma<- arima.sim(model = list(ma = c(0.5)), n = 100)
fit_7_ma <- arima(data_7_ma, order = c(0,0,1), include.mean = T)
fit_7_ma$aic
```

```
## [1] 285.6012
```

```r
fit_7_arma_1 <- arima(data_7_ma, order = c(1,0,1), include.mean = T)
fit_7_arma_1$aic
```

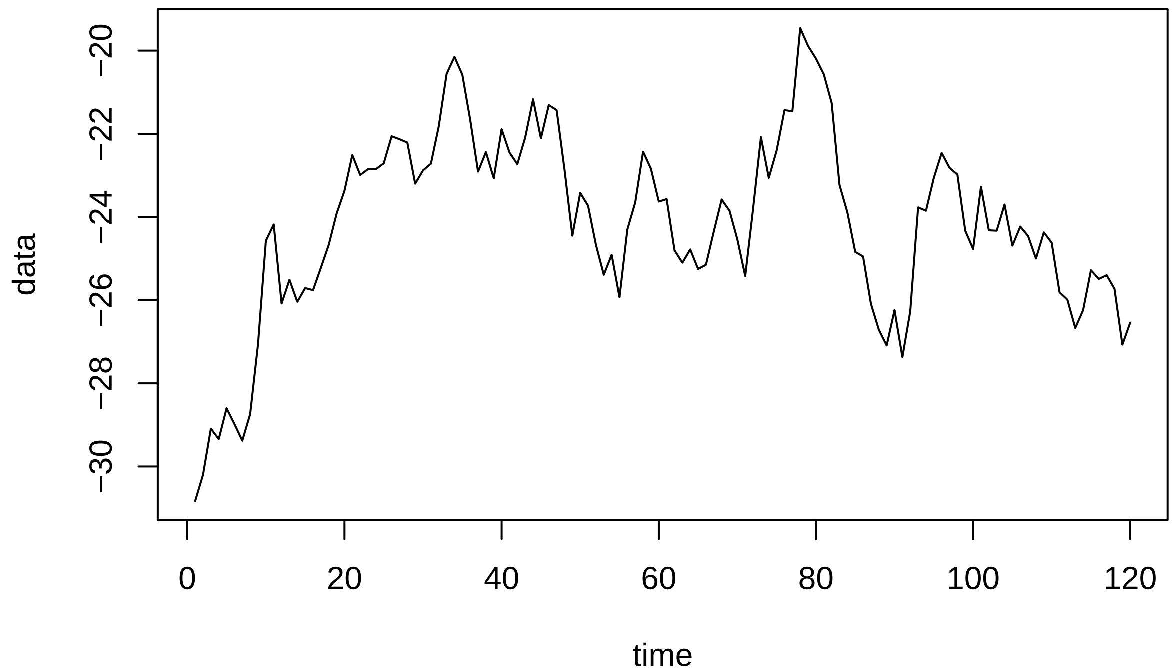
```
## [1] 287.5638
```

-> ma로 fitting한 aic값이 arma fitting시보다 작다.
-> 두 경우 모두 aic 기준 ar(1), ma(1)로 적합하는 것이 더 낫다. ->arma(1)으로 적합하는 것은 overfitting으로 생각할 수 있다.

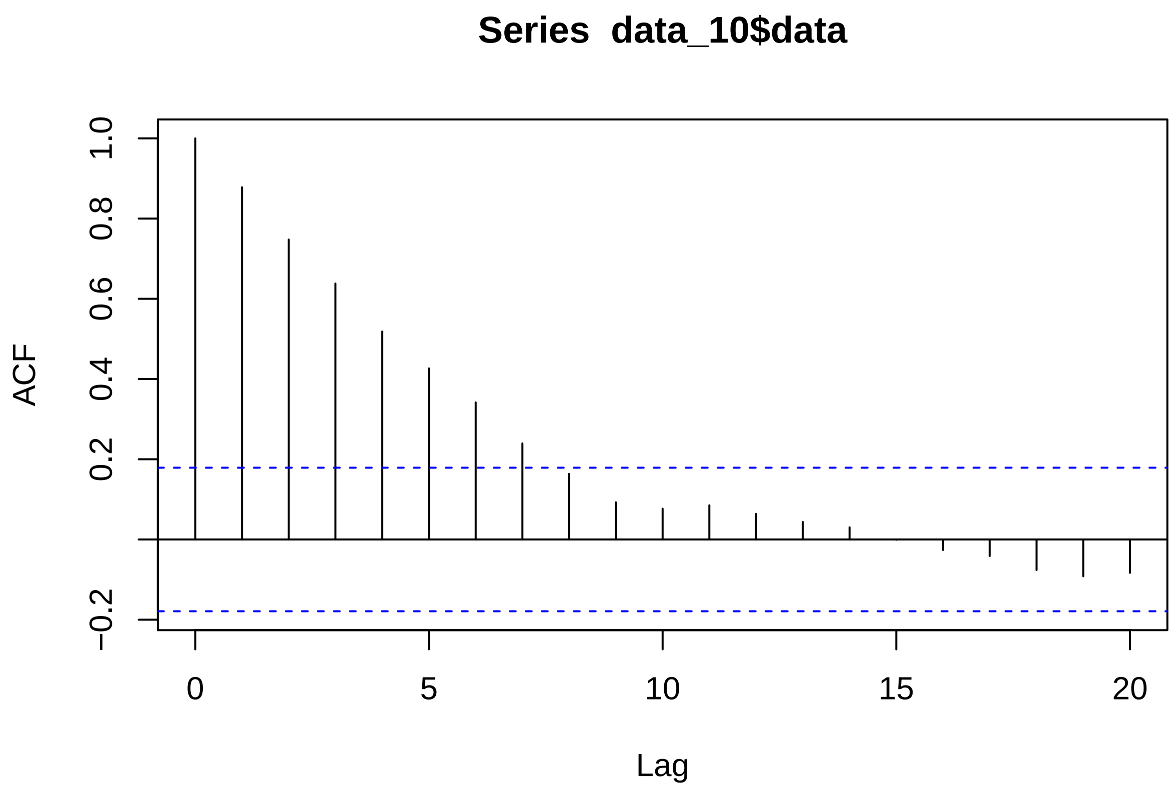
```r
#### 10번

data_10 <- read.csv("ex_ch4_10.txt")

##### (1)
plot(data_10$data, type = 'l', xlab = "time", ylab = "data")
```




```
acf_10 <- acf(data_10$data)
```




```
pacf_10 <- pacf(data_10$data)
```

-> acf를 통해서는 모수값이 매우커진 모델을 생각해야 하지만 pacf를 통해서는 2 이상부터는 0에 매우 가까운 값을 가지고 있으므로 ar(1) 모델로 생각할 수 있다.

```r
##### (2)
adf.test(data_10$data)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data_10$data
## Dickey-Fuller = -3.0271, Lag order = 4, p-value = 0.15
## alternative hypothesis: stationary
```

```r
adf.test(data_10$data, alternative = "explosive")
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data_10$data
## Dickey-Fuller = -3.0271, Lag order = 4, p-value = 0.85
## alternative hypothesis: explosive
```

-> p-value가 0.15로 유의수준 0.05보다 크므로 귀무가설을 기각할 수 없다. 대립가설을 채택할 만한 증거가 되지

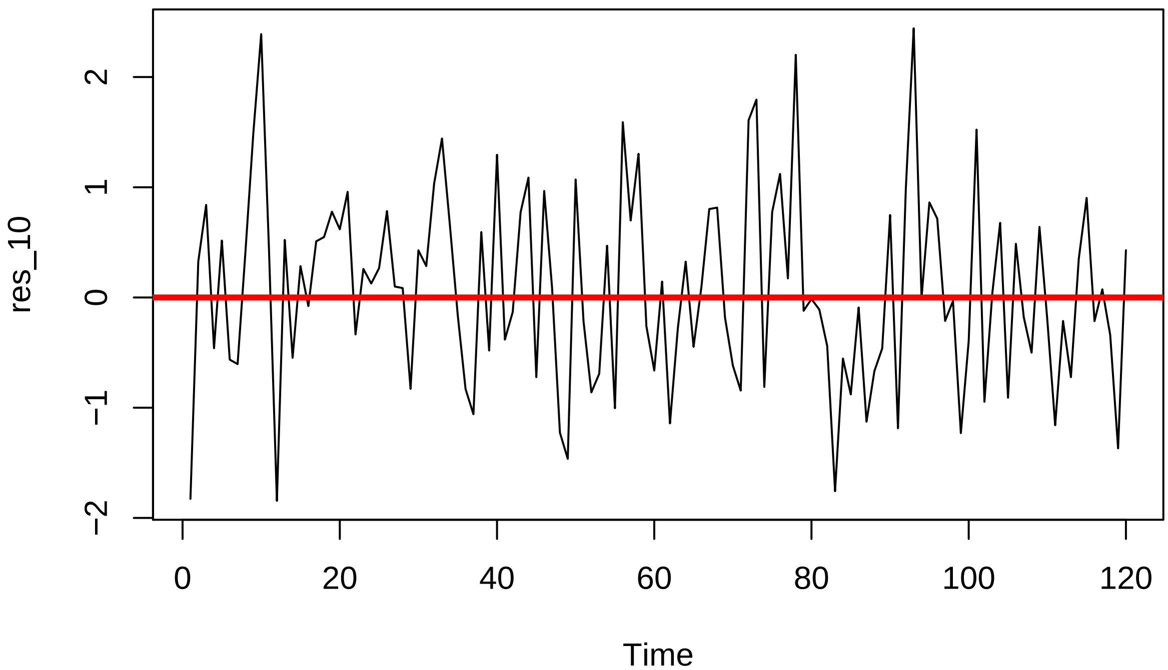
못합니다. 따라서 unit root를 가진다고 볼 수 있다.
-> 만일 data가 stationary하지 않다고 생각해서 검정을 진행한 결과 p-value가 0.85로 귀무가설을 기각할 수 없다. 대립가설을 채택할 만한 증거가 되지 못합니다. 따라서 unit root를 가진다고 볼 수 있다.

```r
##### (3)
auto.arima(data_10$data)
```

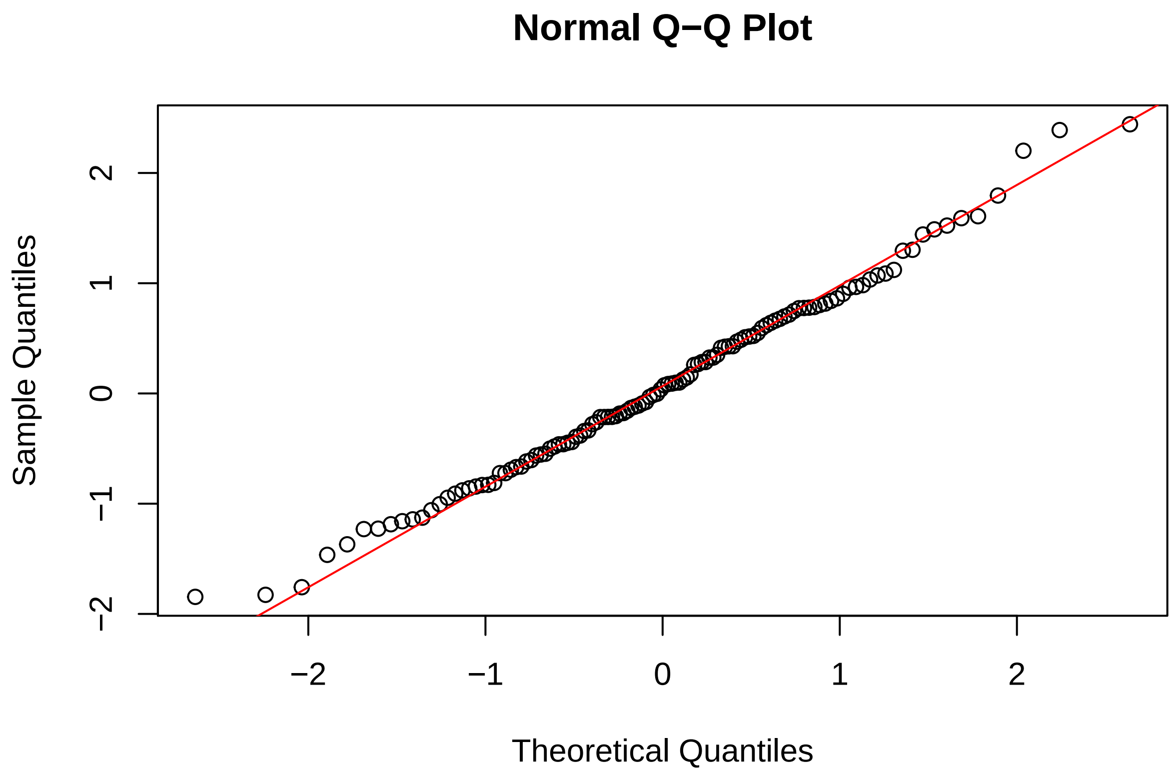
```
## Series: data_10$data
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##         ar1      mean
##       0.946  -25.1935
## s.e.  0.033    1.3969
##
## sigma^2 estimated as 0.7768:  log likelihood=-155.24
## AIC=316.48   AICc=316.69   BIC=324.84
```

-> arima(1,0,0) model이 선택되었으며 이는 (1)에서 생각한 ar(1) 모델의 결과와 같다.

```r
##### (4)
fit_10 <- Arima(data_10$data, order = c(1,0,0))
res_10 <- fit_10$residuals
plot(res_10)
abline(h = 0, col = 'red', lwd = 3)
```




```r
qqnorm(res_10)
qqline(res_10, col = "red")
```




```r
jarque.bera.test(res_10)
```

```
##
##  Jarque Bera Test
##
## data:  res_10
## X-squared = 1.0901, df = 2, p-value = 0.5798
```

```r
shapiro.test(res_10)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_10
## W = 0.99141, p-value = 0.6649
```

-> residual의 분포가 0을 기준으로 매우 random하게 분포되어 있어서 독립적이라고 생각할 수 있다.
-> QQplot을 그렸을 때 Y=X 직선과 매우 유사하므로 잔차가 정규분포를 따른다고 볼 수 있다.
-> jarque bera test와 shapiro test 모두 p-value가 유의수준 0.05보다 큰 값을 가지므로 귀무가설을 기각할 수 없다.
-> 따라서 잔차가 정규분포를 따른다고 생각할 수 있다.

```r
auto.arima(data_10$data)$coef
```

```
##        ar1  intercept
##   0.946014 -25.193511
```

```r
Arima(data_10$data, order = c(2,0,0))$coef
```

```
##         ar1         ar2   intercept
##   1.0574542  -0.1237787 -24.9383117
```

```r
Arima(data_10$data, order = c(1,0,1))$coef
```

```
##         ar1         ma1   intercept
##   0.9268554   0.1305274 -24.9637649
```

```r
pt(Arima(data_10$data, order = c(2,0,0))$coef[2] / 0.0936, df = 119)
```

```
##       ar2
## 0.0942819
```

```r
pt(Arima(data_10$data, order = c(1,0,1))$coef[2] / 0.1005, df = 119)
```

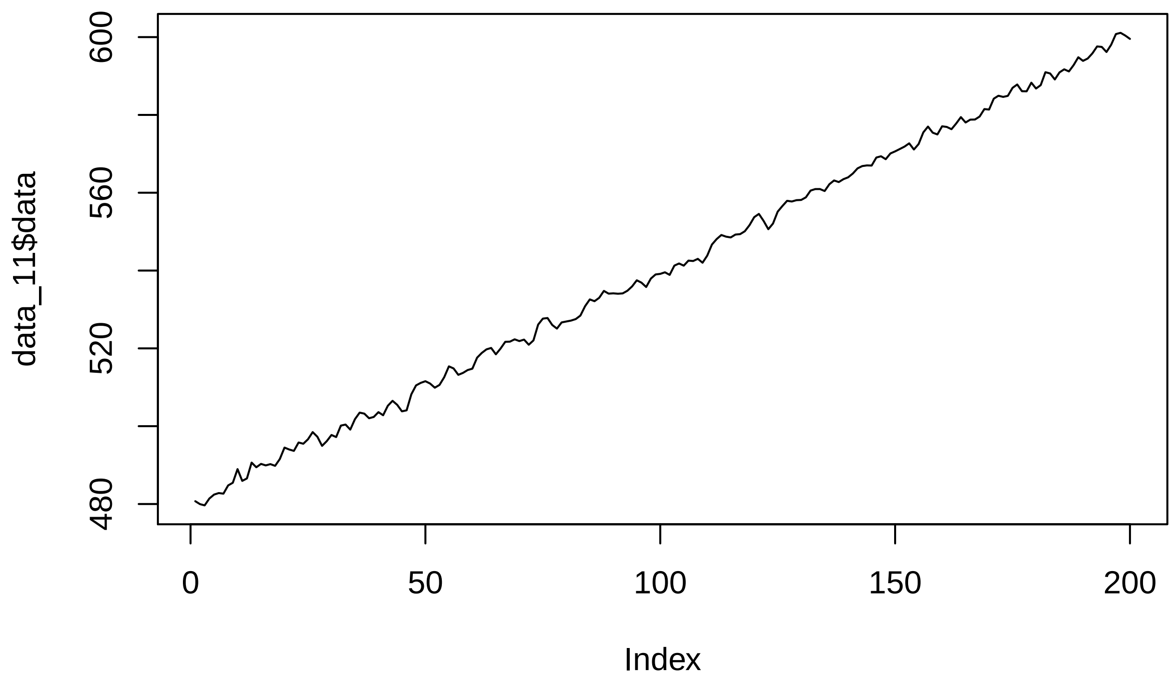
```
##       ma1
## 0.9017348
```

-> 기존 ar(1)모델에서 모수를 1씩 증가시킨 ar(2)모델과 arma(1,1)모델에서 ar1애 대한 계수의 변화가 크지 않다.
-> 새로 추가된 모수들은 유의수준 0.05에 대해 0이라는 귀무가설을 기각할 수 없으므로 0으로 생각할 수 있다.
-> 따라서 과적합 진단에 의해서도 ar(1) 모형을 취할 수 있다.

```r
#### 11번

data_11 <- read.csv("ex_ch4_11.txt")

##### (1)
plot(data_11$data, type = "l")
```

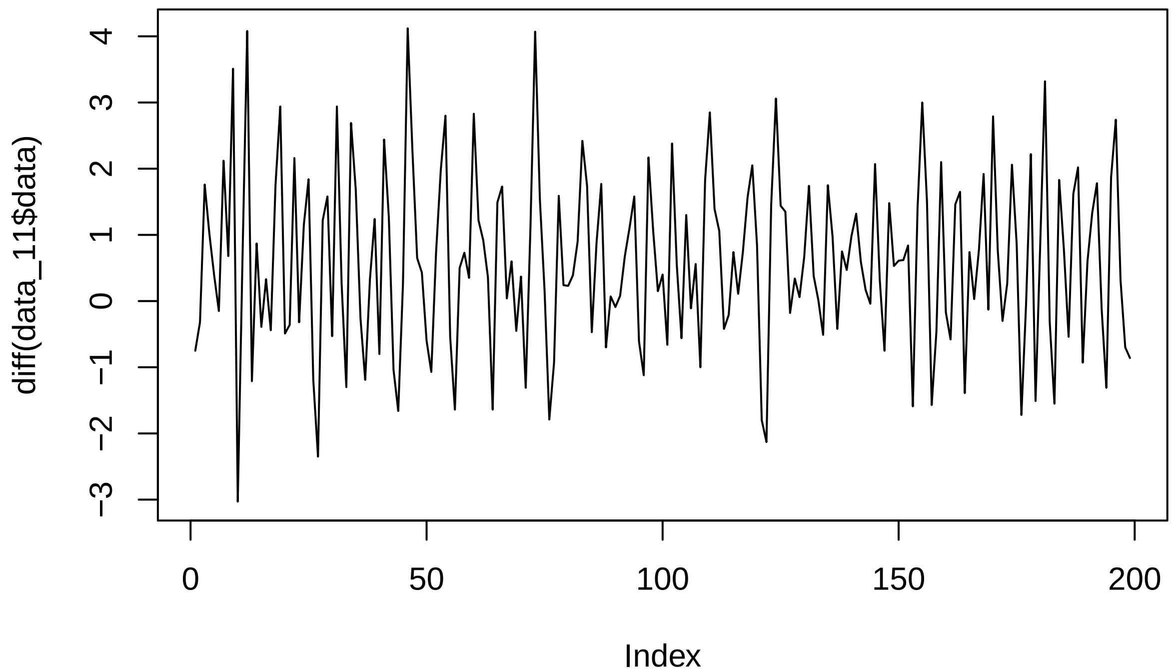


-> 계속하여 증가하고 있으므로 정상 시계열로 볼 수 없다.

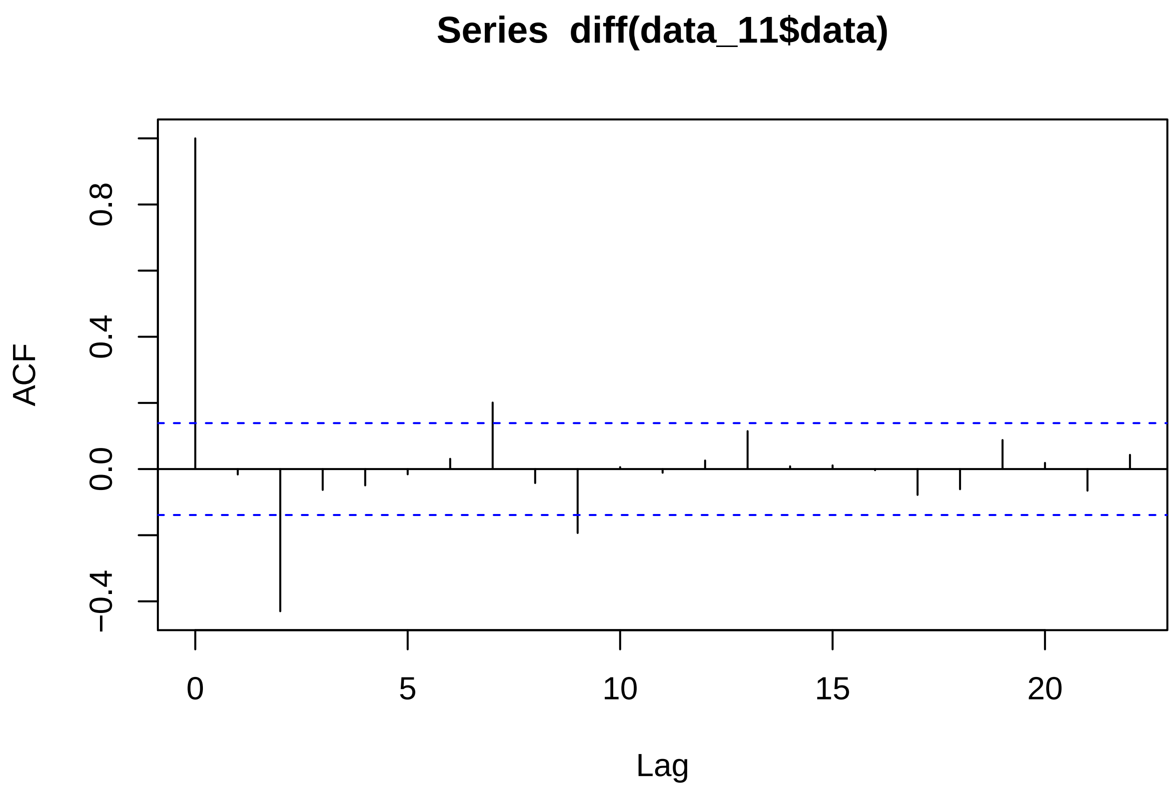
```r
##### (2)
```

-> 선형적으로 증가하기 때문에 차분을 해줘야 할 것 같다.

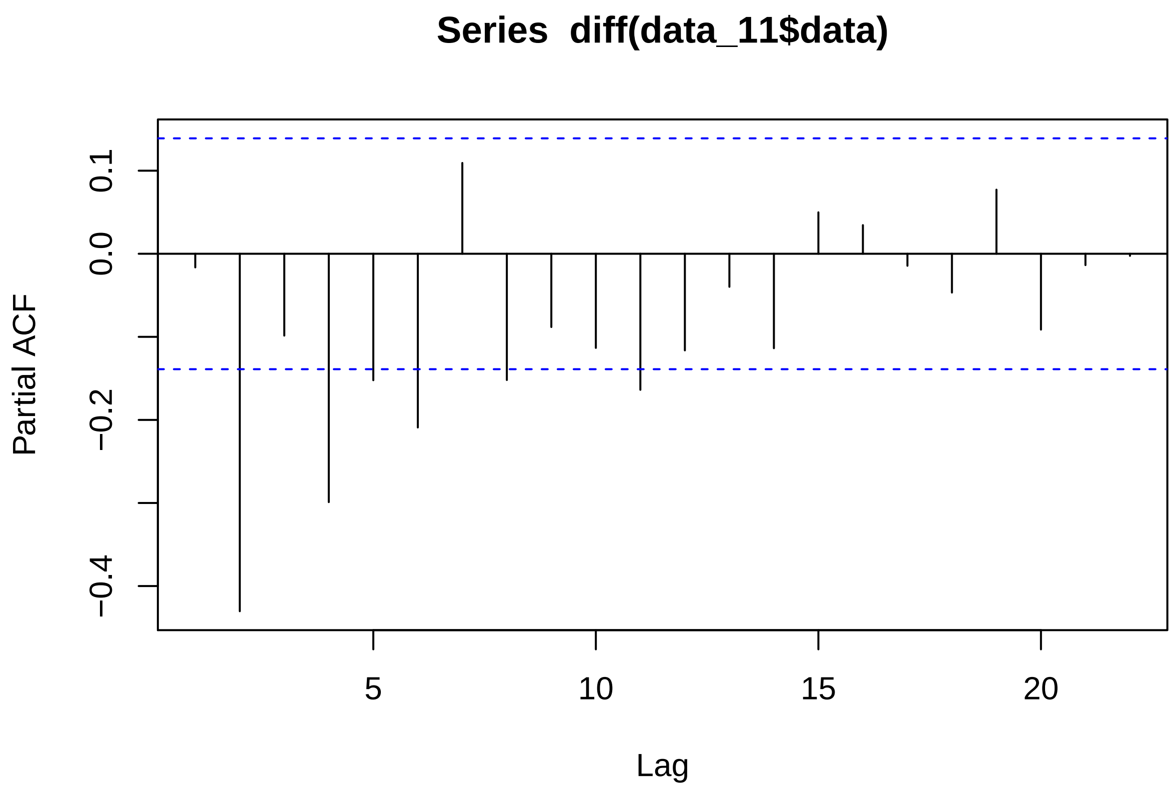
```r
##### (3)
plot(diff(data_11$data), type = 'l')
```




```
acf_11 <- acf(diff(data_11$data))
```




```
pacf_11 <- pacf(diff(data_11$data))
```



-> sacf와 spacf 모두 양수와 음수값이 반복되는 경향을 가지고 있다.

-> 두 값모두 0으로 근사해가고 있지만 그 값들이 큰 값들이기 때문에 만일 이를 기준으로 적합을 한다고 생각하면 arma 모델 이용하여 적합을 해야 할 것 같다.

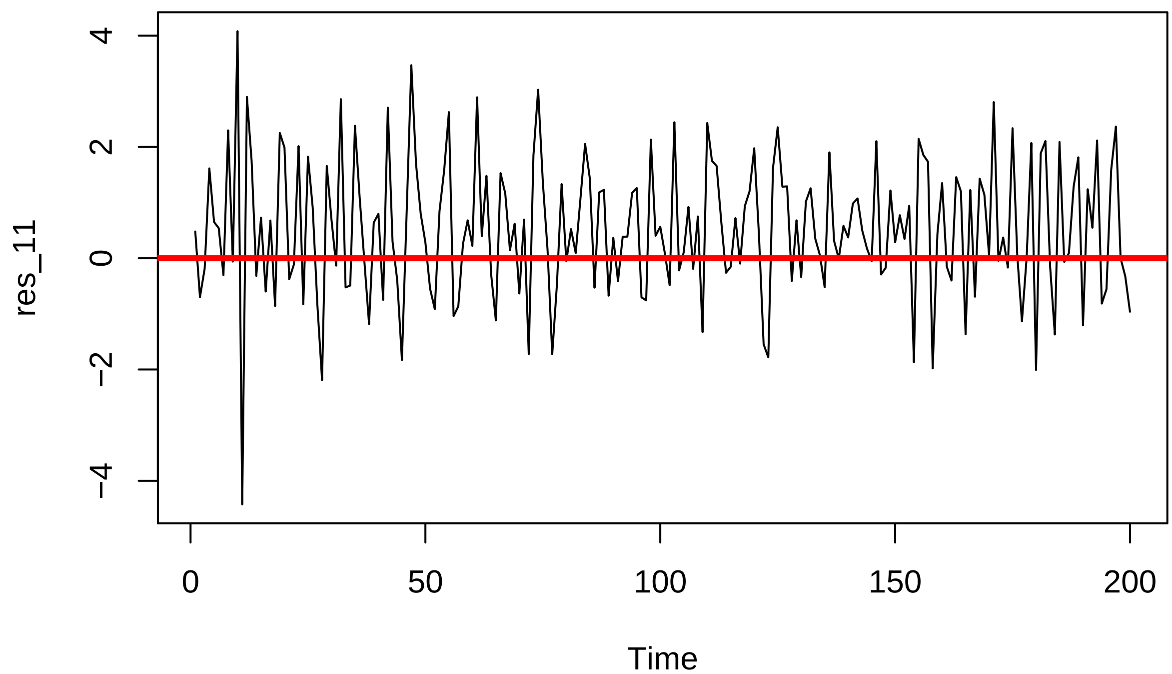
-> ar과 ma를 단독으로 사용하기에는 무리가 있어보인다.

```
##### (4)
auto.arima(data_11$data)
```

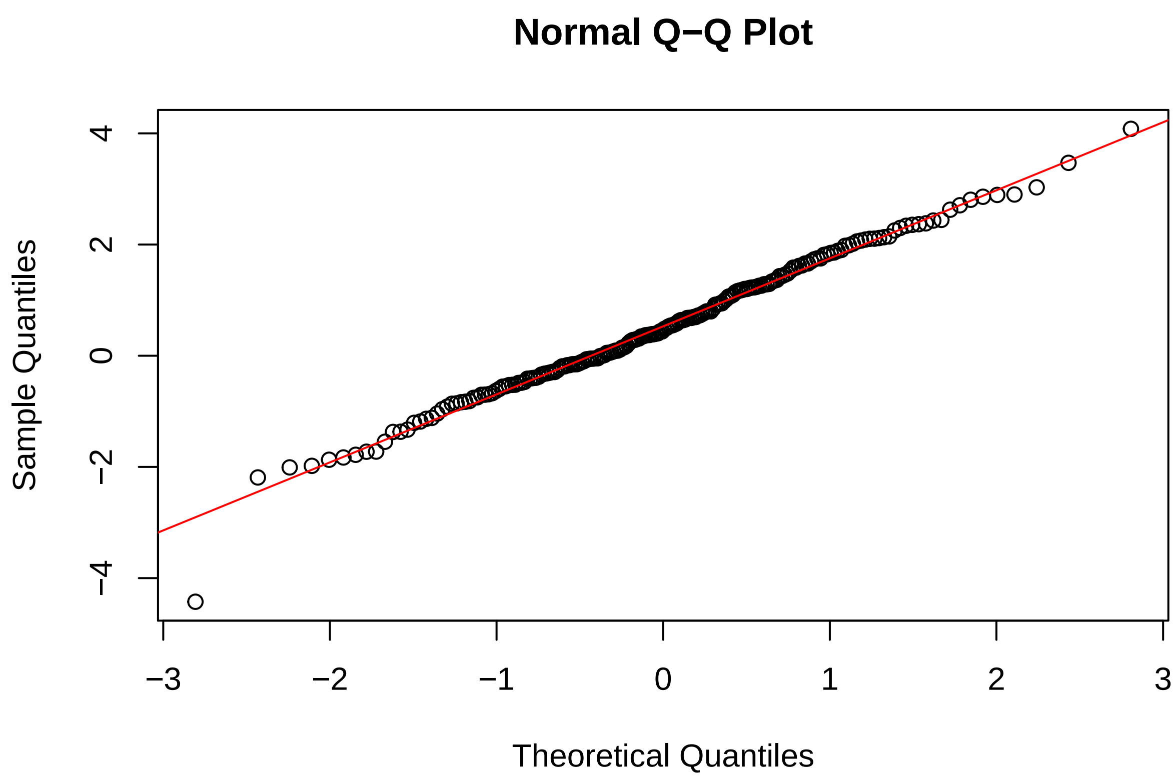
```
## Series: data_11$data
## ARIMA(2,1,1) with drift
##
## Coefficients:
##          ar1      ar2      ma1   drift
##       0.6318  -0.3314  -0.9632  0.6021
## s.e.  0.0743   0.0729   0.0522  0.0048
##
## sigma^2 estimated as 1.197:  log likelihood=-299.34
## AIC=608.68   AICc=608.99   BIC=625.15
```

-> ARIMA(2,1,1)로 적합해야 한다.

```r
fit_11 <- Arima(data_11$data, order = c(2,1,1))
res_11 <- fit_11$residuals
plot(res_11)
abline(h = 0, col = 'red', lwd = 3)
```




```r
qqnorm(res_11)
qqline(res_11, col = "red")
```




```r
jarque.bera.test(res_11)
```

```
##
##  Jarque Bera Test
##
## data:  res_11
## X-squared = 3.5951, df = 2, p-value = 0.1657
```

```r
shapiro.test(res_11)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_11
## W = 0.99099, p-value = 0.2481
```

-> residual의 분포가 0을 기준으로 매우 random하게 분포되어 있어서 독립적이라고 생각할 수 있다.
-> QQplot을 그렸을 때 Y=X 직선과 매우 유사하므로 잔차가 정규분포를 따른다고 볼 수 있다.
-> jarque bera test와 shapiro test 모두 p-value가 유의수준 0.05보다 큰 값을 가지므로 귀무가설을 기각할 수 없다. 따라서 잔차가 정규분포를 따른다고 생각할 수 있다.

```r
auto.arima(data_11$data)$coef
```

```
##        ar1        ar2        ma1      drift 
##  0.6317678 -0.3313508 -0.9631692  0.6020612
```

```r
Arima(data_11$data, order = c(3,1,1))$coef
```

```
##        ar1        ar2        ar3        ma1 
##  0.9734126 -0.4039792  0.4305425 -0.9957922
```

```r
Arima(data_11$data, order = c(2,1,2))$coef
```

```
##        ar1        ar2        ma1        ma2 
## -1.6411153 -0.6582000  1.9450730  0.9539454
```

-> 기존 모델에서의 계수들과 모수를 1씩 늘렸을 때의 모델의 모수의 값을 비교했을 때 이미 충분한 차이가 있다.
-> 따라서 과적합진단 에서는 잠정 모형에 이상이 있음을 알 수 있다.
-> 과적합 진단으로는 적절한 모형을 직접적으로 찾을 수는 없다. 하지만 ARIMA의 대부분의 모형이 과적합 진단을 통해서 적절하지 않다는 결론을 얻을 수 있다. 따라서 ARIMA 모형이 아닌 GARCH 모형 등을 고려해 보는 것도 좋을 것 같다.