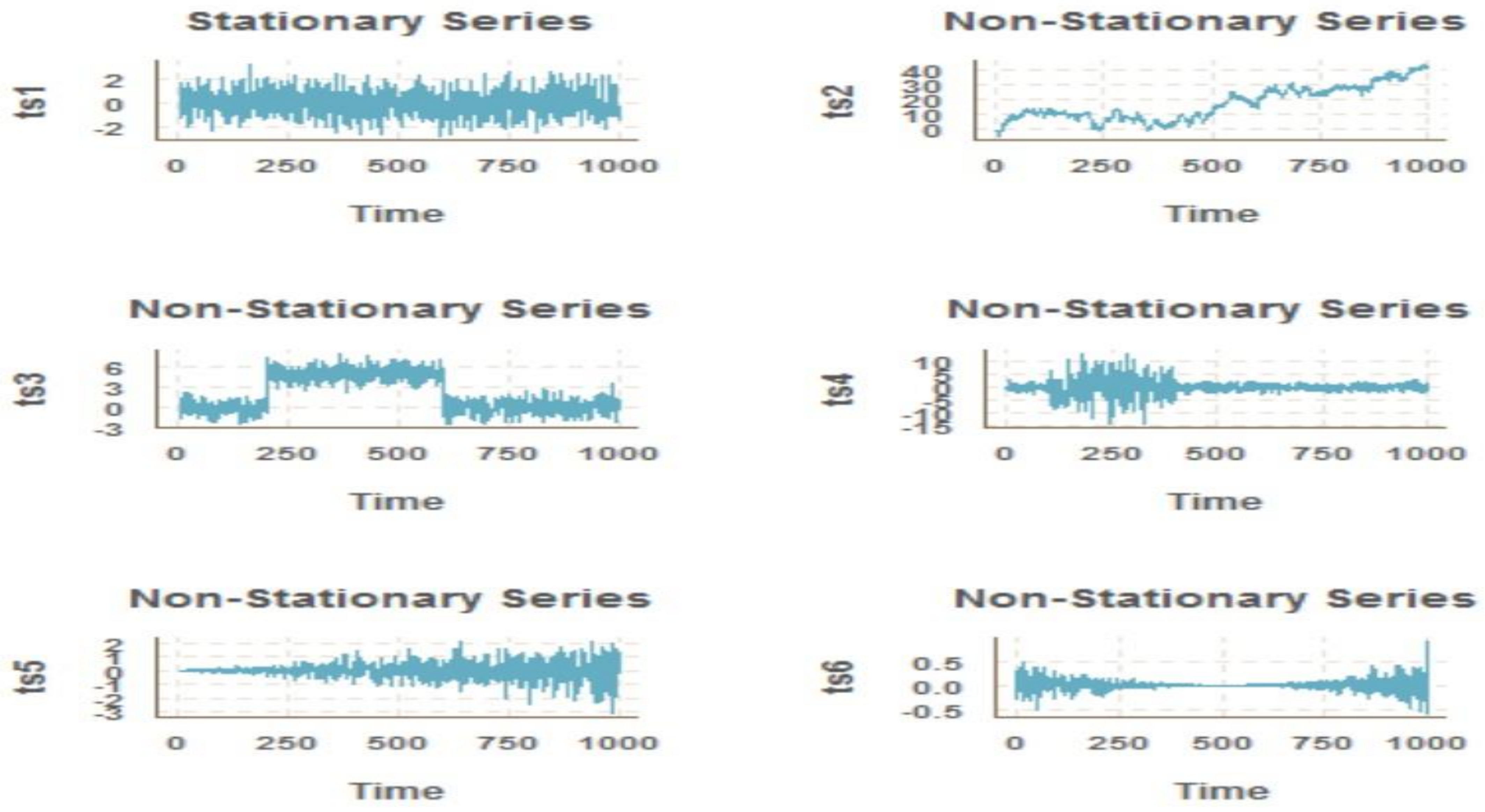
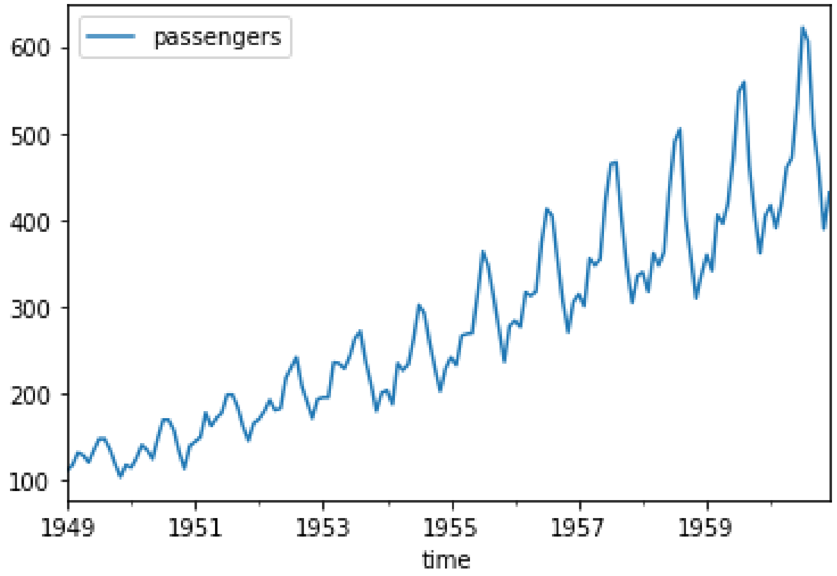
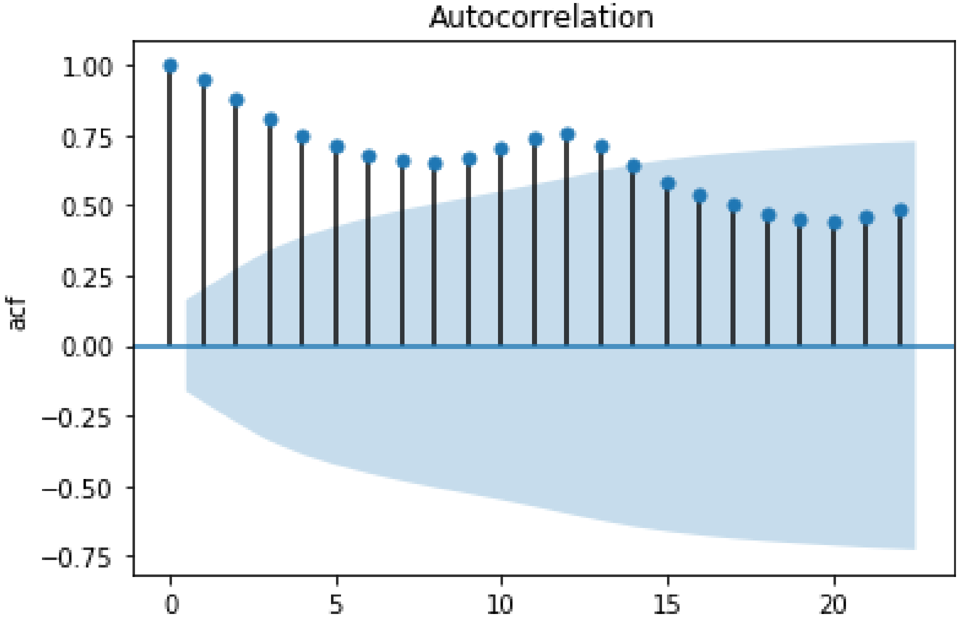
시계열예측

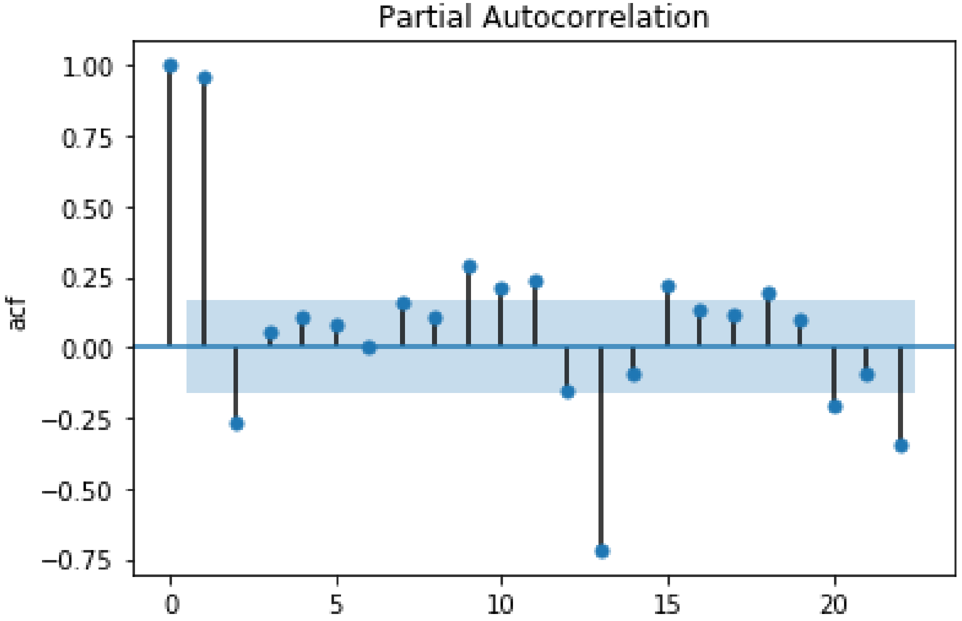


이미지출처: https://stats.stackexchange.com/questions/182764/stationarity-tests-in-r-checking-mean-variance-and-covariance

- stationary :평균, 분산이 일정
- non stationary:평균, 분산이 일정하지 않다.


passengers
600
500
400
300
200
100
1949
1951
1953
1955
1957
1959
time


Autocorrelation
acf
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0
5
10
15
20

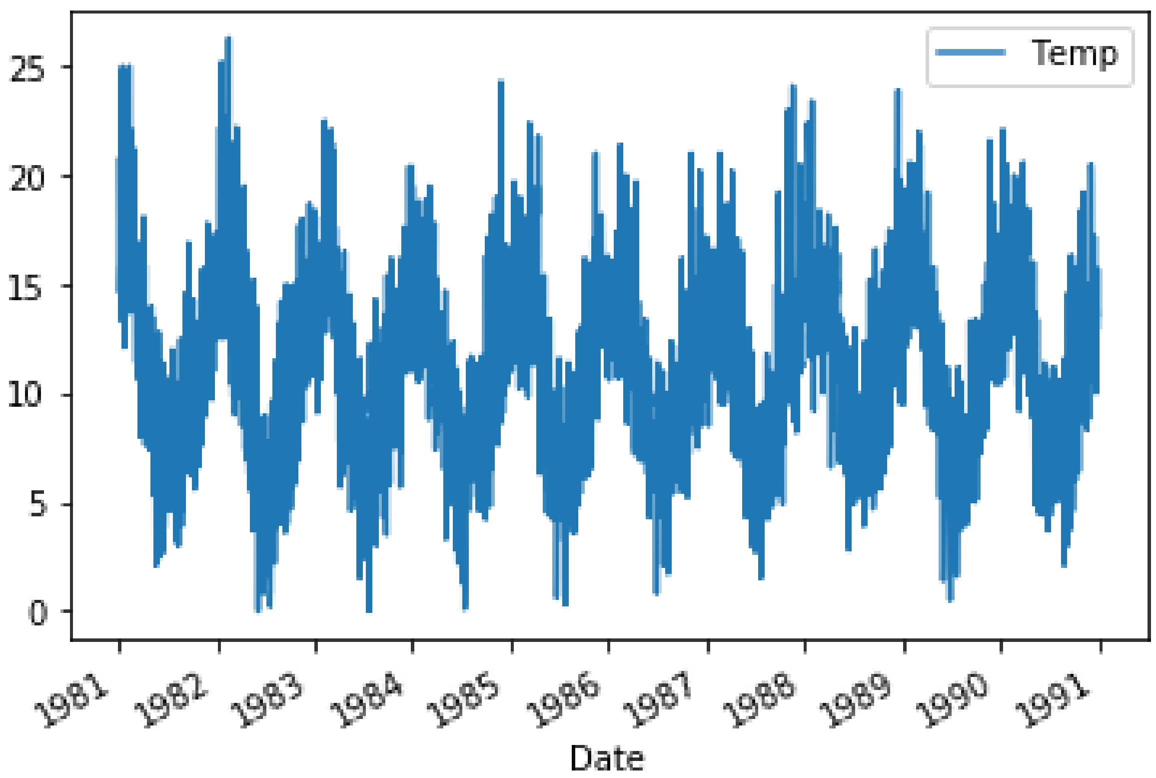

Partial Autocorrelation
acf
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0
5
10
15
20

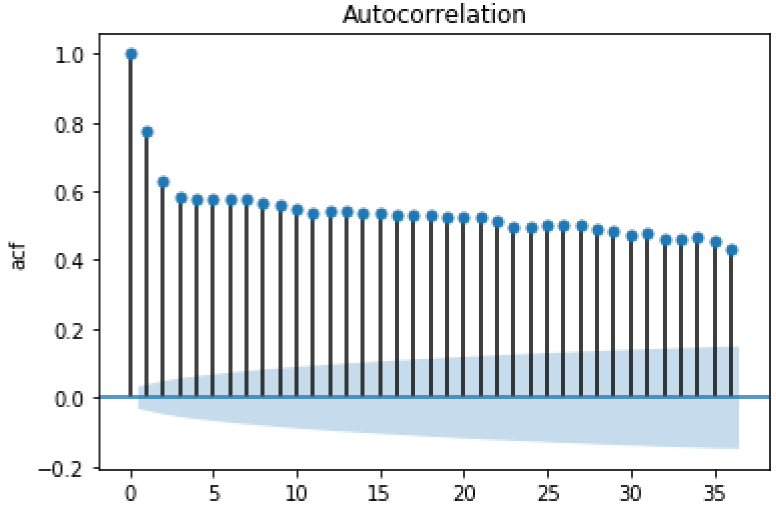
자기상관 함수(ACF : Autocorrelation function)

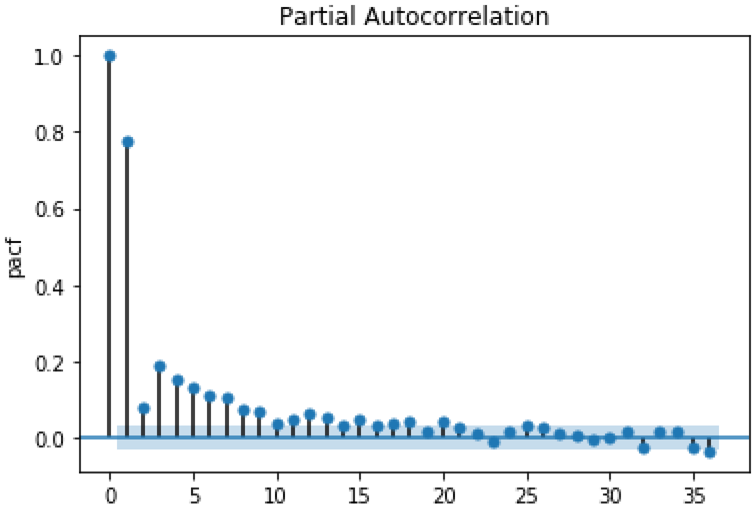
$$\rho(h) = \frac{Cov(y_t, y_{t-h})}{\sqrt{Var(y_t)}\sqrt{Var(y_{t-h})}}, \; h = 0, \pm 1, \pm 2, \cdots$$

편자기상관 함수(PACF: Partial autocorrelation function)

$$\frac{Cov(y_t, y_{t-h} | y_{t-1}, y_{t-2}, \cdots, y_{t-h+1})}{\sqrt{Var(y_t | y_{t-1}, y_{t-2}, \cdots, y_{t-h+1})}\sqrt{Var(y_{t-h} | y_{t-1}, y_{t-2}, \cdots, y_{t-h+1})}}$$


Temp
25
20
15
10
5
0
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
Date


Autocorrelation
acf
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
25
30
35


Partial Autocorrelation
pacf
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35

AR(Autoregressive)

목표 예상 변수(forecast variable)에 대해 과거 값을 이용

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

MA(Moving average)

과거 예측 오차(forecast error)을 이용

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

ARMA
(Autoregressive and Moving average)

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

ARIMA Model

Y_t → AR filter → Integration filter → MA filter → ε_t

(long term) (stochastic trend) (short term) (white noise error)

ARIMA (2,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \varepsilon_{t-1}$

ARIMA (3,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + b_1 \varepsilon_{t-1}$

ARIMA (1,1,0) $\Delta y_t = a_1 \Delta y_{t-1} + \varepsilon_t$, where $\Delta y_t = y_t - y_{t-1}$

ARIMA (2,1,0) $\Delta yt = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \varepsilon_t$ where $\Delta yt = yt - yt\text{-}1$

Integration filter:
Non Stationary (차분)→ Stationary

AR I MA

= = = = = = = = = = =

p d q

p: AR 파라미터 갯수
d: 차분 갯수
q: MA 파라메터의 갯수

ARIMA(p, d, q)로 구성할 수 있다. 만약 ARIMA(1,2,1) 이라면
AR과 MA를 1개만큼의 과거를 window로 활용하고,
차분은 2 만큼을 활용

차분

2
7
10
5
8
6

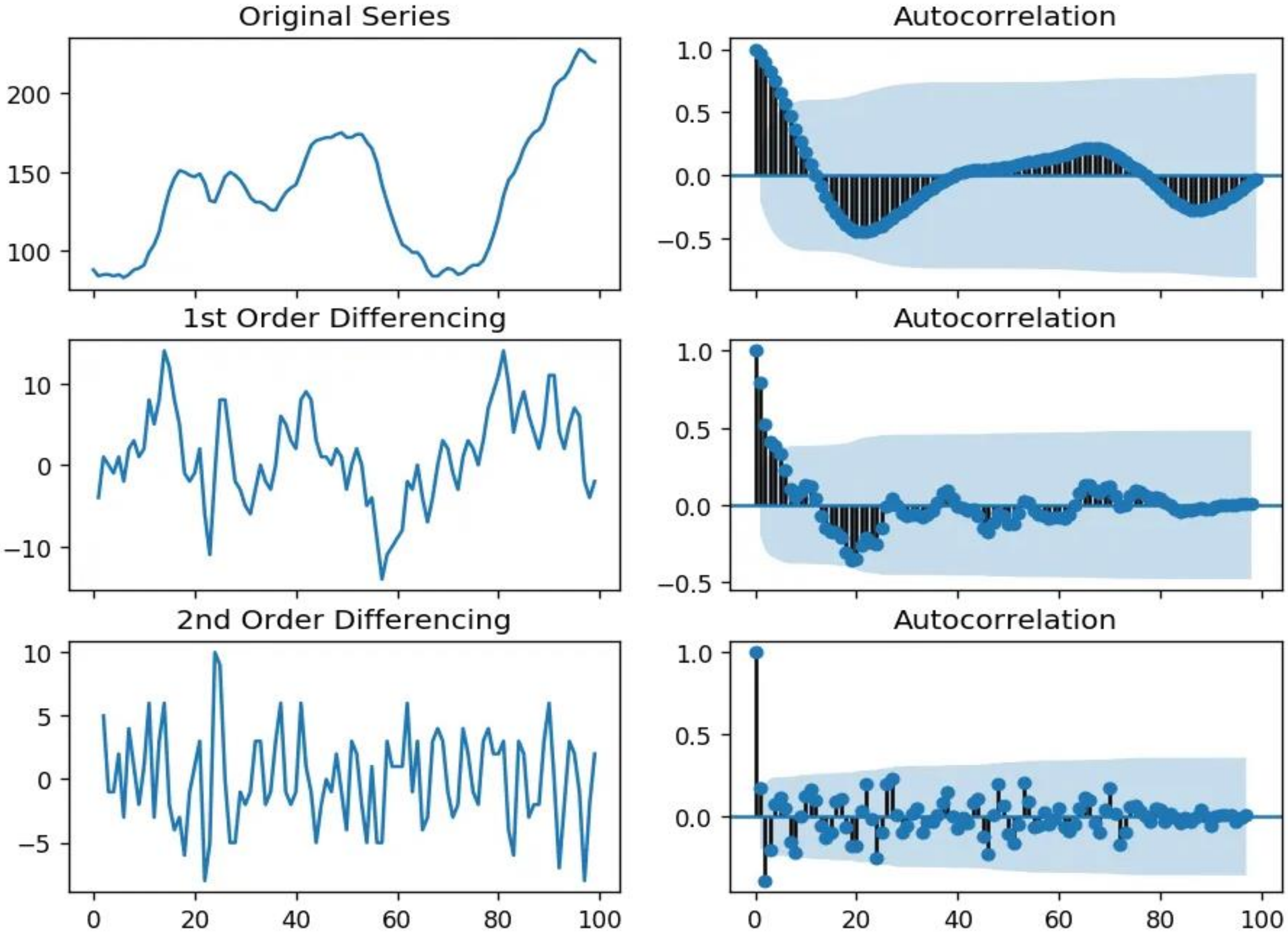
2
7
10
5
8
6

5
3
-5
3
-2

- 1차 :$Y_t = X_t - X_{t-1}$
- 2차 :$Y_t = X_t - X_{t-2}$

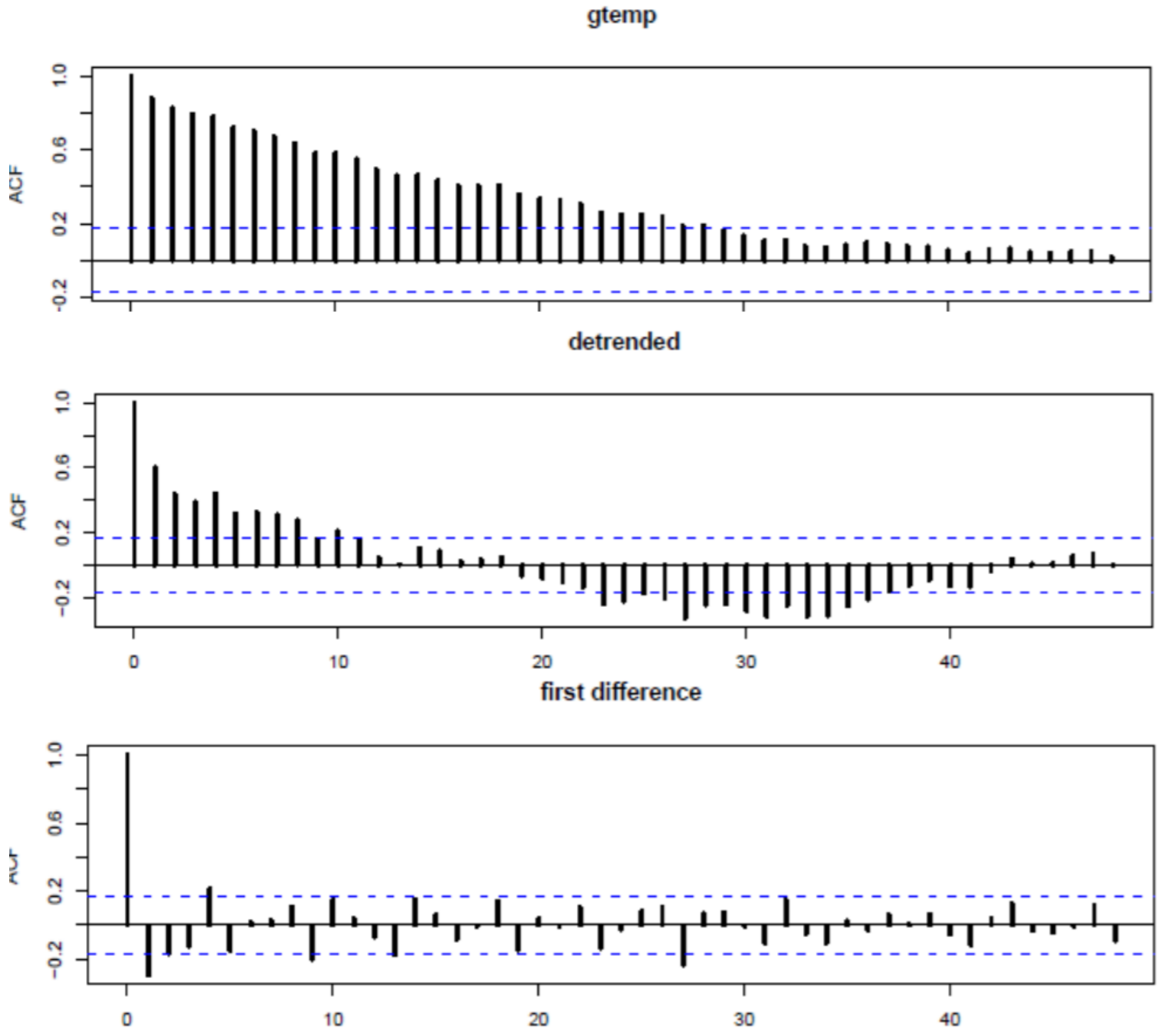
현시점 데이터에서 n시점 데이터를 빼서
stationary로 변환

대부분 1차또는 2차 차분으로 stationary 가 되며 그이상인 경우는 ARMA 모델로 적합하지
않은것로 본다



Original Series
Autocorrelation
1st Order Differencing
Autocorrelation
2nd Order Differencing
Autocorrelation

Global Temperature 데이터의 Autocorrelation Function(ACF)



Original signal X_t

빨리 **decay** 하지 않는다

잔차 **Residual Y_t**

빨리 **decay** 한다

1 step 차이 **$X_t - X_{t-1}$**

더 빨리 **decay** 한다

ACF의 모양으로 적절한 **ARIMA** 모형 선택 가능

1) acf 확인
2) stationary 변환