WEEK3 (1)

(Recap)

- **정상 시계열의 조건**

1) $E(e_t) = 0$ ← 잔차의 평균은 0 임

2) $Var(e_t) = \sigma_e^2$ ← 잔차의 분산은 일정함 (상수)

3) $Var(X_t) = Var(X_{t-1})$ ← 현재 데이터의 분산과 이전 데이터의 분산은 동일함

4) $Cov(X_{t-1}, e_t) = 0$ ← 이전 데이터와 잔차는 서로 무관함, 공분산 = 0

5) $Cov(e_t, e_{t-1}) = 0$ ← 이전 잔차와 현재의 잔차는 서로 무관함, 공분산 = 0

1. **ARMA 모형의 식별**
   - 모형의 식별 및 추정 과정
   a. 그래프를 그려, 정상성 여부 검토(비정상적인 경우 추세제거, 계절성 제거, 분산안정화 등을 거쳐 정상적 시계열로 변환)
   b. 표본 ACF/PACF를 구하고 정상성 여부 확인
   c. 표본값과 이론적 ACF/PACF를 비교하여 모형의 차수 p와 q를 구함
   d. 모수 추정 및 잔차 확인
   e. 잔차분석 → 백색잡음을 따르는지 확인

- 표본 ACF 및 PACF 산출:

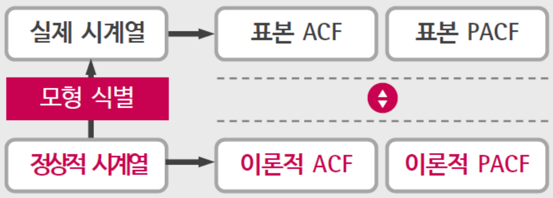
$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-k}(Z_t - Z)(Z_{t+k} - Z)}{\sum_{t=1}^{n}(Z_i - Z)^2}$$

$$\hat{P}(k) = \hat{\phi}_{kk} = \frac{\hat{\rho}(k) - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\hat{\rho}(k-j)}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\hat{\rho}(j)}$$

→ 시차별로 해당 값을 구할 수 있음
→ 시차별 PACF는 ACF로 표현되므로 ACF를 추정하여 표본 PACF를 구할 수 있음

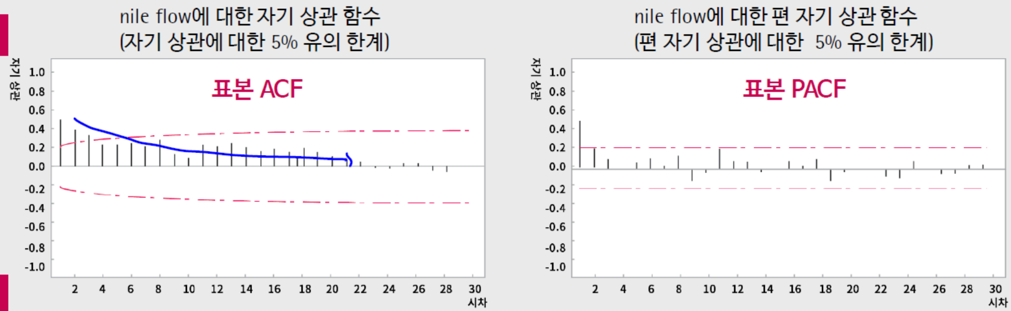
- 그래프와 표본 ACF/PACF 값을 활용하여 어떤 모형을 따르는지 확인

모형	ACF	PACF
AR(p)	지수적으로 감소	시차 p 이후 절단
MA(q)	시차 q 이후 절단	지수적으로 감소
ARMA(p, q)	시차 (q-p) 이후 지수적 감소	시차 (p-q) 이후 지수적 감소



- 예 3.2 계속 (나일강 유량)

단계2



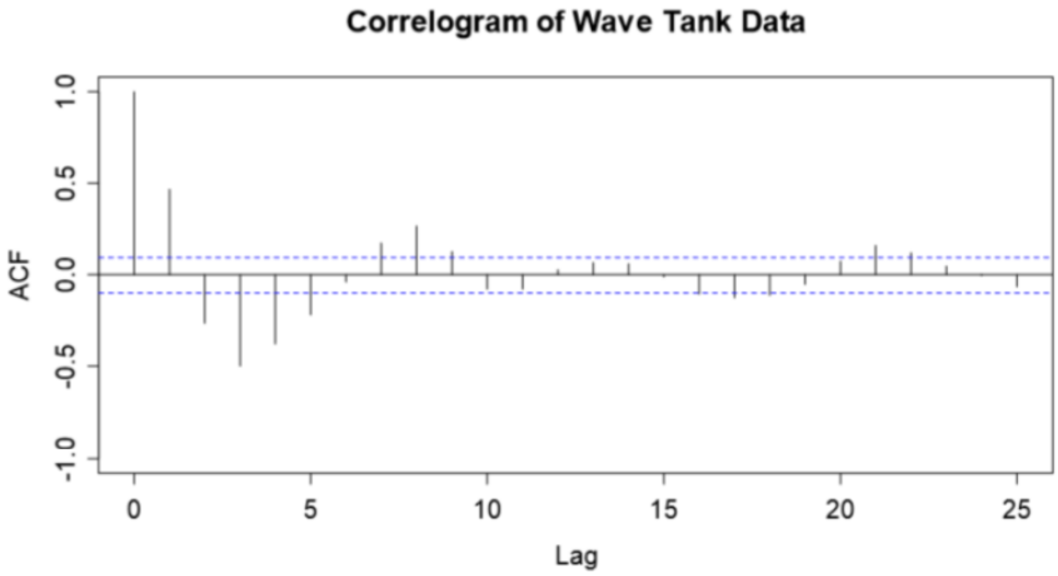
단계3

- 표본 ACF는 지수적으로 감소하고 표본 PACF는 시차 1또는 2 이후 절단되므로 AR(1) 또는 AR(2) 모형으로 보임
- 즉, p=1 또는 2; q=0

→ AR

#1

- 실제 활용되는 모형 식별 방법: Correlogram(자기상관성도표)
  - ACF/PACF 함수 형태를 보고 시차 k가 달라짐에 따른 함숫값을 확인하여 모형 식별
  - x축에 시차(lag), y축에 상관계수를 놓고 plotting



2. **ARMA 모형의 추정**
   - 추정방법: MLE 방식 사용
   - 관측치들은 서로 독립이 아님 → 백색잡음을 이용하여 우도함수 구성
   - a1을 계산하기 위해서는 관측되지 않은 값이 필요

$$L(\theta; a_1, \dots, a_n) = (2\pi\sigma_a^2)^{-n/2} exp\left(-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right)$$

1) 조건이 있는 우도함수: 임의로 초기값을 가정하여 사용
   - 초기치를 사용하여 Σa_t^2를 최소로 하는 theta를 찾음

$$a_* = (a_{1-q}, \dots, a_{-1}, a_0) = (0, \dots, 0,0)$$
$$z_* = (z_{1-p}, \dots, z_{-1}, z_0) = (z, \dots, z, z)$$

초기값

2) 조건이 없는 우도함수: 백색잡음의 기대치로 오차항 예측 → 과거 관측치의 예측 필요
   - 과거값 예측을 위해 후방예측 기법 사용
   - 이미 관측한 미래값을 이용하여 과거값을 예측

$$Z_t = \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$
$$\Rightarrow Z_t = \phi_1 Z_{t+1} - \cdots - \phi_p Z_{t+p} + b_t - \theta_1 b_{t+1} - \cdots - \theta_q b_{t+q}$$

(예- AR(1))다음 10개 시계열 데이터가 AR(1)모형을 따른다.

-0.17, 0.24, -0.51, 1.10, 0.08, 0.32, -0.44, -1.16, -0.58, -0.28

$\phi_1 = 0.2$와 $\phi_1 = 0.4$ 일때 조건없는 제곱합을 각각 구하라.

$S(\theta) = \sum_{t=-M}^{n} E^2[a_t|\theta; z_1, \dots, z_n]$

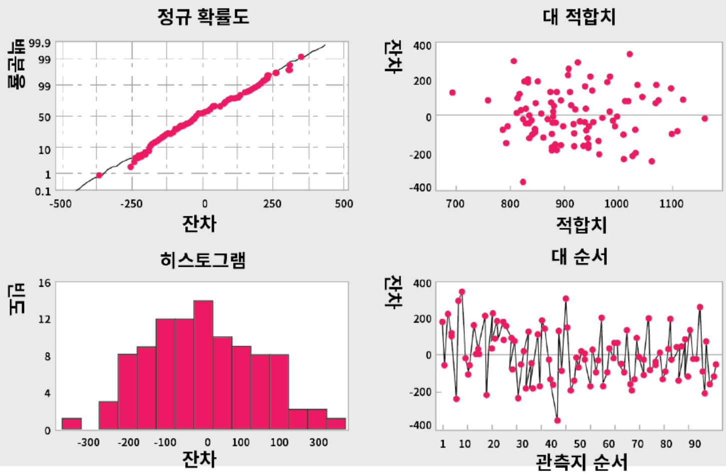
- $Z_t = \phi_1 Z_{t-1} + a_t \Leftrightarrow Z_t = \phi_1 Z_{t+1} + b_t$
- $E[a_t] = Z_t - \phi_1 Z_{t-1}$, $t = 2,3, \dots$
- $E[a_1] = E[Z_1] - \phi_1 E[Z_0] = (1 - \phi_1^2) z_1$
- $E[a_{-j}] = \phi_1^{j+1}(1 - \phi_1^2) z_1$, j= 0,1,2, ...
- $S(\phi_1 = 0.2) = 3.5236$
- $S(\phi_1 = 0.4) = 3.7060$

Z_t		$\phi_1 = 0.2$		$\phi_1 = 0.4$	
		a_t	a_t^2	a_t	a_t^2
-3	-	-0.0040	0.0000	-0.00366	0.0000
-2	-	-0.0100	0.0001	-0.00914	0.0001
-1	-	-0.0250	0.0006	-0.02285	0.0005
0	-	-0.0626	0.0039	-0.05712	0.0033
1	-0.17	-0.1564	0.0245	-0.1428	0.0204
2	0.24	0.2740	0.0751	0.308	0.0949
3	-0.51	-0.5580	0.3114	-0.606	0.3672
4	1.1	1.2020	1.4448	1.304	1.7004
5	0.08	-0.1400	0.0196	-0.36	0.1296
6	0.32	0.3040	0.0924	0.288	0.0829
7	-0.44	-0.5040	0.2540	-0.568	0.3226
8	-1.16	-1.0720	1.1492	-0.984	0.9683
9	-0.58	-0.3480	0.1211	-0.116	0.0135
10	-0.28	-0.1640	0.0269	-0.048	0.0023

- 모형의 검증
  - 오차항에 대한 가정(백색잡음에 대한 가정)
  1) 정규성: q-q plot으로 확인
  2) 등분산성 및 패턴 유무: 잔차도 확인
  3) 랜덤성: ACF/PACF로 시차별 상관계수가 모두 0인지 확인

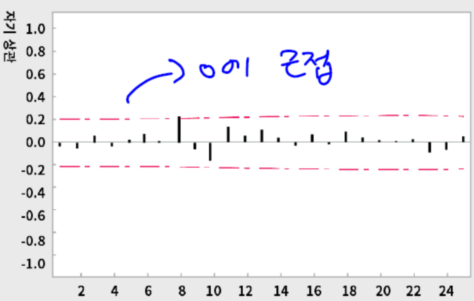
- 예 (나일강 유량)
  - 정규확률도를 볼 때 잔차가 정규분포를 따름
  - 잔차 산점도로 부터 등분산성, 패턴없음이 관찰

nile_flow의 잔차 그림

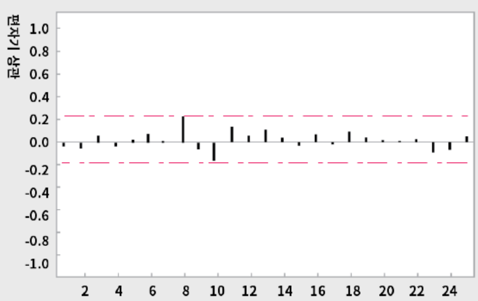


- 예 (나일강 유량)
  - 잔차에 대한 ACF/PACF로 부터 백색잡음 확인

nile_flow에 대한 잔차의 ACF
(자기 상관에 대한 5% 유의 한계)



nile_flow에 대한 잔차의 PACF
(편 자기 상관에 대한 5% 유의 한계)



+포트만토 검정: 시차 수에 근거하여 ‘전체적인’ 무작위성을 시험
  - 시계열의 자기 상관 그룹 중 어떤 그룹이 0과 다른지에 대한 통계적 시험

3. **ARMA 모형의 예측**
   - AR 모형: 평균회귀 모형. 현재 시계열 데이터는 과거 데이터에 종속

- 최소 평균제곱오차 예측치
  - 과거 시계열 관측치의 선형결합으로 예측

$$f_{n,k} = b_k Z_n + b_{k+1} Z_{n-1} +$$
$$\Downarrow \; Z_t = a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \cdots$$
$$\underline{f_{n,k} = c_k a_n + c_{k+1} a_{n-1} + \cdots}$$
→ MA 모형으로 표현

$f_{n,k}$: 시점 n에서 k 시점 이후의 시계열 예측치

→ MSE를 최소로 하는 계수 c들을 찾음(추정치)

$f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \ldots]$: k시점 이후 예측치는 조건부 기대치와 동일

$v_{n,k} = Var[Z_{n+k} | Z_n, Z_{n-1}, \ldots]$: k시점 이후 예측오차분산은 조건부 분산과 동일

- 예측식: 시점 n까지의 과거 데이터를 바탕으로 조건부 기댓값을 적용하여 구함

- k-단계 예측식: $f\_(n,k)=\boldsymbol{E}[Z\_(n+k) \;\; |Z\_n, \boldsymbol{Z\_(n-1)},\cdots], \;\; \boldsymbol{k=1, 2,}\cdots$
- 계수는 추정값 이용, 과거 오차항은 잔차를 이용하여 추정

- AR(1) 모형의 예측 과정 전개

$AR(1): X_n = aX_{n-1} + e_n$

$Forecast: \hat{X}_{n+1} = E[X_{n+1} | X_n]$ ← n+1 시점의 기댓값은 n 시점까지의 데이터를 알고 있는 상태에서 n+1 시점의 기댓값으로 표현.

$= E[a X_n + e_{n+1} | X_n]$

$= a E[X_n | X_n] + E[e_{n+1} | X_n]$ ←

- n 시점까지의 데이터를 알고 있는 상태에서 n 시점의 기댓값은 n 시점에 관찰된 데이터 임. $E[X_n | X_n] = X_n$
- 모형이 적합하다면 n+1 시점의 잔차도 정규분포 특성을 가져야 하므로 잔차의 평균은 0 이 됨. $E[e_{n+1} | X_n] = 0$

$= a X_n$ ← 1단계 전방 예측값

$\hat{X}_{n+2} = E[X_{n+2} | X_n]$

$= E[a X_{n+1} + e_{n+2} | X_n]$

$= a E[X_{n+1} | X_n] + E[e_{n+2} | X_n]$

$= a^2 X_n$ ← 2단계 전방 예측값

- AR(2) 모형의 예측:

AR(2) 모형: $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$

한단계 예측식: $f_{n,1} = E[Z_{n+1} | Z_n, Z_{n-1}, \ldots] = \phi_1 Z_n + \phi_2 Z_{n-1}$

2-단계 예측식: $f_{n,2} = E[Z_{n+2} | Z_n, Z_{n-1}, \ldots] = \phi_1 f_{n,1} + \phi_2 Z_n$

(예- AR(2) 모형) 시점 1에서 20까지의 시계열데이터가 다음과 같다.
시점 1-10: -1.356, -1.567, -0.994, -0.417, 0.840, -0.991, 0.166, 0.889, 0.514, -0.491
시점 11-20: -0.766, -1.936, -2.223, -1.395, -1.512, -0.582, 1.204, 1.706, -0.768, -0.313

- 과거데이터를 바탕으로 시계열모형은 다음의 AR(2)를 따르는 것으로 추정되었다.

$$Z_t = 0.7Z_{t-1} - 0.2Z_{t-2} + a_t \; (\sigma_a^2 = 1.0)$$

- 시점 10에서부터 한단계이후 예측치를 구하고 다음 시점의 실제값과 비교하여 예측오차를 산출하라.

(풀이) 한단계이후 예측치는 다음과 같다.

$$f_{n,1} = 0.7Z_n - 0.2Z_{n-1}$$
$$f_{11,1} = 0.7Z_{11} - 0.2Z_{10}$$

n	Z_n	예측치	오차
10	-0.491	-	-
11	-0.766	-0.4465	-0.3195
12	-1.936	-0.4380	-1.4980
13	-2.223	-1.2020	-1.0210
14	-1.395	-1.1689	-0.2261
15	-1.512	-0.5319	-0.9801
16	-0.582	-0.7794	0.1974
17	1.204	-0.1050	1.3090
18	1.706	0.9592	0.7468
19	-0.768	0.9534	-1.7214
20	-0.313	-0.8788	0.5658

→ 11,1의 의미: 11인 시점에서 한 단계 이후, 즉 n=12일 때의 예측치를 구하는 것

- MA 모형의 예측: 현재 시계열 데이터는 과거의 잔차들의 가중 평균으로 구성

- MA(2) 모형: $Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

- 한단계 예측식: $f_{n,1} = E[Z_{n+1} | Z_n, Z_{n-1}, \ldots] = -\theta_1 a_n - \theta_2 a_{n-1}$ → 과거의 백색잡음
- 2-단계 예측식: $f_{n,2} = E[Z_{n+2} | Z_n, Z_{n-1}, \ldots] = -\theta_2 a_n$

→ AR 모형과 달리, 과거의 백색잡음 값을 통하여 예측함

→ 백색잡음 역시, 정상성이 높고 평균 회귀 특성들이 있으므로 MA 모형도 평균회귀의 특성 지님

cf) AR 모형과 MA 모형의 가역성

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

- **AR(p) 모형을 MA(∞)으로 표현 가능**

Example: AR(1)

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &\cdots \end{aligned}$$

- **제약조건 (가역성: MA(q)을 AR(∞)로 표현하기 위한 조건, q=1 인 경우)**

Provided $-1 < \phi_1 < 1$:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots$$

- 전개 과정:

MA(1) 모형의 전방예측

$$MA(1): X_n = e_n + b_0 e_{n-1}$$

$$\begin{aligned} Forecast: \hat{X}_{n+1} &= E[X_{n+1} | X_n] \\ &= E[e_{n+1} + b_0 e_n | X_n] \\ &= E[e_{n+1} | X_n] + E[b_0 e_n | X_n] \\ &= b_0 e_n \longleftarrow \text{1단계 전방 예측값} \end{aligned}$$

$$\begin{aligned} \hat{X}_{n+2} &= E[X_{n+2} | X_n] \\ &= E[e_{n+2} + b_0 e_{n+1} | X_n] \\ &= E[e_{n+2} | X_n] + E[b_0 e_{n+1} | X_n] \\ &= 0 \longleftarrow \text{2단계 전방 예측값은 존재하지 않음} \end{aligned}$$

MA(2) 모형의 전방예측

$$MA(2): X_n = e_n + b_0 e_{n-1} + b_1 e_{n-2}$$

$$\begin{aligned} Forecast: \hat{X}_{n+1} &= E[X_{n+1} | X_n] \\ &= E[e_{n+1} + b_0 e_n + b_1 e_{n-1} | X_n] \\ &= E[e_{n+1} | X_n] + E[b_0 e_n | X_n] + E[b_1 e_{n-1} | X_n] \\ &= b_0 e_n + b_1 e_{n-1} \longleftarrow \text{1단계 전방 예측값} \end{aligned}$$

$$\begin{aligned} \hat{X}_{n+2} &= E[X_{n+2} | X_n] \\ &= E[e_{n+2} + b_0 e_{n+1} + b_1 e_n | X_n] \\ &= E[e_{n+2} | X_n] + E[b_0 e_{n+1} | X_n] + E[b_1 e_n | X_n] \\ &= b_1 e_n \longleftarrow \text{2단계 전방 예측값} \end{aligned}$$

- AR/MA 모형 모두 현재 시점에서 먼 값을 예측할수록 예측 오차가 커짐