

## ○ ARIMA, AR, MA, ACF, PACF, 정상성에 대한 정리

### 1. 시계열은 반드시 고정된(fixed) 시간 구간의 관측치여야 한다

즉, 시계열이 불규칙적인 시간 구간이어서는 안된다는 것이다. 어떤 구간은 Daily 구간이었다가, 어떤 구간은 Monthly, 또 어떤 구간은 Yearly... 이렇게 다양한 구간이 동시에 존재해서는 안된다는 것

### 2. 정상성 : 평균, 분산이 시간에 따라 일정한 성질. 즉, 시계열 데이터의 특성이 시간의 흐름에 따라 변하지 않음을 의미

ex) 추세나 계절성이 있는 시계열은 정상 시계열이 아님. 이렇게 정상성을 나타내지 않는 데이터는 복잡한 패턴을 모델링하여 분석하기 어렵기때문에 정상성을 갖도록 로그변환, 차분 등 전처리 후 분석을 시행한다.

1) 정상시계열(Stationary) : 평균과 표준편차가 일정하다는 조건이 선행되어야 분석이 가능하다.

2) 비정상시계열 : 차분이나 log함수를 씌워 정상시계열로 변환 후 분석을 해야 한다.

### 3. 정상 시계열 변환 : 정상성을 나타내지 않는 데이터를 정상 시계열로 변환하는 방법은 다음과 같다.

1) 변동폭이 일정하지 않은 경우 : 로그 변환

2) 추세, 계절성이 존재하는 경우 : \* 차분(differencing)

\* 단, 1차 차분으로 정상성을 띄지 않으면, 차분을 반복한다.

### 4. 시계열 예측 모델

1) AR(Auto Correlation)모형 : 예측하고자 하는 특정 변수의 과거 관측값의 선형결합으로 해당 변수의 미래값을 예측하는 모형

AR(p) 모형의 식은 다음과 같다.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

$y_t$ 는  $t$ 시점의 관측값,  $c$ 는 상수,  $\phi$ 는 가중치,  $\varepsilon_t$ 는 오차항을 의미한다.

2) MA(Moving Average)모형 : 예측 오차를 이용하여 미래를 예측하는 모형

MA(q) 모형의 식은 다음과 같다.

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

3) ARIMA모형 : d차 차분한 데이터에 위 AR(p)모형과 MA(q)모형을 합친 모형

ARIMA(p,d,q) 모형은 d차 차분한 데이터에 위 AR(p) 모형과 MA(q) 모형을 합친 모형으로, 식은 다음과 같다.

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

4) ACF(자기상관함수) 및 PACF(편자기상관함수) : 정상성을 판단하는 함수

- 추세를 띠는 경우 정상성을 띄지 않는 것임

- 자세한 내용 : 링크 참조

<출처>

<https://leedakyeong.tistory.com/entry/ARIMA%EB%9E%80-ARIMA-%EB%B6%84%EC%84%9D%EA%B8%B0%EB%B2%95-AR-MA-ACF-PACF-%EC%A0%95%EC%83%81%EC%84%B1%EC%9D%B4%EB%9E%80>