**시계열분석(ARIMA)**

과거 데이터의 패턴을 분석하여 미래의 값을 예측하는 방법으로, 과거의 패턴이 미래에도 지속된다는 데이터의 안정성이 기본적인 가정으로 필요하다.

**시계열 분해법**

시계열 분해법이란 시계열 데이터를 추세/순환/계절/불규칙 요소로 분해하는 기법이다.

* 추세요소 : 데이터가 장기적으로 증가하거나 감소하는 것
* 순환요소 : 일정 주기가 없으며 장기적인 변환 현상
* 계절성요소 : 주, 월, 분기, 반기 등 특정 시간의 주기로 나타나는 패턴
* 불규칙요소 : 예측 불가능한 임이의 변동

덧셈 분해 : y = s + t +r

곱셈 분해 : y = s\*t\*r

덧셈 분해와 곱셈 분해의 차이점은 덧셈은 Trend와 Seasonal이 별개이고, 곱셈은 Trend에 따라 Seansonal이 변화한다.

**ARIMA(Autoregressive Intergrated Moving Average)**

**정상성(Stationary)**

ARIMA는 AR과 MA 모형을 합친 모형으로 시계열 데이터의 정상성을 가정하고 있다.

여기서 정상성이란 평균, 분산이 시간에 따라 일정한 성질을 말하는데, 예로 추세나 계절성이 있는 시계열은 정상 시계열이 아니다.

따라서 정상성을 나타내지 않는 데이터를 정상 시계열로 변환하는 방법은 아래와 같다.

1. 변동폭이 일정하지 않은 경우 : 로그변환 np.log1p(x)
2. 추세, 계절성이 존재하는 경우 : 차분 data.diff(n)

**모형 설명 (AR, MA, ARIMA)**

AR 모형이란 : 자기 회귀 모형으로 자기상관성을 시계열 모형으로 구성하였으며, 예측하고자 하는 특정 변수의 과거 관측값의 선형결합으로 해당 변수의 미래값을 예측하는 모형이다.

MA 모형이란 : 예측 오차를 이용하여 미래를 예측하는 모형이다.

ARIMA 모형이란 : ARIMA(p,d,q) 모형은 d차 차분한 데이터에 AR(p), MA(q) 이 두 모형을 합친 것이다.

**P와 q는 각각 PACF 그래프와 ACF 그래프를 그려서 확인한다.**

**모형 선택 지표 (ACF, PACF)**

* ACF(자기상관함수, AutoCorrelation Function)이란:

시차에 따른 일련의 자기상관을 의미하며, 시차가 커질수록 ACF는 0에 가까워진다.

정상 시계열은 상대적으로 빠르게 0에 수렴하며, 비정상 시계열은 천천히 감소하고, 종종 큰 양의 값을 가진다.

* PACF(편자기상관함수, Partial AutoCorrelation Function)

시차에 따른 일련의 편자기상관이며, 시차가 다른 두 시계열 데이터 간의 순수한 상호 연관성이다.

모델링을 마친 후 잔차의€ ACF 그래프를 그려, 정상성을 따르는지 확인해야 한다. 만약, 정상성을 따르지 않는다면, p, d, q의 파라미터를 재조정해야한다.