

18기 정규세션

ToBig's 17기강의자
신은아

Time Series Analysis

시계열 분석

Contents

Unit 01 | 시계열 데이터

Unit 02 | 고전적 분석 - 평활법 & 분해법

Unit 03 | 확률적 분석 - AR, MA, ARMA

Unit 04 | Box-Jenkins 방법론

Contents

Unit 01 | 시계열 데이터

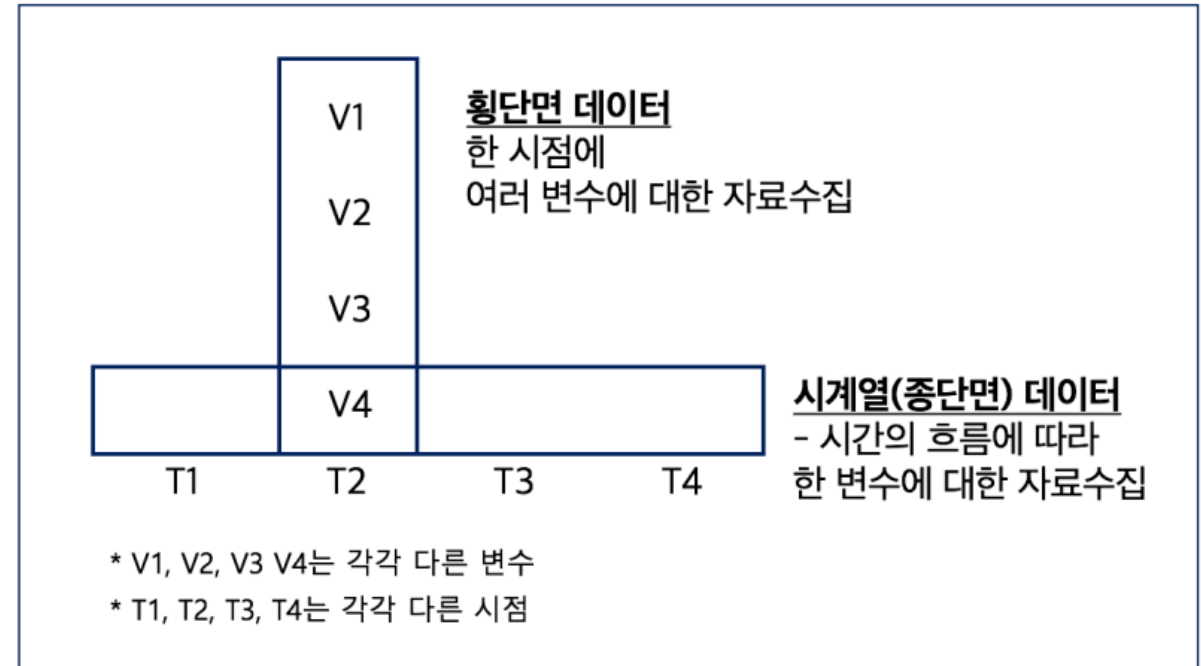
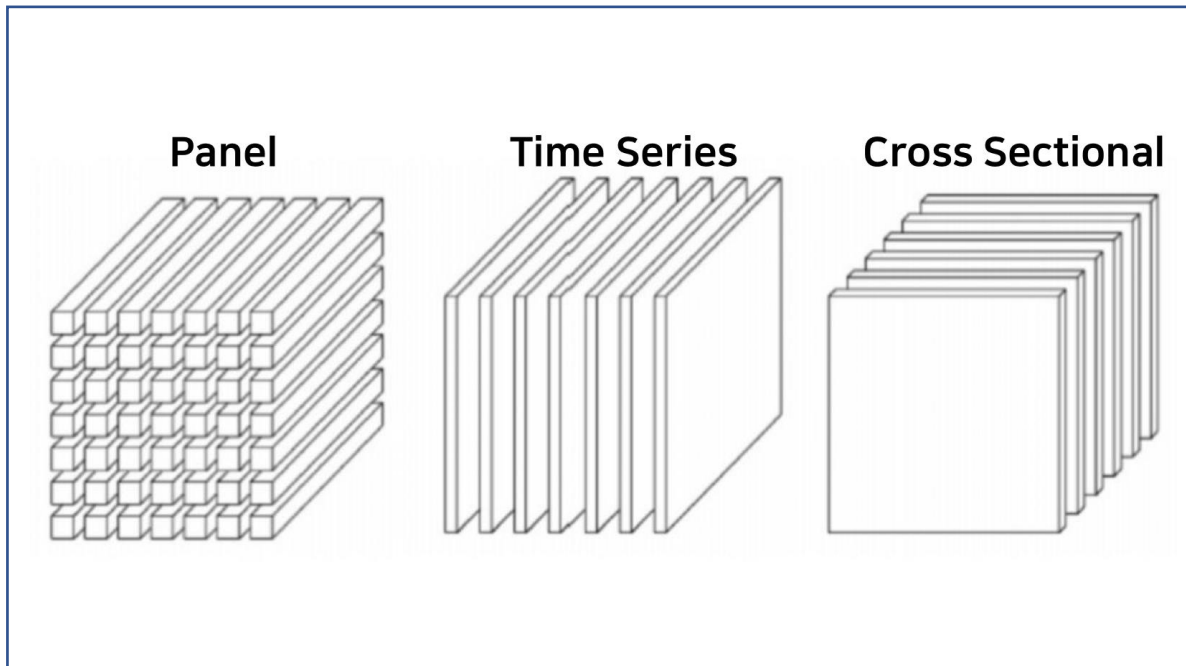
Unit 02 | 고전적 분석 - 평활법 & 분해법

Unit 03 | 확률적 분석 - AR, MA, ARMA

Unit 04 | Box-Jenkins 방법론

Unit 01 | 시계열 자료

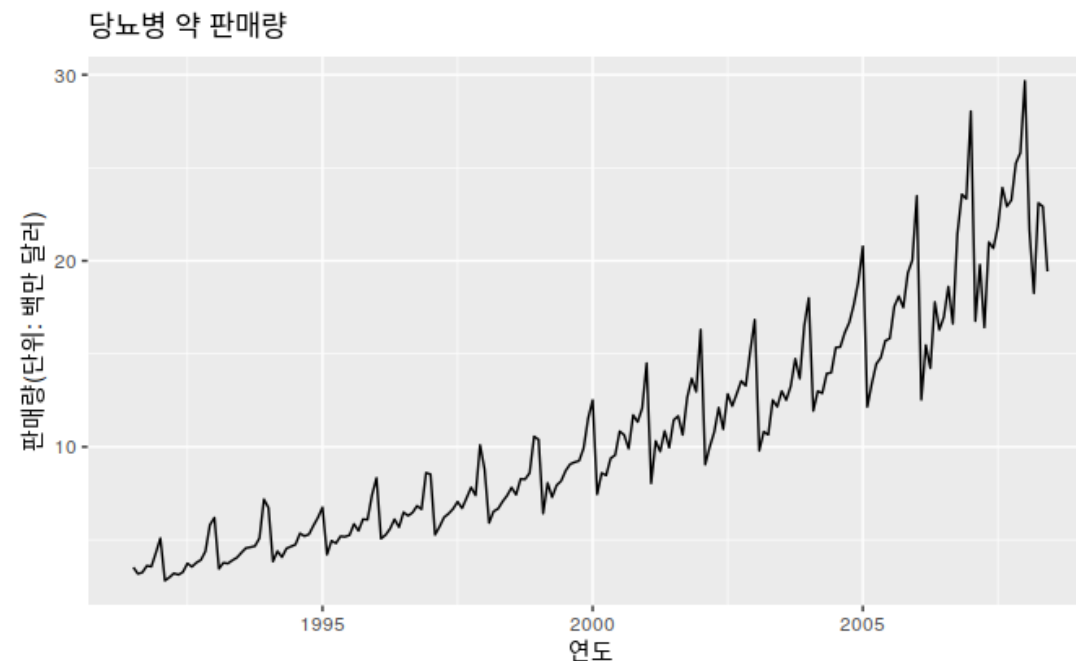
패널, 횡단면, 시계열



Unit 01 | 시계열 자료

시계열 자료 (Time Series Data)

- 시간의 흐름에 따라 관측된 자료
- Ex. 국민 총생산, 물가지수, 상품의 판매량, 종합주가지수, 강우량 등
- 일반적으로 시간 t 에 대해 Z_t 와 같이 표현함
- 시계열 자료 분석 목적
 - 미래에 대한 예측
 - 시스템 또는 확률과정에 대한 이해와 제어



- 시계열 분석 시 시계열그림(time series plot)을 그려보는 것이 가장 선행되어야 함.

Unit 01 | 목차1

시계열 자료의 형태

- 전통적으로 시계열 자료를 분석하기 위해 시계열이 여러 가지 성분으로 구성되어 있다고 보았다.

추세 성분

시간이 경과함에 따라
관측값이 지속적으로
증가하거나 감소하는
추세를 갖는 경우

계절 성분

주별, 월별, 계절별과 같은
주기적인 성분에 의한
변동

Ex. 강우량, 기온, 아이스
크림

순환 성분

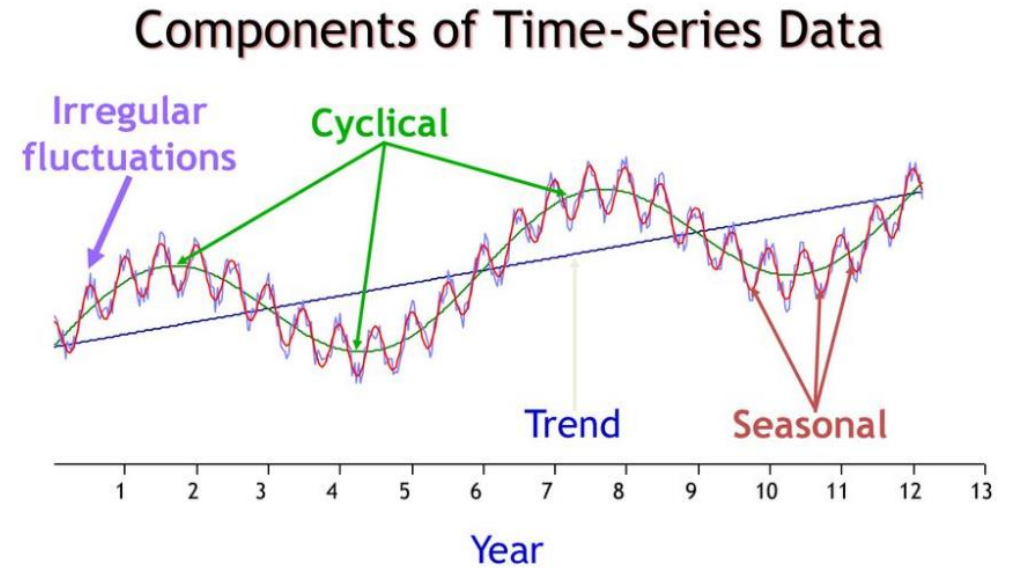
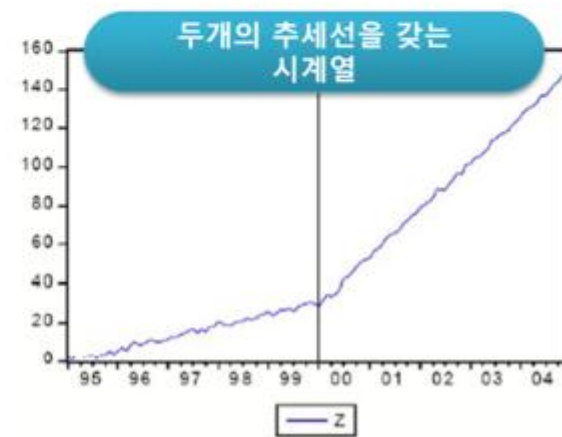
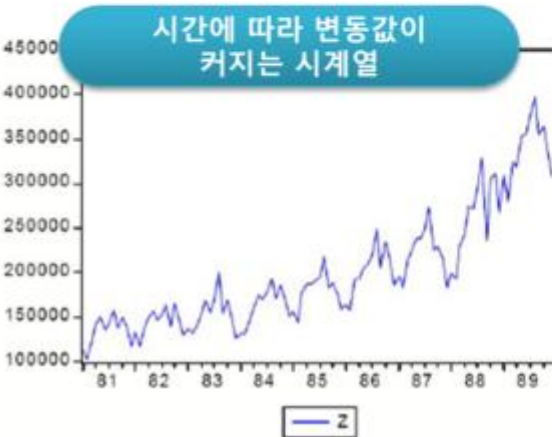
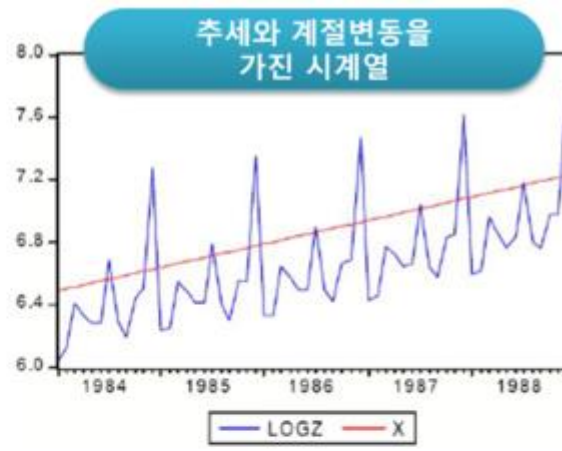
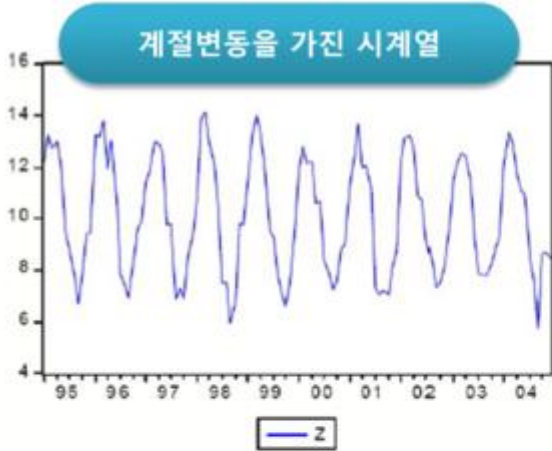
주기가 긴 경우의 변동

Ex. 태양 흑점 수, 남극 빙
하의 변화

불규칙 성분

시간에 따른
규칙적 움직임에 무관하게
랜덤한 원인에 의해
나타나는 변동성

Unit 01 | 시계열 자료



Contents

Unit 01 | 시계열 데이터

Unit 02 | 고전적 분석 - 평활법 & 분해법

Unit 03 | 확률적 분석 - AR, MA, ARMA

Unit 04 | Box-Jenkins 방법론

Unit 02 | 고전적 분석 - 평활법 & 분해법

평활법 (Smoothing Method)

- 평활법 : 과거 및 현재 데이터의 불규칙 변동을 부드럽게 평활(smoothing)하여 미래 값 예측
- 평활법의 종류
 - 목측법(Eye-measurement) : 눈짐작으로 선 긋기. 정확성은 낮으나 신속한 경향 파악 가능.
 - 이동평균법(Moving Average Method) : 최근 특정 기간의 관측치 몇 개를 산술평균한 값을 이용해 미래 시점의 예측치로 사용. 계절성분 또는 불규칙성부를 제거하고 전반적인 추세를 파악하게 해줌.
 - 지수평활법(Exponential Smoothing) : 모든 데이터를 활용하되, 최근 관측치에 더 큰 가중치를 준 가중평균 값을 미래 시점의 예측치로 사용
 - 최소제곱법 : 편차의 제곱합을 최소로 하는 경향선을 구해 미래 값 예측 (회귀분석)

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - 단순지수평활법

- 지수평활법 : 모든 데이터 활용, 최근 관측치에 더 큰 가중치를 준 가중평균 값을 미래 시점의 예측치로 사용
- 시계열 구조 변화에 유연하게 대응 가능, 예측의 갱신이 쉬우며 직관적이고 사용이 용이함
- **단순 지수평활법** : 추세나 계절성이 없는, 수평적 패턴의 데이터에 유용
 - 시계열 $\{Z_n\}$ 의 시점 n 에서의 지수평활치
: $S_n = \alpha Z_n + \alpha(1 - \alpha)Z_{n-1} + \alpha(1 - \alpha)^2 Z_{n-2} + \dots$
 - 시계열 $\{Z_n\}$ 의 시점 $n + 1$ 에서의 지수평활치 :
: $S_{n+1} = \alpha Z_{n+1} + (1 - \alpha)S_n$

Notation

 Z_n : 시점 n 에서의 관측값 α : 지수평활상수 (Smoothing Constant)

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - 단순지수평활법

• 초기 평활값 S_0 선택

- Brown(1962), Montgomery & Johnson(1976) : $\bar{Z} = \frac{1}{n} \sum_{t=1}^n Z_t$ ($n = 6$ or $n = \frac{T}{2}$)
- Makridakis & Wheelwright(1978) : Z_1

• 지수평활상수 α 선택

- Brown(1962) : 0.05와 0.3 사이의 값
- Montgomery(1976) : $1 - 0.8^{1/trend}$
- MSE를 최소로 하는 α

*이 때 trend는 평활의 차수. 단순지수평활의 경우 trend = 1

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - 단순지수평활법

지수평활상수 α

- α 의 값이 작으면 평활의 효과가 커 S_n 은 지엽적인 변화에 둔감하게 반응
- α 의 값이 크면 평활의 효과가 작아 S_n 은 최근의 관측값에 의해 크게 영향을 받아 시계열의 지엽적인 변화에 민감하게 반응

⇒ 시계열의 수준변화가 큰 경우 : $\alpha \approx 1$ 초기평활값 Z_1 선택

시계열의 수준변화가 완만할 때 : $\alpha \approx 0$ 표본평균 \bar{Z} 선택

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - 이중지수평활법

- 선형 추세가 있는 시계열에 단순 지수평활법을 적용하면, 추세를 늦게 따라가는 문제 발생
- **이중 지수평활법** : 지수평활을 두 번 적용, 시계열이 선형 추세가 있는 경우에 적합
 - 지수평활값과 이중지수평활값:
 - : $SM_n = \alpha Z_n + (1 - \alpha)SM_{n-1}$ (지수평활값)
 - : $SM'_n = \alpha SM_n + (1 - \alpha)SM'_{n-1}$ (이중지수평활값)

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - 이중지수평활법

- 시점 t 에서의 한 단계 이후 $t + 1$ 시점의 예측치

$$F_{t+1} = E[Z_{t+1} | Z_t, Z_{t-1}, \dots, Z_1] = a + b(t + 1)$$

$$\widehat{F}_{t+1} = \hat{a} + \hat{b}(t + 1) = (2SM_t - SM'_t) + 1\hat{b}, \quad \hat{b} = \frac{\alpha}{1-\alpha} (SM_t - SM'_t)$$

- 시점 t 에서의 l 단계 이후 예측치

$$\widehat{F}_{t+l} = \hat{a} + \hat{b}(t + l) = (2SM_t - SM'_t) + \hat{b}l$$

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - 이중지수평활법

- 지수평활상수 $\alpha = 0.2$, 초기값 $SM_0 = Z_1 = 10$ 일 때, 이중 지수평활법 기반 예측값

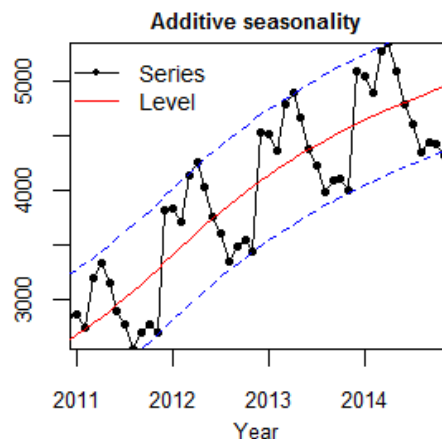
표 3.8 | 가상자료에 대한 브라운의 선형지수평활예측값

t	원계열(Z_t)	지수평활값(SM_t)	이중지수평활값(SM_t)	예측값(F_t)
1	10	10,000	10,000	$F_{1+1} = 2 \times 10 - 10 + \frac{0.2}{0.8} \times (10 - 10) = 10$
2	18	11,600	10,320	$F_{2+1} = 2 \times 11.60 - 10.32 + \frac{0.2}{0.8} \times (11.60 - 10.32) = 13.2$
3	20	13,280	10,912	$F_{3+1} = 2 \times 13.28 - 10.912 + \frac{0.2}{0.8} \times (13.28 - 10.912) = 16.24$
4	28	16,224	11,974	$F_{4+1} = 2 \times 16.224 - 11.794 + \frac{0.2}{0.8} \times (16.224 - 11.794) = 21.536$
5	30	18,379	13,255	$F_{5+1} = 2 \times 18.379 - 13.255 + \frac{0.2}{0.8} \times (18.379 - 13.255) = 24.784$
6	38	22,303	15,065	$F_{6+1} = 2 \times 22.303 - 15.065 + \frac{0.2}{0.8} \times (22.303 - 15.065) = 31.354$
7	37	25,242	17,101	$F_{7+1} = 2 \times 25.242 - 17.101 + \frac{0.2}{0.8} \times (25.242 - 17.101) = 38.183$

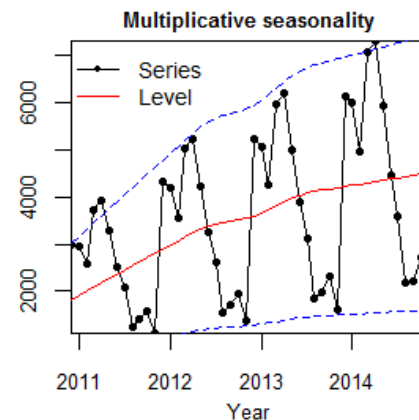
Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - Holt - Winters 계절지수평활법

- 월별, 분기별 자료 : **계절변동**(추세+계절성+수준(자료변동)) 존재
- **계절변동의 두 가지 형태**
 - 가법적 계절변동 : 시간에 따른 계절적 진폭의 크기 일정
 - 승법적 계절변동 : 계절적 진폭의 크기에 평균 수준에 비례하여 점차적으로 증가 or 감소



가법적 계절변동



승법적 계절변동

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - Holt - Winters 계절지수평활법

- **가법적 계절 지수평활법** : 가법적 계절변동이 있을 때, 가법적으로 각 성분(변동) 갱신

$$Z_{n+l} = T_{n+l} + S_{n+l} + I_{n+l}$$

T_{n+l} : 추세성분

S_{n+l} : 계절주기 s 를 가지는 계절성분

I_{n+l} : 오차항으로서 불규칙성분

- 각 성분들을 평활법에 의해 추정해 갱신

계절성분

$$S_i = S_{i+s} = S_{i+2s} = \cdots, \quad \sum_{i=1}^s S_i = 0$$

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - Holt - Winters 계절지수평활법

- 주기가 p 인 시계열 $\{y_t\}$ 가 있다고 할 때, 가법적 계절지수평활법의 갱신 알고리즘은 아래와 같다.

$$\hat{y}_{t+h|t} = \hat{\mu}_t + h \cdot b_t + s_{t+h-p}, \quad h \leq p$$

(p : 계절성의 주기)

여기서 μ_t, b_t, s_t 는 시점 t 에서 추정된 수준, 기울기, 계절효과를 의미한다.

$$\begin{aligned} \cdot \hat{\mu}_t &= \alpha(y_t - s_{t-p}) + (1 - \alpha)(\hat{\mu}_{t-1} + b_{t-1}) \\ \cdot b_t &= \beta(\hat{\mu}_t - \hat{\mu}_{t-1}) + (1 - \beta)b_{t-1} \\ \cdot s_t &= \gamma(y_t - \hat{\mu}_t) + (1 - \gamma)s_{t-p} \\ \cdot \alpha, \beta, \gamma &: \text{전체, 추세, 계절조정을 위한 평활모수} \end{aligned}$$

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - Holt - Winters 계절지수평활법

- **승법적 계절 지수평활법** : 승법적 계절변동이 있을 때, 승법적으로 각 성분(변동) 갱신

$$Z_{n+l} = T_{n+l} \cdot S_{n+l} + I_{n+l}$$

T_{n+l} : 추세성분

S_{n+l} : 계절주기 s 를 가지는 계절성분

I_{n+l} : 오차항으로서 불규칙성분

- 각 성분들을 평활법에 의해 추정해 갱신

계절성분

$$S_i = S_{i+s} = S_{i+2s} = \cdots, \quad \sum_{i=1}^s S_i = 0$$

Unit 02 | 고전적 분석 - 평활법 & 분해법

지수평활법 - Holt - Winters 계절지수평활법

- 주기가 p 인 시계열 $\{y_t\}$ 가 있다고 할 때, 승법적 계절지수평활법의 갱신 알고리즘은 아래와 같다.

$$\hat{y}_{t+h|t} = (\hat{\mu}_t + h \cdot b_t) \times s_{t+h-p}, \quad h \leq p$$

(p : 계절성의 주기)

여기서 μ_t, b_t, s_t 의 의미는 위와 같으나, 수준과 계절효과의 세부적인 식이 다르다.

$$\cdot \hat{\mu}_t = \alpha \left(\frac{y_t}{s_{t-p}} \right) + (1 - \alpha)(\hat{\mu}_{t-1} + b_{t-1})$$

$$\cdot b_t = \beta(\hat{\mu}_t - \hat{\mu}_{t-1}) + (1 - \beta)b_{t-1}$$

$$\cdot s_t = \gamma \left(\frac{y_t}{\hat{\mu}_t} \right) + (1 - \gamma)s_{t-p}$$

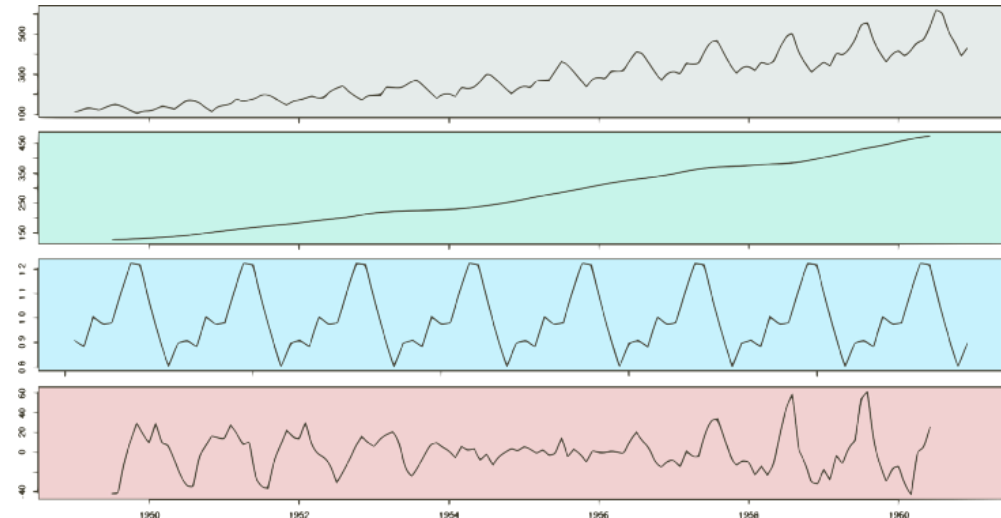
· α, β, γ : 전체, 추세, 계절조정을 위한 평활모수

출처 : <https://be-favorite.tistory.com/62>

Unit 02 | 고전적 분석 - 평활법 & 분해법

분해법 (Decomposition Method)

- 20세기 초에 경제학자들이 경기변동을 예측하기 위해 시도한데서 비롯된 전통적인 방법
- 시계열을 **개별 변동성분(추세, 계절성, 순환성)**으로 각각 분해한 후, 분해된 각 성분을 개별적으로 예측
→ 이후 이를 다시 결합하여 예측
- 성분들은 결정적(deterministic)이고 서로 독립적이라고 가정
- 목적 : 예측(forecasting), **계절조정(seasonal adjustment)**
- 가법모형(additive model) : $Z_t = T_t + S_t + C_t + I_t$
- 승법모형(multiplicative model) : $Z_t = T_t \cdot S_t \cdot C_t \cdot I_t$



Unit 02 | 고전적 분석 - 평활법 & 분해법

이동평균법 - 단순이동평균법

- 이동평균법 : 최근 특정 기간의 관측치 몇 개를 산술평균한 값을 이용해 미래 시점의 예측치로 사용
- **단순 이동평균법** : 가장 최근에 얻어진 관측값 m 개의 산술평균값을 이용
 - 시계열 $\{Z_n\}$ 에서, 시점 t 에서의 관측값을 Z_t 라 할 때, 시점 n 에서의 단순 이동평균

$$: MA_n = \frac{Z_n + Z_{n-1} + \cdots + Z_{n-m+1}}{m} = \frac{1}{m} \sum_{t=n-m+1}^n Z_t$$

- 시점 n 에서 추정한 시점 $n+1$ 의 예측값을 F_{n+1} 이라고 하면, $F_{n+1} = MA_n$

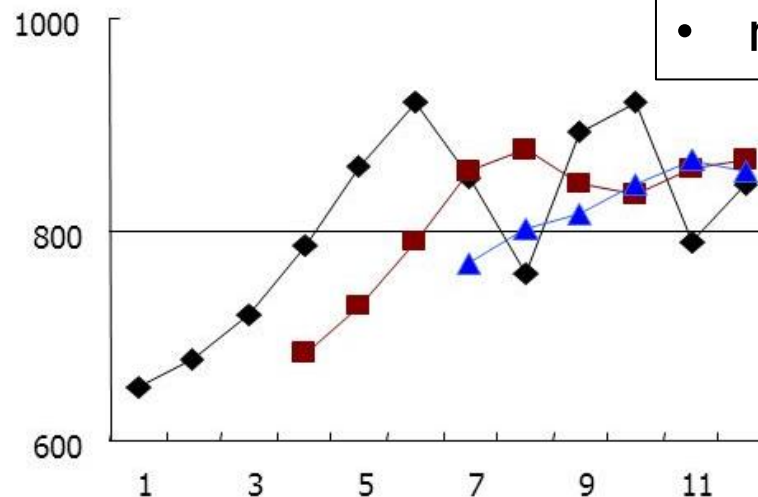
Unit 02 | 고전적 분석 - 평활법 & 분해법

이동평균법 - 단순이동평균법

Week	Demand	3-week MA	6-week MA
1	650		
2	678		
3	720		
4	785	683	
5	859	728	
6	920	788	
7	850	855	769
8	758	876	802
9	892	843	815
10	920	833	844
11	789	857	867
12	844	867	855

$$= (650 + 678 + 720) / 3$$

$$= (650 + 678 + 720 + 785 + 859 + 920) / 6$$



- 데이터 변동이 클수록, 최근 데이터의 경향을 많이 반영해야 함
→ **m이 작아야 함**
- 데이터의 변동이 완만하다면,
→ m이 커도 됨
- m의 선택 : MSE를 최소로 하는 m

Unit 02 | 고전적 분석 - 평활법 & 분해법

이동평균법 - 이중(선형)이동평균법

- 선형 추세가 있는 시계열에 단순 이동평균법을 적용하면, 추세를 늦게 따라가는 문제 발생
- **이중(선형) 이동평균법** : 이동평균을 두 번 적용한 방법

- 시계열 $\{Z_n\}$ 에서, 시점 t 에서의 관측값을 Z_t 라 할 때, 시점 n 에서의 **단순 이동평균**

$$: MA_n = \frac{Z_n + Z_{n-1} + \cdots + Z_{n-m+1}}{m} = \frac{1}{m} \sum_{t=n-m+1}^n Z_t$$

- 시계열 $\{Z_n\}$ 에서, 시점 t 에서의 관측값을 Z_t 라 할 때, 시점 n 에서의 **이중(선형) 이동평균**

$$: MA'_n = \frac{1}{m} (MA_n + MA_{n-1} + \cdots + MA_{n-m+1})$$

Unit 02 | 고전적 분석 - 평활법 & 분해법

대칭 이동평균 & 중심화 이동평균

- 시점 t 에서의 관측값을 Z_t 라 하고, 이동평균을 MA_t 라 하면,

- **대칭 이동평균** (m 이 홀수) : 해당 시점과 인접한 전후 시점들의 평균

$$: MA_t = \frac{1}{m} \sum_{j=-(m-1)/2}^{(m-1)/2} Z_{t+j}$$

- **중심화 이동평균** (m 이 짝수) : 짝수 기간의 이동평균은 두 시점 사이의 값에 대응 \rightarrow 재차 이동평균

$$: MA_t = \frac{1}{2} \left(MA_{t-\frac{1}{2}} + MA_{t+\frac{1}{2}} \right)$$

$$MA_{t-1/2} = \frac{1}{m} \sum_{j=-m/2}^{m/2-1} Z_{t+j} \quad MA_{t+1/2} = \frac{1}{m} \sum_{j=-m/2+1}^{m/2} Z_{t+j}$$

Unit 02 | 고전적 분석 - 평활법 & 분해법

대칭 이동평균 & 중심화 이동평균

가상적 자료에 대한 이동평균

출처: 시계열분석과 예측, 이누리 저, 제 2판 p77

기간	관측값 $\{Z_t\}$	$3MA$	관측값 $\{Z_t\}$	$4MA$	$2 \times 4MA$
1	20	—	20	—	—
2	10	20 MA_2	10	—	—
3	30	25 MA_3	30	23.75 $MA_{2.5}$	22.500 MA_3
4	35	25	35	21.25 $MA_{3.5}$	27.500
5	10	35	10	33.75	36.250
6	60	40	60	38.75	39.375
7	50	50	50	40.00	50.000
8	40	60	40	60.00	60.625
9	90	65	90	61.25	61.875
10	65	70	65	62.50	72.500
11	55	80	55	82.50	—
12	120	—	120	—	—

대칭 이동평균

중심화 이동평균

Unit 02 | 고전적 분석 - 평활법 & 분해법

분해법에 의한 예측 절차 (가법 모형 전제)

$$\text{원 계열} - \text{추세} = \text{계절성} + \text{불규칙 성분}$$

1. 추세 및 추세 조정된 계열의 추정

- ✓ 원 계열에 추세선을 적합시켜 추세를 추정 → 이 추세를 원 시계열에서 빼서 추세 조정된 계열 얻기

2. 계절성 및 계절 조정된 계열의 추정

- ✓ 추세 조정된 계열에 대하여 계절성의 길이만큼 이동평균 → 계절성 제거 (불규칙 변동만이 남음)
- ✓ 다시 추세 조정된 계열에서 추정된 불규칙 변동 제거 → 시점별 계절성 추정
- ✓ 이렇게 추정된 계절성별로 계절별 평균을 구하고, 계절평균의 합이 0이 되도록 조정 → 계절지수

3. 예측

- ✓ 분해된 성분들을 개별적으로 예측하여 원 계열의 미래값 예측

Contents

Unit 01 | 시계열 데이터

Unit 02 | 고전적 분석 - 평활법 & 분해법

Unit 03 | 확률적 분석 - AR, MA, ARMA

Unit 04 | Box-Jenkins 방법론

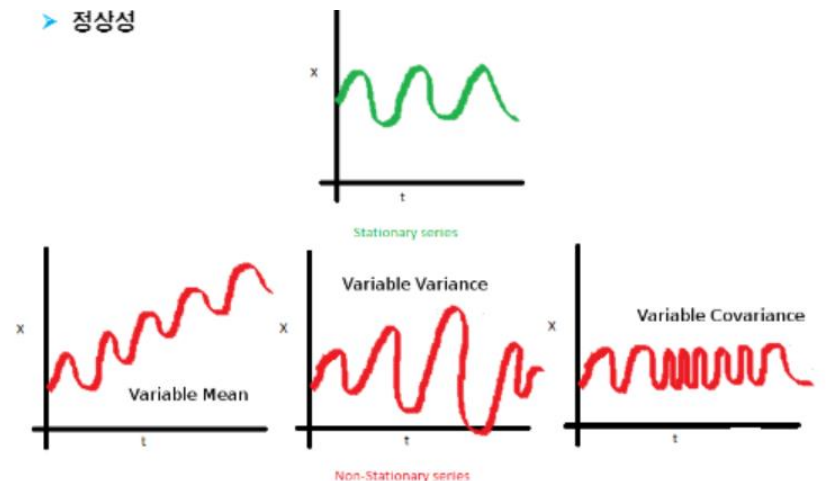
Unit 03 | 확률적 분석 - AR, MA, ARMA

정상 시계열 (Stationary Time Series)

- **정상 시계열** : 시간에 관계없이 평균과 분산이 일정한 시계열
 - 모든 t 에 대해서 $E(Z_t) = \mu, \text{Var}(Z_t) = \sigma^2$
 - 두 시점 사이의 자기공분산은 시점 t 와 무관하며, 시차(time lag) k 에만 의존 \rightarrow 시간 불변성

- **정상 시계열의 형태**

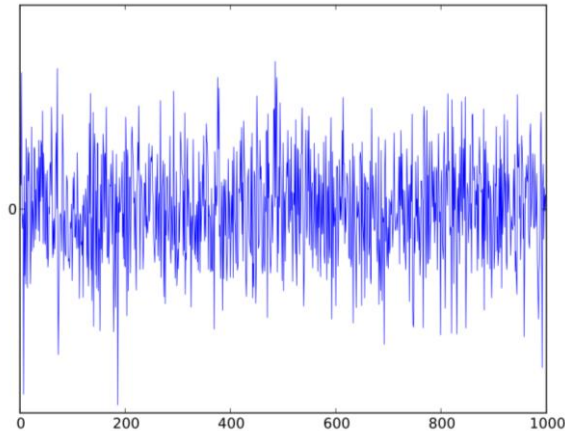
- 추세, 계절성, 순환성 등의 패턴이 보이지 않음
- 자료 변화의 폭이 일정함
- 시간에 따라 달라지는 자기상관적 패턴을 나타내는 구간이 없음



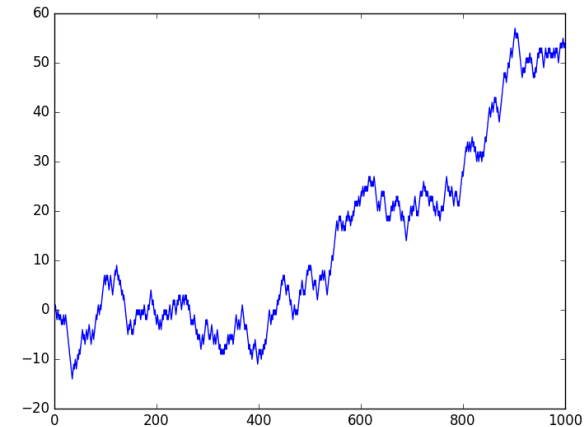
Unit 03 | 확률적 분석 - AR, MA, ARMA

정상 시계열, 비정상 시계열 예시

- **정상시계열** : 백색잡음과정 (White Noise)
 - ✓ 서로 독립이면서 동일한 분포를 따르는(i.i.d.) 확률변수들의 계열로 구성
 - ✓ $\{\varepsilon_t\} : WN(0, \sigma_\varepsilon^2), Z_t = \varepsilon_t$



- **비정상시계열** : 확률보행과정 (Random Walk)
 - ✓ $\{\varepsilon_t\} : WN(0, \sigma_\varepsilon^2), Z_t = Z_{t-1} + \varepsilon_t$
 - ✓ 평균은 0으로 일정하나, 분산이 시간에 종속



Unit 03 | 확률적 분석 - AR, MA, ARMA

정상 시계열 여부 판단

- **판단 기준**
 - ✓ X축을 lag(현재 데이터와의 시점 차)로, Y축을 ACF(Autocorrelation Function)로 시각화
 - ✓ 특정 패턴이 없다면 정상 시계열, 있다면 비정상 시계열
- ✓ **[참고] Autocorrelation Function** (ACF, 자기상관함수)
 - ✓ Correlation(두 변수 간 관계를 -1~1 값으로 표현하는 척도)에 Auto 개념이 추가된 것
 - ✓ 시계열적 관점에서, time shifted된 자기 자신과의 correlation

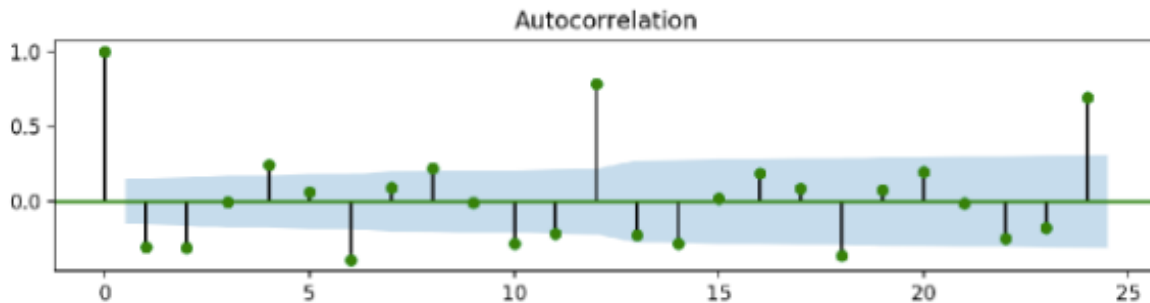
Unit 03 | 확률적 분석 - AR, MA, ARMA

정상 시계열 여부 판단

• ACF (자기상관함수)

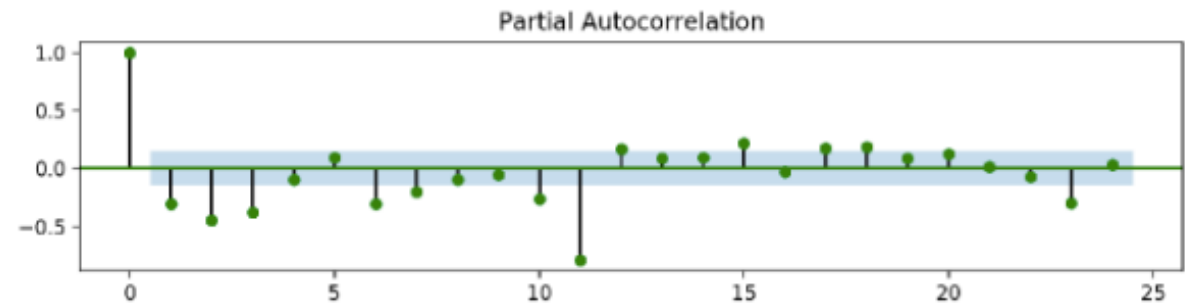
- ✓ 시차 k 에 대해서, 관측치 간 상관계수 함수

$$ACF(k) = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}$$



• PACF (Partial ACF, 편자기상관함수)

- ✓ 시차 k 에 대해서, Z_t 와 Z_{t+k} 사이에 있는 관측치와의 상호 의존성을 제거한 함수
- ✓ 순수한 상관관계 파악 가능



Unit 03 | 확률적 분석 - AR, MA, ARMA

Box-Jenkins 세 가지 모형

- 자기회귀모형 (Auto Regressive) : AR(p) 모형

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t$$

- 이동평균모형 (Moving Average) : MA(q) 모형

$$Z_t - \mu = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$$

- 자기회귀이동평균 모형 (Auto Regressive Moving Average) : ARMA(p, q) 모형

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}$$

Unit 03 | 확률적 분석 - AR, MA, ARMA

자기회귀모형(Auto Regressive) : AR(p)

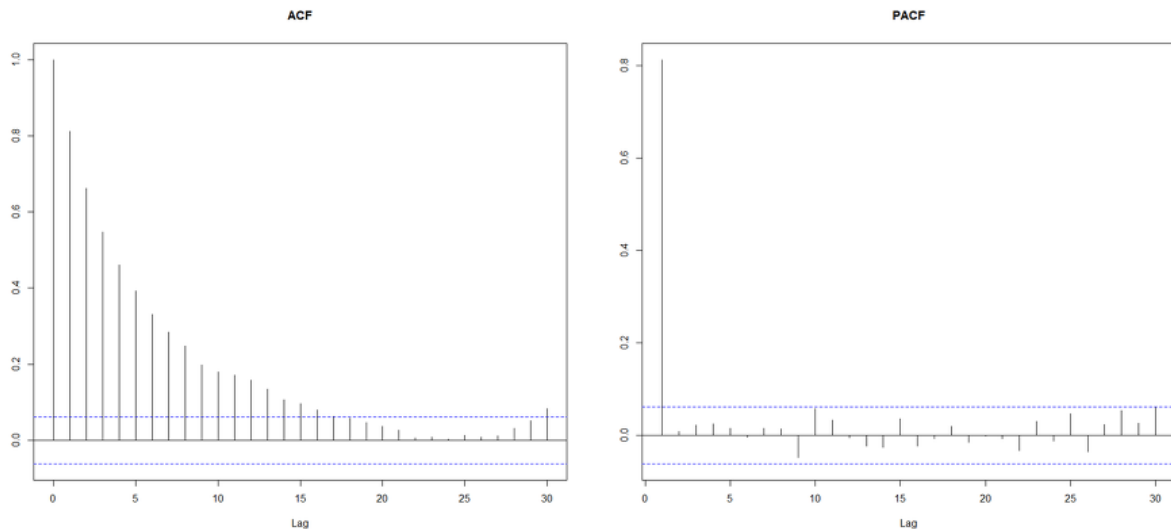
$$\begin{aligned} Z_t - \mu &= \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \cdots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t \\ &= \sum_{j=1}^p \phi_j(Z_{t-j} - \mu) + \varepsilon_t \end{aligned}$$

- 자기 자신을 종속변수로 하고, 그 이전 시점의 시계열을 독립변수로 하는 모형
 - 즉, 시계열 Z_t 를 그 이전 시점의 계열로 회귀
 - p 차 AR 모형은 Z_t 가 $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$ 의 선형결합임을 가정, **항상 가역성 만족**
- p 는 하이퍼파라미터 (ACF, PACF 확인 후 결정)

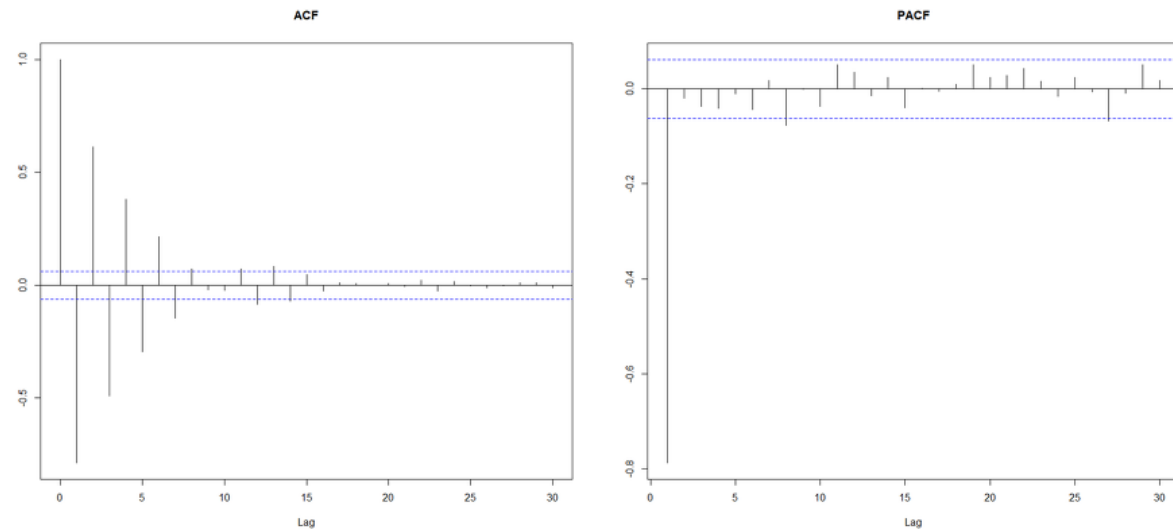
Unit 03 | 확률적 분석 - AR, MA, ARMA

AR(1) 모형의 ACF, PACF

$$\text{AR}(1) : Z_t = \phi_1 Z_{t-1} + \varepsilon_t$$

1. Simulation of AR(1), $\phi > 0$ 

$$\phi > 0$$

2. Simulation of AR(1), $\phi < 0$ 

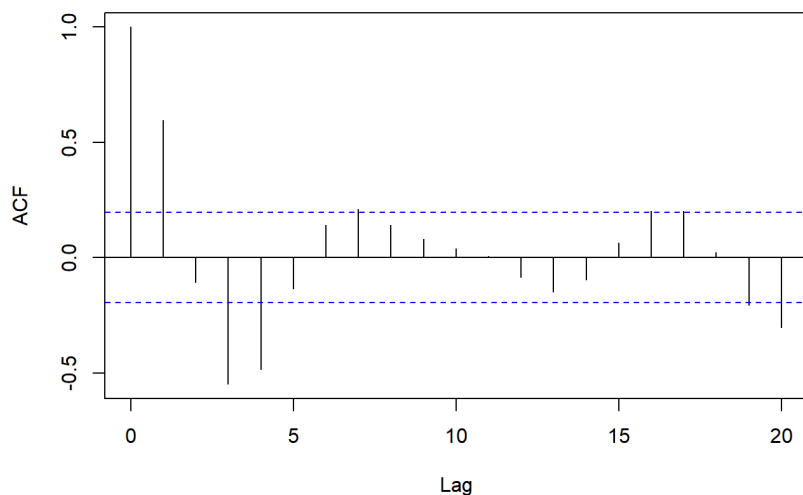
$$\phi < 0$$

Unit 03 | 확률적 분석 - AR, MA, ARMA

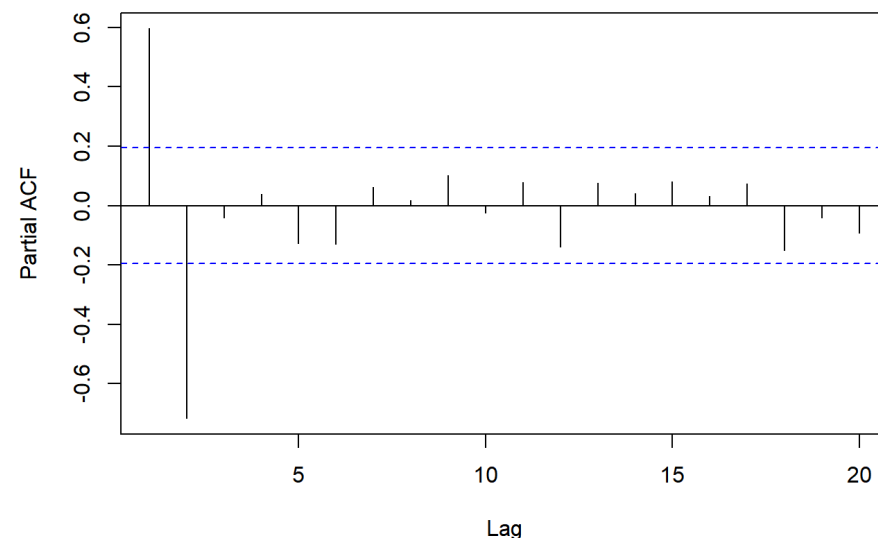
AR(2) 모형의 ACF, PACF

$$\text{AR}(2) : \dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \varepsilon_t$$

AR(2) phi1=0.9, phi2=-0.6 ACF



AR(2) phi1=0.9, phi2=-0.6 PACF



⇒ AR(p) 과정의 이론적인 ACF와 PACF의 일반적인 형태

- ACF : **지수함수의 형태로** 또는 싸인함수와 같은 곡선의 형태를 가지며 **점차 줄어듦**
- PACF : AR(p) 모형의 차수인 시차 p까지는 0이 아니며 **시차 p 이후에는 0이 됨**

Unit 03 | 확률적 분석 - AR, MA, ARMA

이동평균모형(Moving Average) : MA(q)

$$Z_t - \mu = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

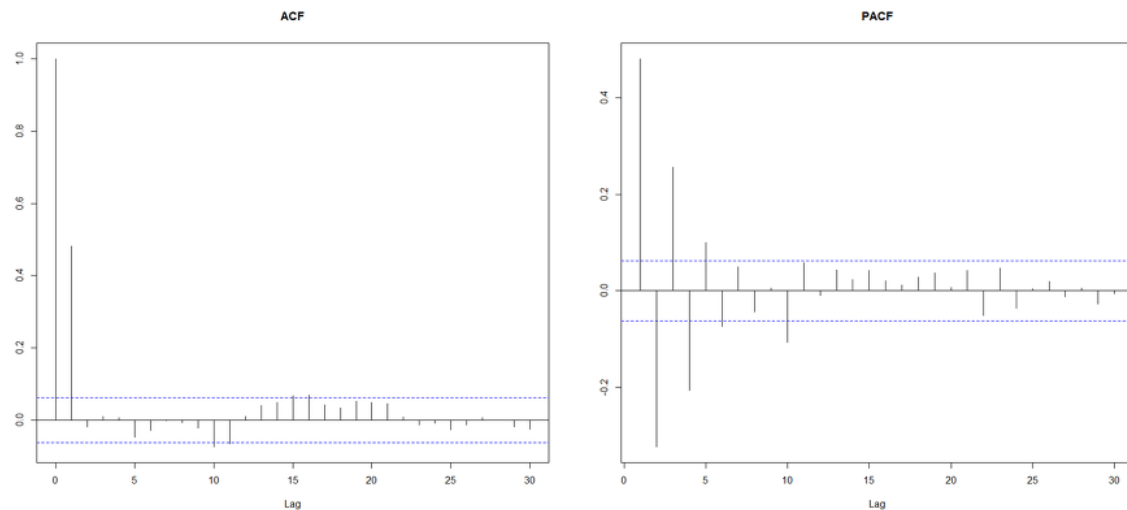
- 자기 자신을 종속변수로 하고, 그 이전 시점의 white noise process를 독립변수로 하는 모형
 - 즉, 시계열 Z_t 를 그 이전 시점의 백색잡음 계열 $\{a_t, a_{t-1}, a_{t-2}, \dots\}$ 로 회귀
 - q차 MA 모형은 **항상 정상성 만족**
- q는 하이퍼파라미터 (ACF, PACF 확인 후 결정)

Unit 03 | 확률적 분석 - AR, MA, ARMA

MA(1) 모형의 ACF, PACF

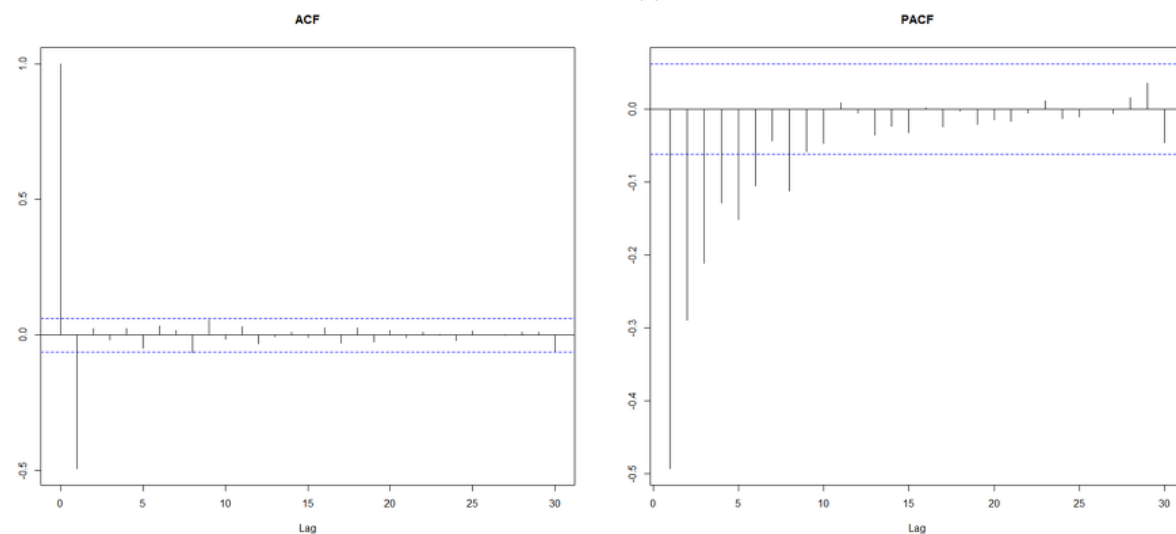
$$\text{MA}(1) : \dot{Z}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

3. Simulation of MA(1), theta>0



$$\theta > 0$$

4. Simulation of MA(1), theta<0

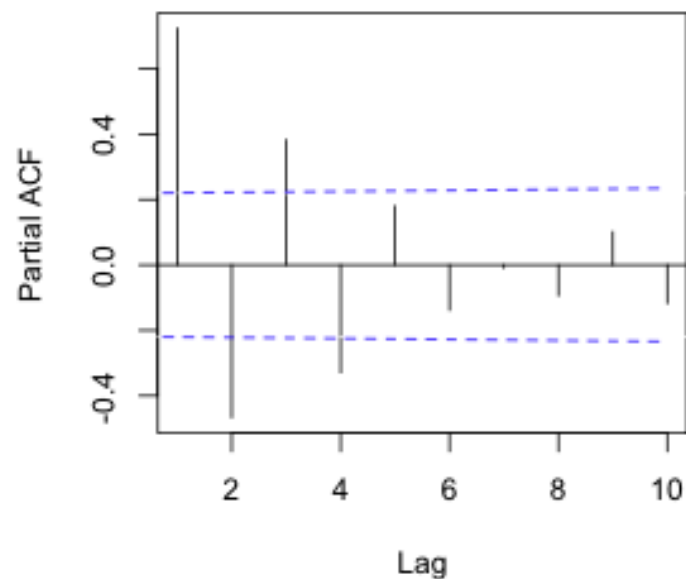
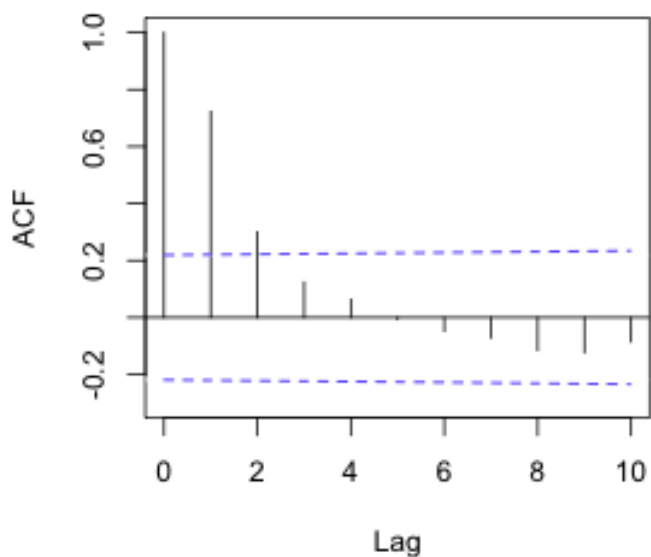


$$\theta < 0$$

Unit 03 | 확률적 분석 - AR, MA, ARMA

MA(2) 모형의 ACF, PACF

$$\text{MA}(2) : \dot{Z}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$



⇒ MA(q) 과정의 이론적인 ACF와 PACF의 일반적인 형태

- ACF : MA(q) 모형의 차수인 시차 q까지는 0이 아니며 **시차 q 이후에는 0이 됨**
- PACF : **지수함수의 형태로** 또는 싸인함수와 같은 곡선의 형태를 가지며 **점차 줄어듦**

Unit 03 | 확률적 분석 - AR, MA, ARMA

AR과정과 MA과정의 쌍대성(duality)

- AR(p), MA(q) 모형의 Theoretical ACF, PACF 특징

	AR(p)	MA(q)
ACF	Die out (지수적 감소, 소멸하는 sine 함수 형태)	Cut off after lag q (q+1 시점부터 0으로 절단)
PACF	Cut off after lag p (p+1 시점부터 0으로 절단)	Die out (지수적 감소, 소멸하는 sine 함수 형태)

Unit 03 | 확률적 분석 - AR, MA, ARMA

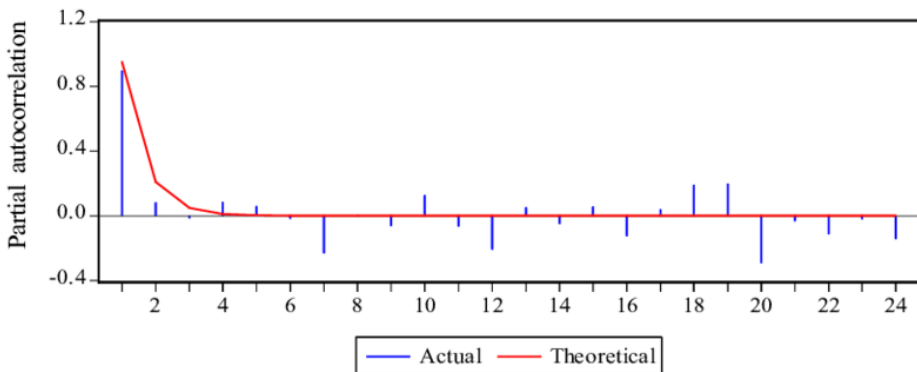
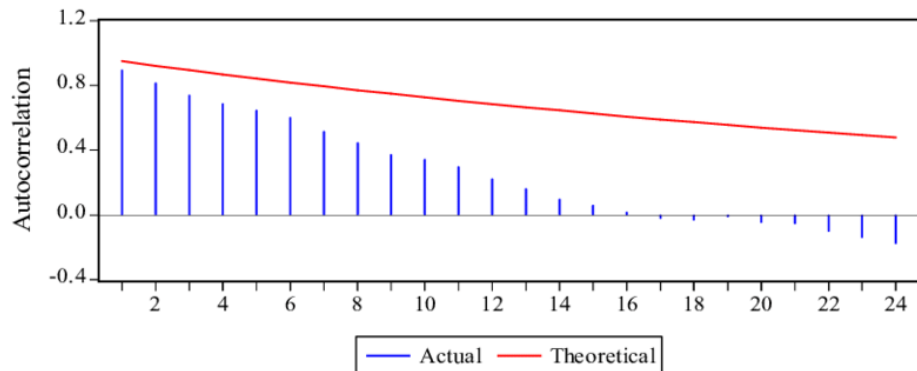
자기회귀-이동평균모형: ARMA(p,q)

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}$$

- ARMA(p, q) 모형은 AR(p) 모형과 MA(q) 모형을 결합시킨 모형
- 정상성과 가역성을 만족하기 위해서는,
 - ✓ AR 부분 특성 방정식의 모든 근의 절댓값이 1보다 커야 함 → 정상성 만족
 - ✓ MA 부분 특성 방정식의 모든 근의 절댓값이 1보다 커야 함 → 가역성 만족
- p, q는 하이퍼파라미터 (ACF, PACF 확인 후 결정)

Unit 03 | 확률적 분석 - AR, MA, ARMA

ARMA(1,1) 모형의 ACF, PACF



- ARMA(1,1) 모형의 ACF와 PACF
 - ACF : AR 모형의 ACF처럼 Die out(지수적 감소)
 - PACF : MA 모형의 PACF처럼 Die out(지수적 감소)

Unit 03 | 확률적 분석 - AR, MA, ARMA

AR(p), MA(q), ARMA(p, q) 모형 비교

- AR(p), MA(q), ARMA(p, q) 모형의 Theoretical ACF, PACF 특징

	AR(p)	MA(q)	ARMA(p, q)
ACF	Die out (지수적 감소, 소멸하는 sine 함수 형태)	Cut off after lag q (q+1 시점부터 0으로 절단)	Die out (q-p+1 시점부터 지수적 감소, 소멸하는 sine 함수 형태)
PACF	Cut off after lag p (p+1 시점부터 0으로 절단)	Die out (지수적 감소, 소멸하는 sine 함수 형태)	Die out (p-q+1 시점부터 지수적 감소, 소멸하는 sine 함수 형태)

Unit 03 | 확률적 분석 - AR, MA, ARMA

비정상 시계열 - ARIMA(p, d, q) 모형

- 추세나 계절성이 포함된 비정상 시계열의 경우, 해당 시계열을 정상화시켜 모형을 적합해야 한다.
- 시계열 정상화 방법
 - 로그변환 : 분산 일정하지 않을 때
 - 추세분석 : 결정적 추세 존재할 때
 - 차분 : 확률적 추세 존재할 때
- 참고) 비정상성 판단
 - ✓ Augmented Dickey-Fuller Test : 단위근 검정
 - ✓ 검정통계량 값(ADF)이 임계값보다 작으면(p-value가 신뢰수준보다 작으면) stationary
 - ✓ [Python] `from statsmodels.tsa.stattools import adfuller`

비정상시계열의 대표적인 특성

- 시계열의 수준이 시간대에 따라 다름
- 시계열이 추세를 가짐
- 시계열이 계절성을 보임
- 시계열의 분산이 시간에 따라 변함

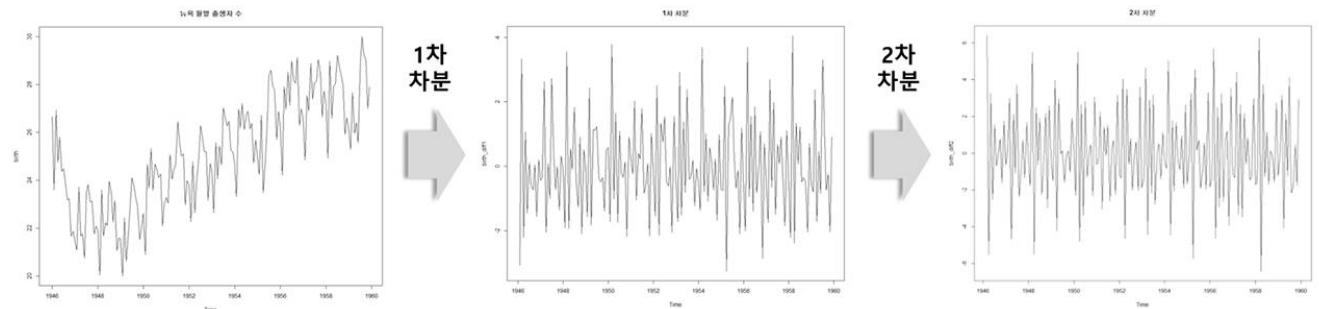
Unit 03 | 확률적 분석 - AR, MA, ARMA

비정상 시계열 - ARIMA(p, d, q) 모형

- 차분(Differencing) : 연이은 관측값들의 차이를 계산하는 것
- 차분을 통해 시계열을 정상화할 수 있는데, 과대차분은 ACF를 복잡하게 만들거나 분산을 크게 함

1차 차분 $y'_t = y_t - y_{t-1} = \nabla y_t$

2차 차분 $y''_t = y'_t - y'_{t-1}$
 $= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$
 $= y_t - 2y_{t-1} + y_{t-2} = \nabla^2 y_t$



Unit 03 | 확률적 분석 - AR, MA, ARMA

비정상 시계열 - ARIMA(p, d, q) 모형

- ARIMA(p, d, q) : 누적(Integrated) 자기회귀이동평균 모형
 - ✓ d차 차분으로 변환된 시계열이 ARMA(p, q)를 따르면, 원 시계열은 ARIMA(p, d, q)를 따른다.
- p, d, q는 하이퍼파라미터
- 참고) 계절성 시계열 데이터
 - ✓ 계절성이 있는 시계열 데이터는 비정상
 - ✓ 단순 차분을 반복하여 적용함으로써 정상 시계열로 전환되기도 하나, 과대차분의 위험
 - ✓ 일반적으로 계절 차분이 효과적 $\rightarrow Z_t \sim \text{SARIMA}(p, d, q) \times (P, D, Q)_s$

Unit 03 | 확률적 분석 - AR, MA, ARMA

튜토리얼 : <https://hyperconnect.github.io/2020/03/09/prophet-package.html>

참고) 시계열 예측 패키지 Prophet

- Prophet : 페이스북에서 공개한 시계열 예측 라이브러리
 - ✓ 정확도가 높고 빠름
 - ✓ 직관적인 파라미터 → 모델 수정에 용이
- 모델의 주요 구성 요소 : Trend, Seasonality, Holiday
 - $y(t) = g(t) + s(t) + h(t) + \epsilon_i$
 - $g(t)$: Trend. 부분적으로 선형 or 로지스틱 곡선으로 구성.
 - $s(t)$: Seasonality. Trend와 달리 주기적으로 나타나는 패턴 포함.
 - $h(t)$: Holiday. 휴일과 같이 불규칙한 이벤트 나타냄.

Contents

Unit 01 | 시계열 데이터

Unit 02 | 고전적 분석 - 평활법 & 분해법

Unit 03 | 확률적 분석 - AR, MA, ARMA

Unit 04 | Box-Jenkins 방법론

Unit 04 | Box-Jenkins 방법론

식별

- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정

추정

- 최소제곱법, 최대가능도법, 비선형최소제곱법 등 이용

진단

- 잔차의 시계열도, ACF, PACF 및 포트맨토 통계량을 이용한 잔차분석
- AIC, SBC 등의 통계값 등을 이용해 가장 설명력 높은 모형 선택

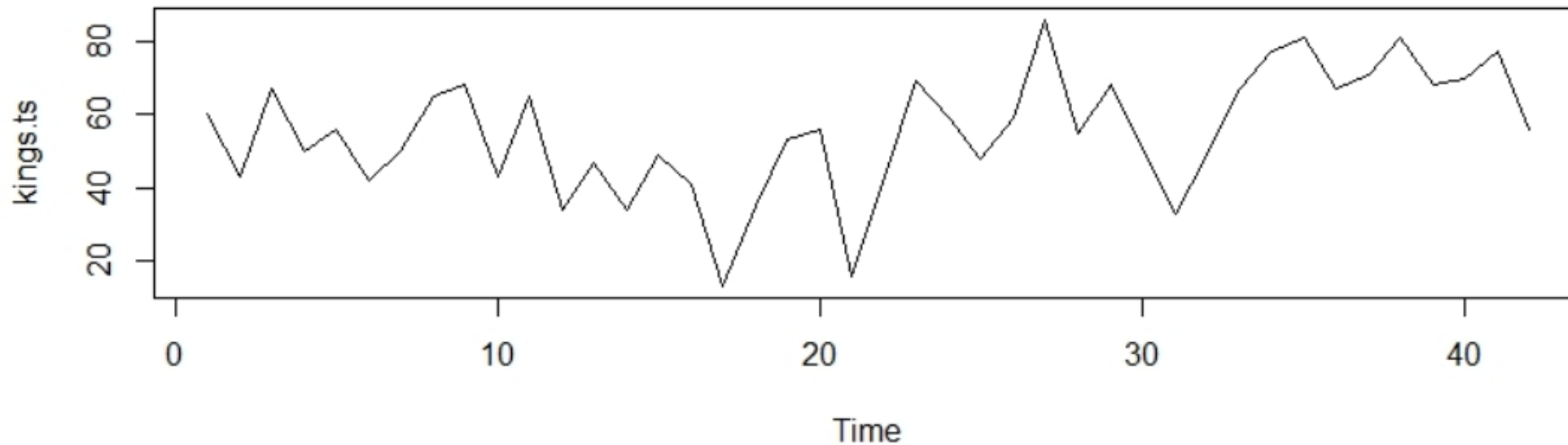
예측

- 적합한 최종모형으로 이후 값 예측

Unit 04 | Box-Jenkins 방법론

식별

- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정

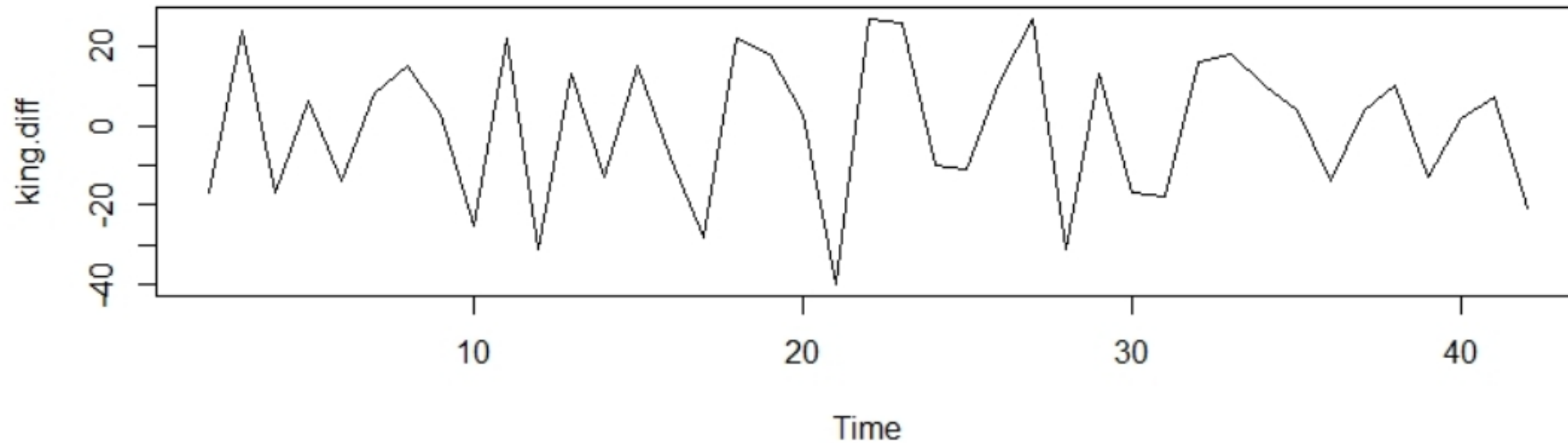
예시 출처 : <https://ckmoong.tistory.com/10>

✓ 시계열그림 그려서 정상성 확인!
평균이 시간에 따라 일정하지 않기 때문에 비정상 시계열이다.

Unit 04 | Box-Jenkins 방법론

식별

- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정

예시 출처 : <https://ckmoong.tistory.com/10>

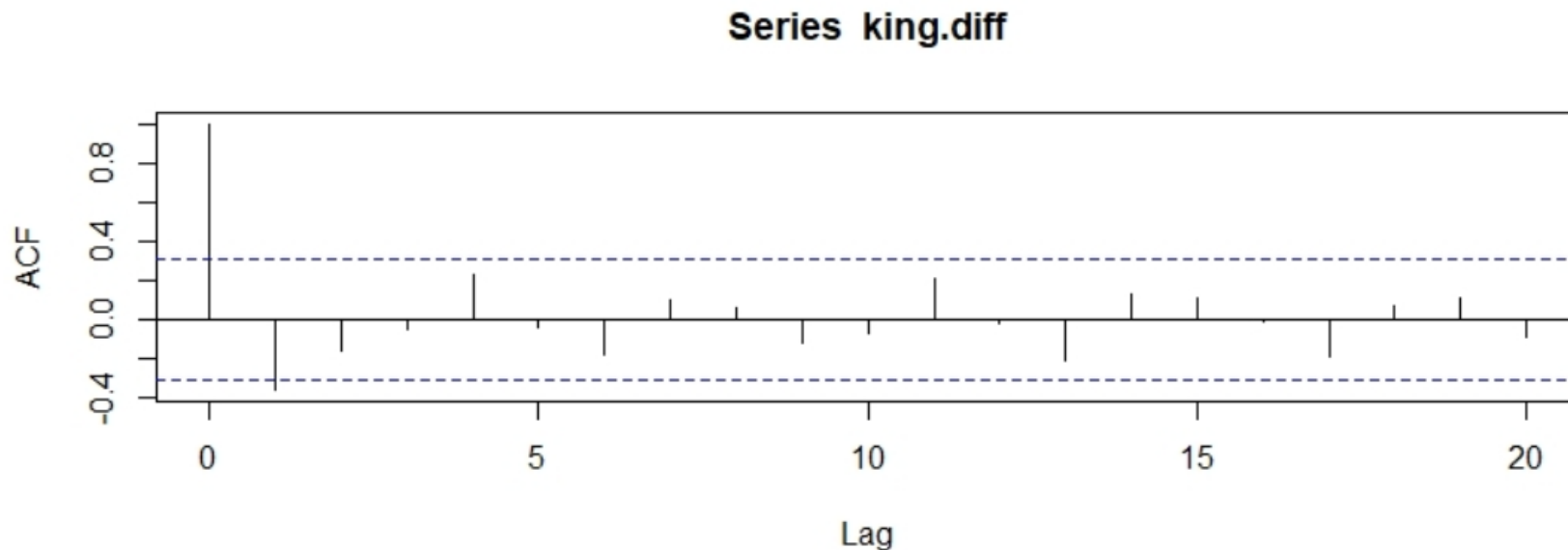
✓ 차분 통해 정상화! : $ARIMA(p,1,q)$

ADF 검정 등을 통해 정상 시계열이 됐는지 확인해보는 것도 좋은 방법!

Unit 04 | Box-Jenkins 방법론

식별

- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정

예시 출처 : <https://ckmoong.tistory.com/10>

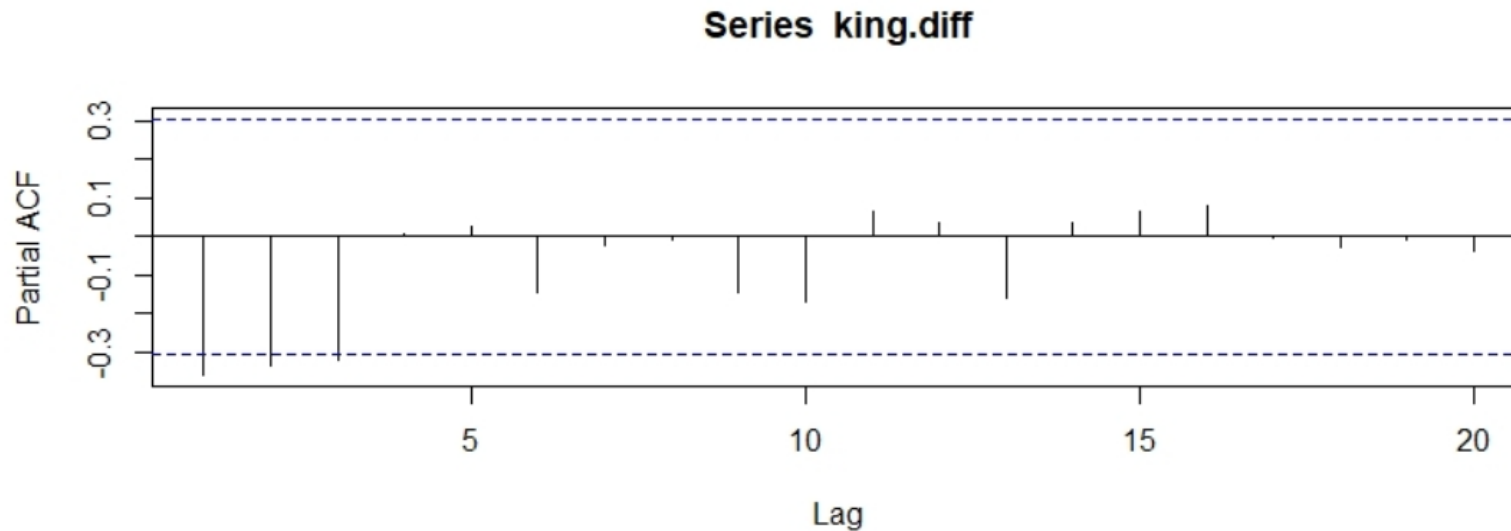
✓ ACF, PACF 통해 모형 식별!

ACF가 Lag 1 이후로 모두 점선 구간 안에 있으므로 (0으로 판단되므로) MA(1) 모형을 잠정모형으로 판단!

Unit 04 | Box-Jenkins 방법론

식별

- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정

예시 출처 : <https://ckmoong.tistory.com/10>

✓ ACF, PACF 통해 모형 식별!

PACF가 Lag 3 이후로 모두 점선 구간 안에 있으므로 (0으로 판단되므로) AR(3) 모형을 잠정모형으로 판단!

Unit 04 | Box-Jenkins 방법론

예시 출처 : <https://ckmoong.tistory.com/10>

추정

- 최소제곱법, 최대가능도법, 비선형최소제곱법 등 이용

잠정모형	모수
ARIMA(3,1,0)	$\phi_1 = -0.6063, \phi_2 = -0.4904, \phi_3 = -0.3284$
ARIMA(0,1,1)	$\theta_1 = -0.7218$
ARIMA(3,1,1)	$\phi_1 = -0.5805, \phi_2 = -0.4778, \phi_3 = -0.3196, \theta_1 = -0.0293$

✓ 잠정모형의 모수 추정

Unit 04 | Box-Jenkins 방법론

예시 출처 : <https://ckmoong.tistory.com/10>

진단

- 잔차의 시계열도, ACF, PACF 및 포트맨토 통계량을 이용한 잔차분석
- AIC, SBC 등의 통계값 등을 이용해 가장 설명력 높은 모형 선택

잠정모형	AIC
ARIMA(3,1,0)	347.7
ARIMA(0,1,1)	344.4
ARIMA(3,1,1)	350.3

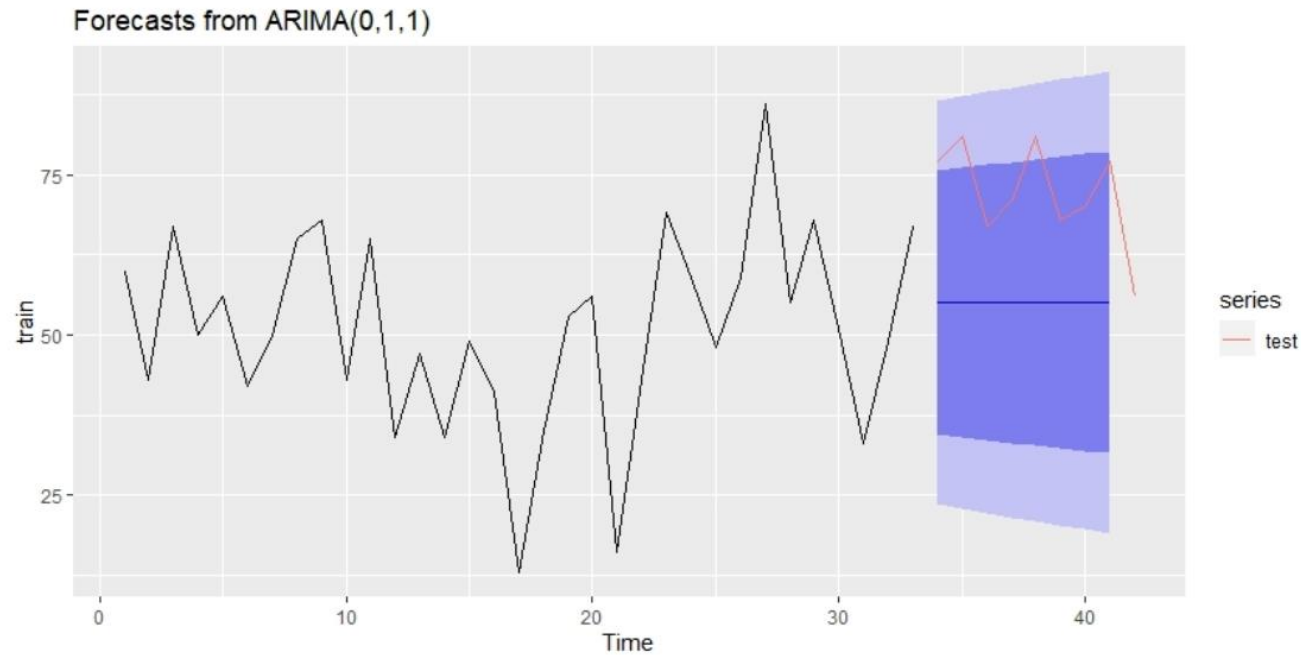
✓ 최종모형 선택

가장 작은 AIC 값을 갖는 ARIMA(0,1,1)을 최종모형으로 채택한다.
잔차분석을 진행하는 것도 좋은 방법이다!

Unit 04 | Box-Jenkins 방법론

예측

- 적합한 최종모형으로 이후 값 예측

예시 출처 : <https://ckmoong.tistory.com/10>

✓ 이후 값 예측

더 자세한 흐름은 시계열분석_참고자료 참고!

과제

[과제] 시계열 분석 실습

- 데이터 : 'Electric_Production.csv'
 - ✓ 출처 : <https://www.kaggle.com/datasets/shenba/time-series-datasets>
 - ✓ Box-Jenkins 방법론(모형 식별, 모수 추정, 모형 적합성 진단, 모형 확정 및 예측) 혹은 Prophet 패키지 사용 등 자유롭게 시계열 분석을 진행해주세요!
- Python, R 중 어느 것을 사용하셔도 무관합니다.

Reference

[강의안]

- 투빅스 14기 이원도님 강의안
- 투빅스 16기 이예림님 강의안
- 연세대학교 응용통계학과 문지은 교수님 <시계열분석> 강의안
- 서울시립대학교 통계학과 오유진 강사님 <시계열분석및실습> 강의안

[교재]

- Robert H. Shumway, David S. Stoffer, <Time Series Analysis and Its Applications>
- 조신섭 외 2인. <SAS/ETS와 R을 이용한 시계열 분석>

[참고 자료]

- [시계열 분해] [Chapter 6 시계열 분해 | Forecasting: Principles and Practice \(otexts.com\)](#)
- [AR, MA, ARMA] [\[머신러닝\]\[시계열\] AR, MA, ARMA, ARIMA의 모든 것 - 개념편 \(velog.io\)](#)
- [AR, MA, ARMA] [10. \[R\] AR모형과 MA모형의 ACF, PACF 시뮬레이션](#)
- [ARIMA] [ARIMA, Python으로 하는 시계열분석 \(feat. 비트코인 가격예측\) \(byeongkijeong.github.io\)](#)
- [Prophet] <https://hyperconnect.github.io/2020/03/09/prophet-package.html>

Q & A

들어주셔서 감사합니다.