시계열자료분석팀

5팀

김나현
강철석
이승아
김재원
이신영

CONTENTS

1

시계열자료분석 개요

시계열 자료의 정의 및 목적

시계열 자료(Time series)

시간에 따라 관측된 자료의 집합

$\{x_t,\ t\ \in T_0\}$

시점 t의 종류에 따른 시계열 자료

이산형 시계열

$\{x_t\}$ if $T_0 \in \mathbb{Z}$ *(정수)*

연속형 시계열

$\{x_t\}$ if $T_0 \in \mathbb{R}$ *(실수)*

1 시계열자료분석 개요

시계열 자료의 정의 및 목적

시계열 자료 분석

시계열 자료와 추세 분석을 다루는 통계 기법으로, 시간 순으로 정렬된 데이터에서 관계를 찾아내고 의미 있는 요약과 통계 정보를 추출하는 과정

시계열 자료 분석의 목적

예측을 위한 분석

시계열 데이터를 활용하여
미래의 값을 예측하고자 하는 분석

추세분석, 평활법, 분해법,
자기회귀누적이동평균(ARIMA) 모형 등

시스템 이해 및 제어를 위한 분석

시계열 패턴을 통해 시스템의
동작 원리와 구조를 이해하고,
이를 바탕으로 시스템을 **제어하거나**
최적화하기 위한 분석

스펙트럼 분석, 개입분석, 전이함수 모형 등

1 시계열자료분석 개요

시계열 자료의 특징

시계열 자료는 시간에 따라 관측되었기 때문에 시간의 흐름이 반영되어
관측치(observation) 간의 **연관성(dependency)**이 존재

특정 시점에 대한 확률 변수 $\{X_t\}$의 분포는 하나의 관측치만을 고려한 것이 아닌,
전체 시점에서의 관측치 집합$\{x_1, x_2, \cdots\}$을 모두 고려한 결합분포(joint distribution)임

1 시계열자료분석 개요

시계열 자료의 특징

시계열 자료는 시간에 따라 관측되었기 때문에 시간의 흐름이 반영되어

관측치(observation) 간의 **연관성(dependency)**이 존재

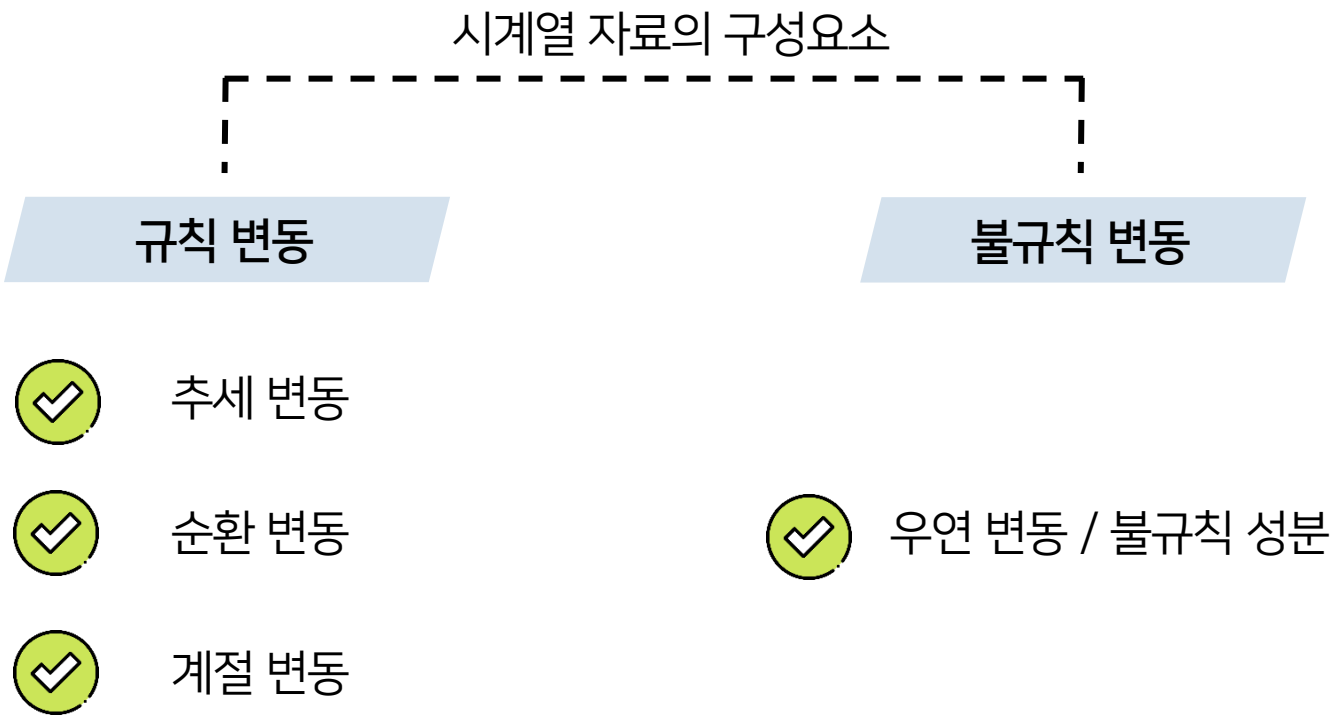
즉, 상당수의 통계적 분포에서 가정하는 **독립성 조건을 만족하지 않음**

→ 데이터 간의 **연관성**을 반영하도록

시간 흐름에 따른 패턴과 변동성을 파악할 수 있는 분석이 요구됨

1 시계열자료분석 개요

시계열 자료의 구성 요소



다음 네 가지 **구성요소들을 분해하여 미래를 예측**하는 것이 **시계열 분석의 목적**

시계열 자료의 구성 요소 | 규칙 요소

추세 변동 (Trend)

시간의 흐름에 따라 관측치가 **증가하거나 감소하는** 추세를 가지는 변동

특별한 충격이 없는 한 지속되는 특성이 있음

순환 변동 (Cycle)

일정한 **주기**를 가지고 변화하지만 **규칙적으로 발생하지 않는** 변동

경제적, 사회적 요인에 의해 발생해 예측이 어려움

주기적인 변화가 있지만 계절에 의한 것이 아니며, 주기가 긴 경우의 변동

시계열 자료의 구성 요소 | 규칙 요소

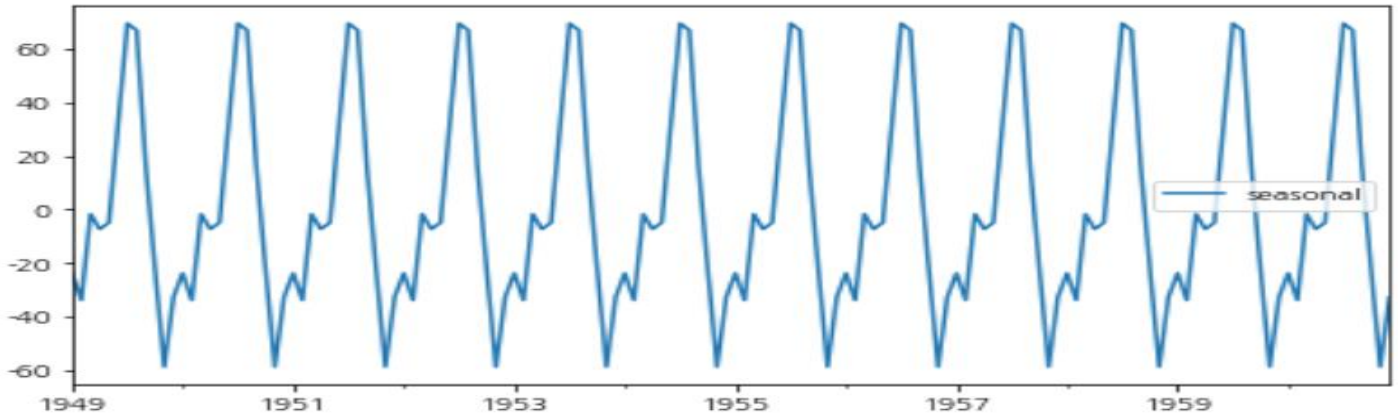
계절 변동 (Seasonal Variation)

규칙적인 주기를 가지고 발생하는 변동

주별, 월별, 계절별과 같이 **특정 시간 간격**을 가지고 반복하는 특징

환경적인 요인에 의해 발생하기 때문에 예측 및 처리에 용이

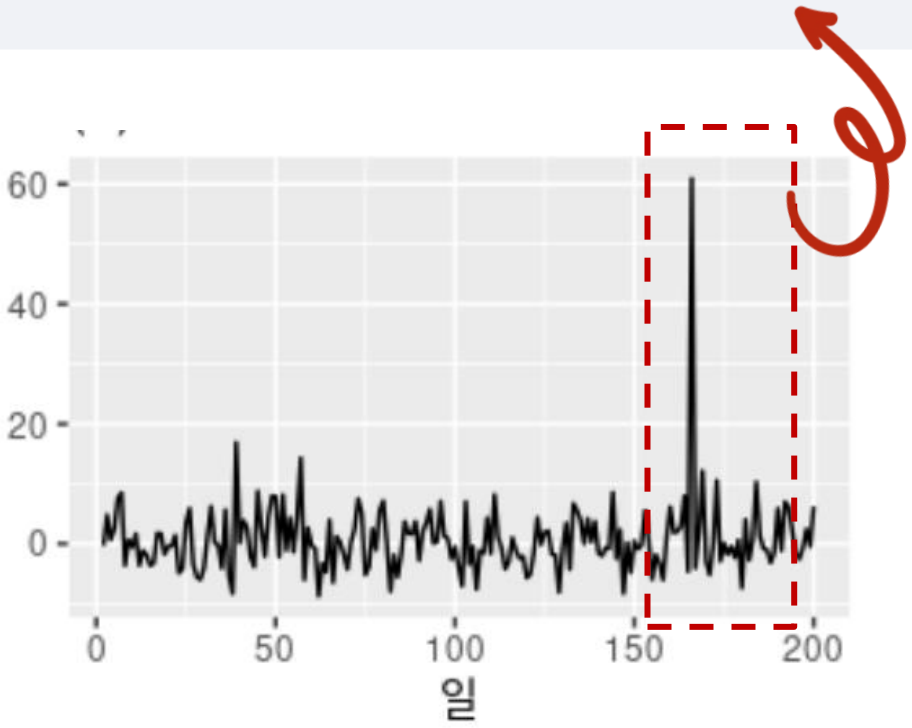
Ex) 계절변동



시계열 자료의 구성 요소 | 불규칙 요소

우연 변동 / 불규칙 성분

무작위 원인에 의해 나타나 일정한 규칙성을 인지할 수 없는 변동



시계열 분해

시계열 분해

시계열 자료 분석의 전통적인 방법 중 하나로 시계열 자료를

비정상 부분과 **정상** 부분으로 분해하여 시계열을 해석하는 방법

Non-Stationary *Stationary*

m_t(추세)와 s_t(계절성) : 비정상 부분

Y_t(오차) : 정상부분

시계열 분해

덧셈 분해

$$X_t = m_t + s_t + Y_t$$

곱셈 분해

$$X_t = m_t * s_t * Y_t$$

1 시계열자료분석 개요

시계열 분해 | 덧셈 분해

덧셈 분해(additive decomposition)

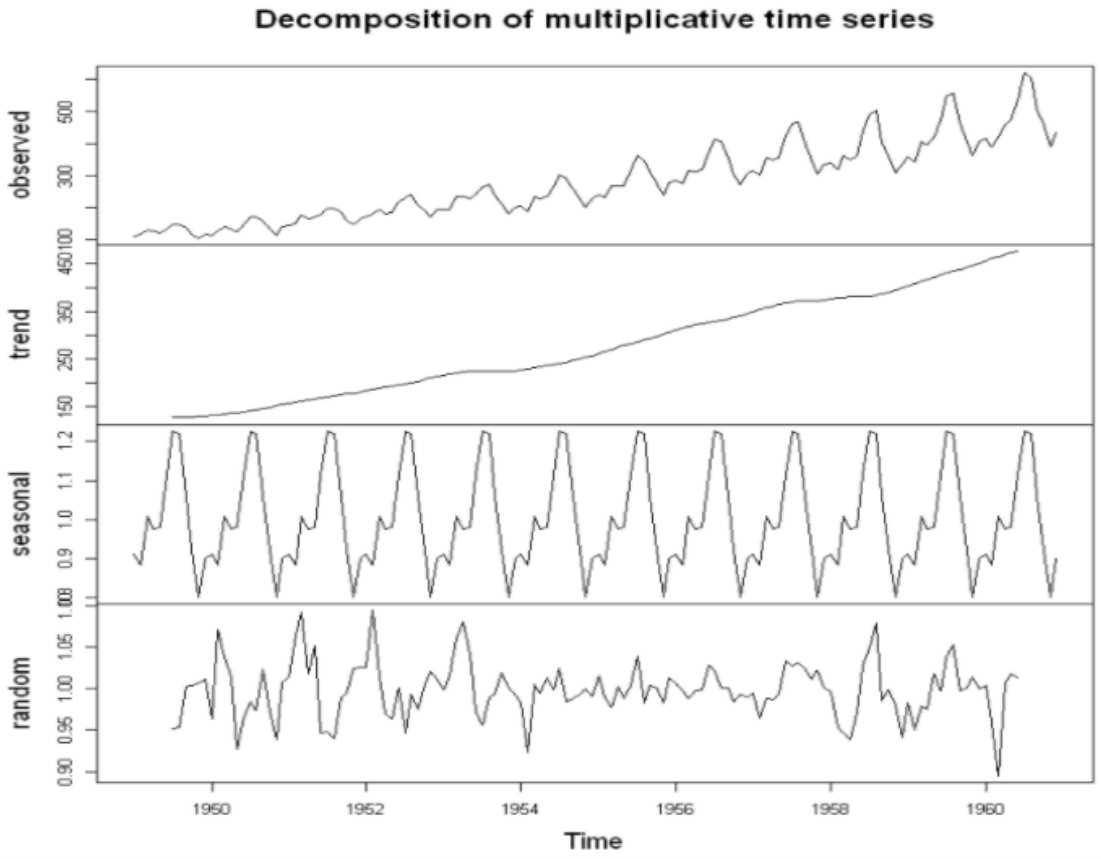
$$X_t = m_t + s_t + Y_t$$

덧셈 분해는 위 식과 같이 시계열 자료를 **덧셈**으로 분해하는 것을 의미함

시계열 분석에서는 m_t(추세)와 s_t(계절성)을 제거한 후
정상성을 만족하는 오차를 이용해 모델링을 진행함

이번 클린업에서는 덧셈 분해를 이용한 시계열 분석을 주로 다룸

시계열 분해 | 덧셈 분해



시계열 데이터

추세

계절성

오차

시계열 분해 | 곱셈 분해

곱셈 분해(additive decomposition)

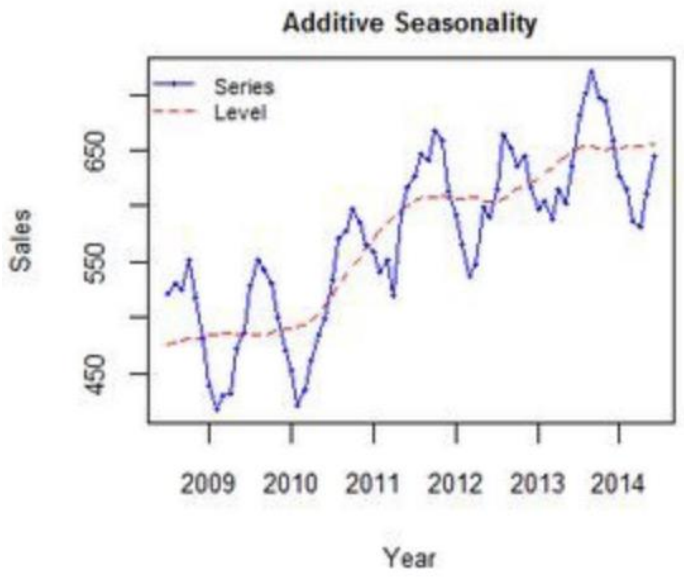
$$X_t = m_t * s_t * Y_t$$

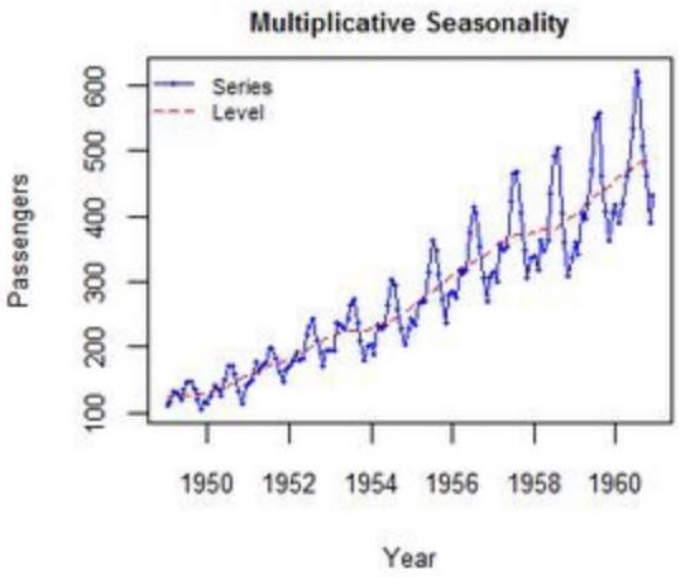
곱셈 분해는 위 식과 같이 시계열 자료를 구성요소의 **곱**으로 분해하는 것을 의미함

덧셈 분해는 추세와 계절성을 **별개**의 구성요소로 보지만,
곱셈 분해는 **추세에 따라 계절성이 변화함**을 가정한다는 것이 차이점임

단, 곱셈이므로 데이터에 0이 포함되어선 안 됨

시계열 분해 | 곱셈 분해

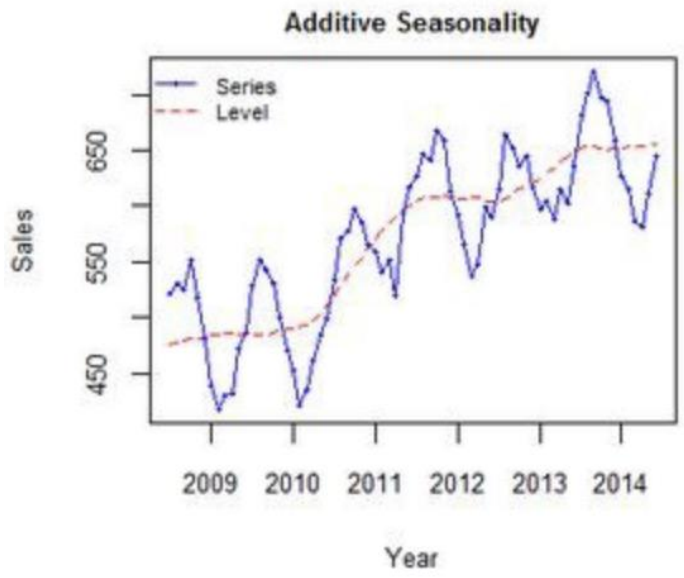


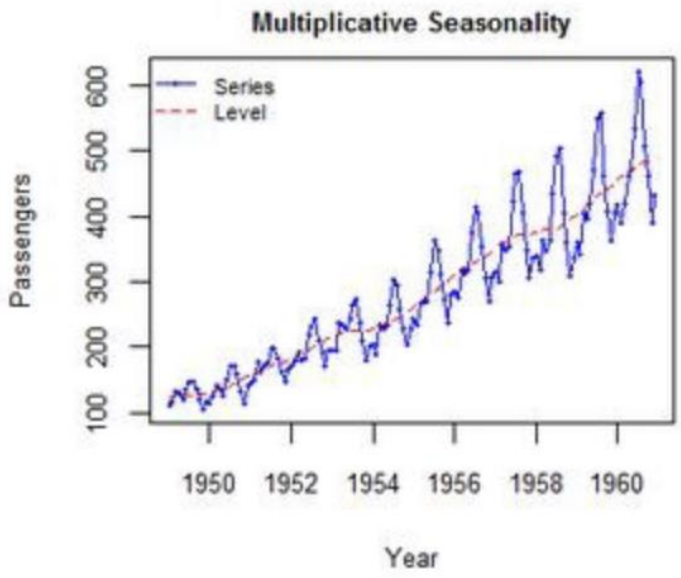


추세와 계절성을 **별개의 구성요소**로 간주 가능

추세에 따라 계절성이 변화한다고 볼 수 있음

시계열 분해 | 곱셈 분해





추세와 계절성을 별개의
덧셈분해 사용
구성요소로 간주 가능

추세에 따라 계절성이
곱셈분해 사용
변화한다고 볼 수 있음

2

정상성

2 정상성

정상성 가정의 필요성

미래의 값을 **예측**하는 목적을 지닌 시계열 모델의 경우,

추정하고자 하는 분포의 **미래 index**를 포함해야 함

$X = (X_1, \cdots, X_n)'$의 결합 분포

$F_X(x_1, \cdots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n,)$ 와 같이 표현

미래의 값을 예측하기 위해서는

무한한 시점들 $(X_1, \cdots, X_n, X_{n+1}, \cdots)'$ 의 **결합 분포**를 고려해야함

정상성 가정의 필요성

미래의 값을 **예측**하는 목적을 지닌 시계열 모델의 경우,

추정하고자 하는 분포의 미래 index를 포함해야 함

즉, **자료는 유한함**에도 불구하고

무한의 dimension에 대한 분포를 계산해야 함

$X = (X_1, \cdots, X_n)'$의 결합 분포

$F_X(x_1, \cdots, x_n) = P(X_1 \le x_1, \cdots, X_n \le x_n,)$ 와 같이 표현

현실적으로 **매우 복잡**하기에 몇 가지의 가정을 통해 데이터를 단순화하며,

이때의 가정을 **정상성(stationarity) 가정**이라 함

미래의 값을 예측하기 위해서는

무한한 시점들 $(X_1, \cdots, X_n, X_{n+1}, \cdots)'$ 의 **결합 분포**를 고려해야함

2 정상성

정상성 개념

정상성

time-invariant

시계열의 확률적 성질이 **시간 흐름에 영향을 받지 않는 것**

즉, 시간에 따른 **평균, 분산 등에 변화가 없음**

시계열 자료의 확률적인 성질이

시점에 의존하지 않고 **시차(lag)**에만 의존

강정상성

강정상성 (Strict Stationarity)

일정한 시차 간격(h)을 가지는 관측치들이 모두 같은 분포를 따른다는 특성

$$(X_{t_1}, \cdots, X_{t_n}) \overset{\mathrm{d}}{=} (X_{t_1+h}, \cdots, X_{t_n+h})$$

시계열 $\{X_t\}$에 대하여, t_1 부터 t_n까지의 n 기간만큼의 분포가
시점을 h만큼 옮긴 t_{1+h} 부터 t_{n+h} 기간의 분포와 동일해야 함

강정상성

즉, 강정상성이란 일정한 시차 간격을 가지는 관측치 집합들이

모두 같은 분포를 따르는 것을 말함

그러나 이를 만족하는 것은 현실적으로 **매우 어려우며 복잡함**

2 정상성

약정상성

모든 h와 n에 대하여 시계열 $\{x_t,\ t \in Z\}$가 다음 세 가지 조건을 만족

→ 해당 시계열은 약정상성을 만족

2차 적률이 존재하고 시점 t에 관계 없이 일정함

$$E[|X_t|]^2 < \infty,\ \forall t \in Z$$

평균이 상수로 시점 t에 관계 없이 일정함

$$E[X_t] = m\ ,\ \forall t \in Z$$

자기 공분산은 시차 h에 의존하고 시점 t와 무관

$$\gamma_X(r,s)=\gamma_X(r+h,s+h),\quad \forall r,s,h \in Z$$

2 정상성

약정상성

모든 h와 n에 대하여 시계열 $\{x_t,\ t \in Z\}$가 다음 세 가지 조건을 만족

→ 해당 시계열은 약정상성을 만족

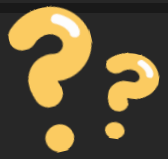
분포 전체가 동일해야 하는 강정상성과 달리

1차 적률과 $\gamma(h)$만 고려하면 된다는 점에서 훨씬 간단함

앞으로 다룰 정상 시계열의 경우 모두 **약정상성**을 가정

2 정상성

약정상성



자기공분산과 자기상관이란?

공분산과 상관계수는 두 변수 간의 관계를 나타내는 **상관성 지표**

공분산은 두 변수 사이의 관계, **상관계수**는 각 공분산을

표준편차로 나눠 두 변수의 선형관계를 -1~1 사이로 나타낸 값

그래서 '자기(auto)'라는 말이 붙는 것!

시계열에서는 두 변수가 아닌 **자기 자신과 몇 시점 떨어진 자기 자신 사이**의

공분산 및 상관 계수를 구함으로써 두 데이터의 연관 정도를 확인함

3

정상화

정상화

추후 배울 시계열 모델은 대부분 **정상성**을 가정하고 전개됨

그러나, 현실 세계의 시계열 자료는 대부분 **정상성 불만족**

회귀분석에서 기본 가정 불만족 시
transformation 과정과 유사

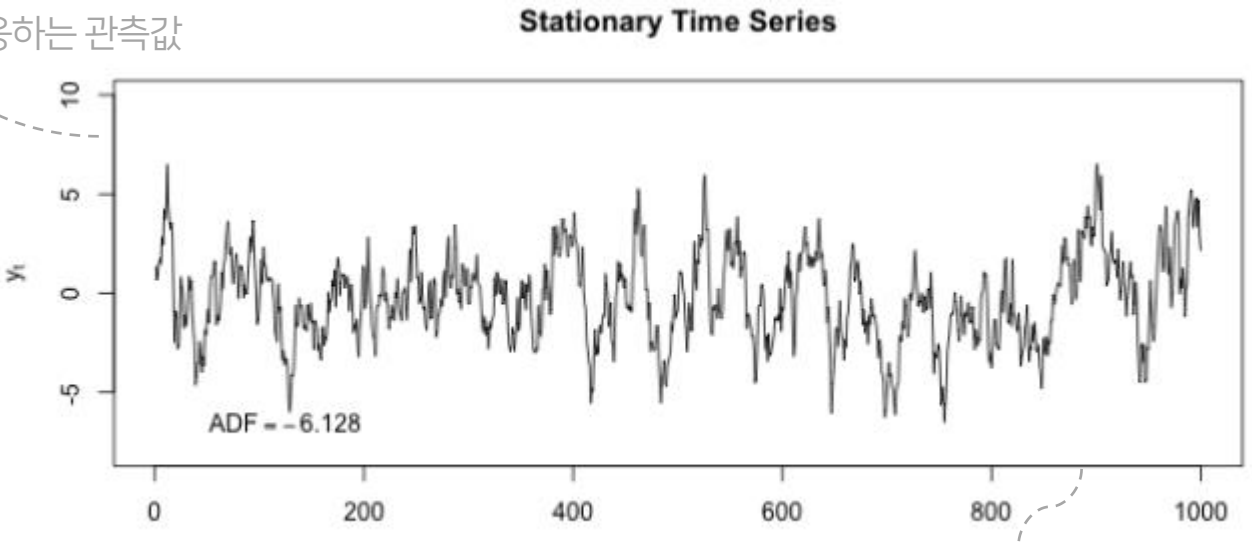
비정상 시계열을 정상 시계열로 변환해주는

정상화 과정이 필요

정상 시계열과 비정상 시계열

시계열(Time Series) plot을 통해 **자료의 정상성 여부**를 파악

Y축 : 각 시점에 대응하는 관측값



X축 : time point t

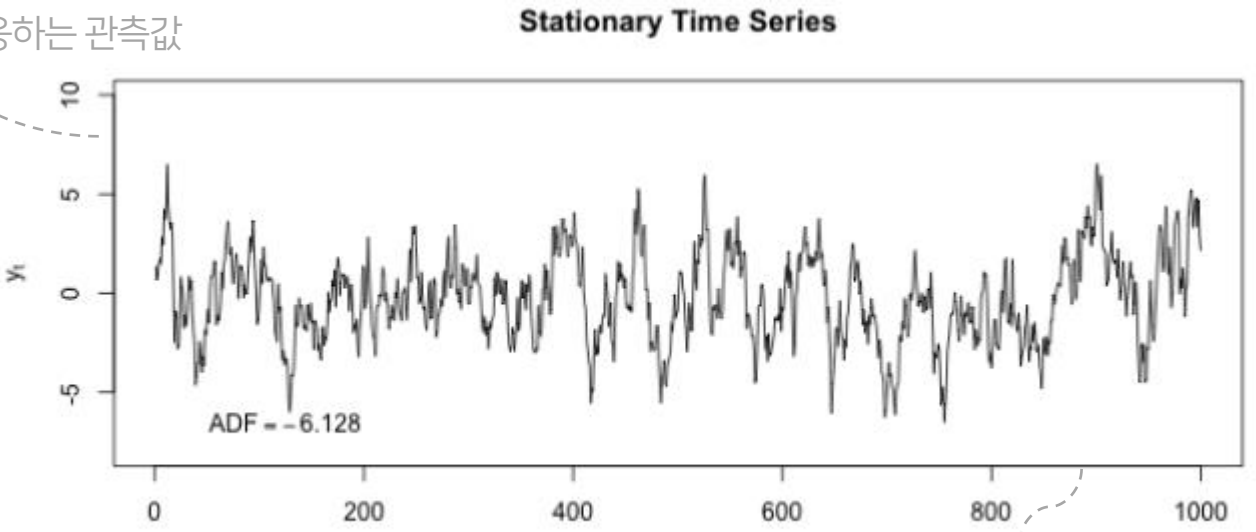
추세가 존재하는지, **돌발적인 변화**가 있는지,

이상치가 존재하는지 등을 파악해야 함

정상 시계열과 비정상 시계열

시계열(Time Series) plot을 통해 **자료의 정상성 여부**를 파악

Y축 : 각 시점에 대응하는 관측값



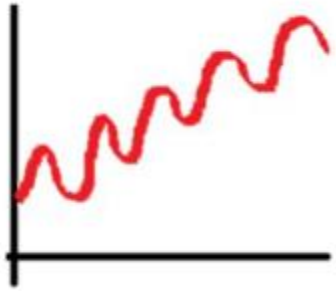
X축 : time point *t*

특별한 추세나 계절성이 없고, 평균과 분산이 일정한 것으로 보아

정상 시계열로 판단 가능

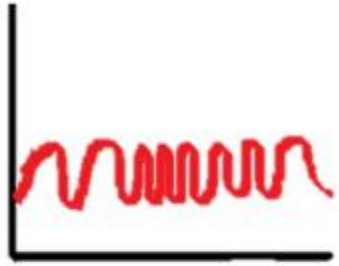
정상 시계열과 비정상 시계열

정상성 조건을 만족하지 못하는 비정상 시계열



평균이 일정하지 않은 경우

분산이 일정하지 않은 경우

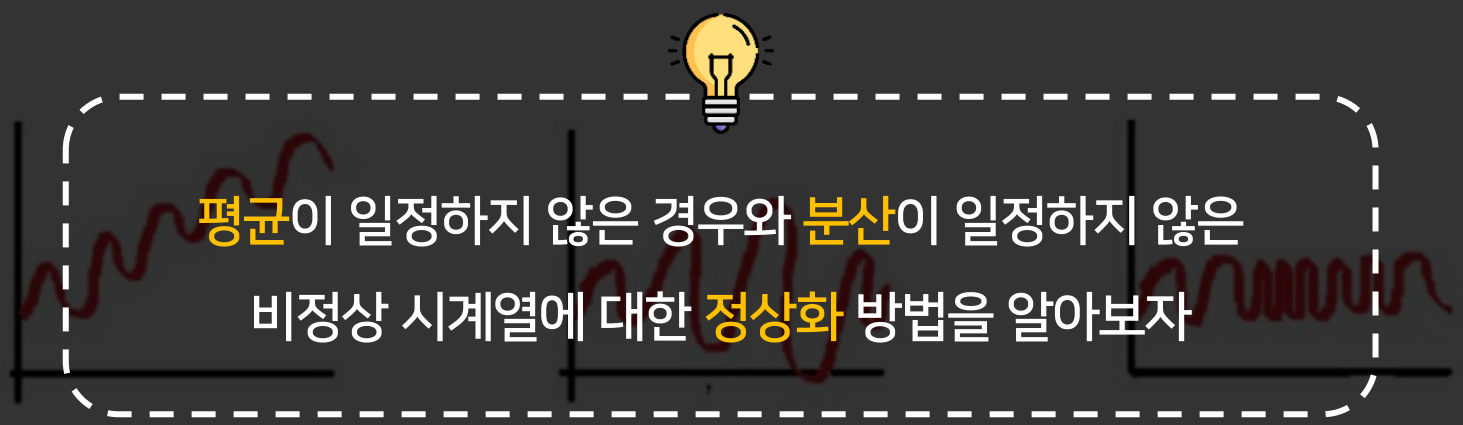


공분산이 시점에 의존하는 경우

3 정상화

정상 시계열과 비정상 시계열

정상성 조건을 만족하지 못하는 비정상 시계열



평균이 일정하지 않은 경우와 분산이 일정하지 않은
비정상 시계열에 대한 정상화 방법을 알아보자

평균이 일정하지 않은 경우

분산이 일정하지 않은 경우

공분산이 시점에 의존하는 경우

3 정상화

분산이 일정하지 않은 경우의 정상화 과정

시간의 흐름에 따라 변동폭이 달라지는

이분산(heteroscedasticity) 자료가 존재함

분산 안정화 변환 (VST, Variance Stabilizing Transformation)

특정 시계열 자료의 분산이 시점 t에 의존하지 않고 일정하도록 만드는 방법

분산이 일정하지 않은 경우의 정상화 과정

분산 안정화 변환

- Box-Cox Transformation

$$f_\lambda(X_t) = \begin{cases} \dfrac{X_t^\lambda - 1}{\lambda} & if\ X_t \geq 0, \lambda \geq 0 \\ \log X_t & if\ \lambda = 0 \end{cases}$$

- Log-Transformation

$$f(X_t) = \log(X_t)$$

- Square root Transformation

$$f(X_t) = \sqrt{X_t}$$

3 정상화

평균이 일정하지 않은 경우의 정상화 과정

시계열 분해식

$$X_t = m_t + s_t + Y_t$$

m_t : 추세, s_t : 계절성, Y_t : 정상성을 만족하는 오차

평균이 일정하지 않은 경우

- (A) 추세만 존재하는 경우
- (B) 계절성만 존재하는 경우
- (C) 추세&계절성이 모두 존재하는 경우

3 정상화

평균이 일정하지 않은 경우의 정상화 과정

시계열 분해식

$$X_t = m_t + s_t + Y_t$$

m_t : 추세, s_t : 계절성, Y_t : 정상성을 만족하는 오차

평균이 일정하지 않은 경우

(A) 추세만 존재하는 경우

(B) 계절성만 존재하는 경우

(C) 추세와 계절성이 모두 존재하는 경우

회귀/평활/차분의 세 가지 방법을 통해
비정상 부분을 추정 및 제거하여 정상화 진행

평균이 일정하지 않은 경우의 정상화 과정 I ① 회귀

[(A) 추세만 존재하는 경우] Polynomial Regression

추세만 존재하도록 가정!

시계열 가정

$$X_t = m_t + Y_t,\ E(Y_t) = 0$$

추세 성분을 t에 대한 선형회귀식으로 표현

$$m_t = c_0 + c_1 t + c_2 t^2 + \cdots + c_p t^p$$

OLS를 통해 추세에 대한 회귀계수 추정

$$(\widehat{c_0}, \cdots, \widehat{c_p}) = \underset{c}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - m_t)^2$$

평균이 일정하지 않은 경우의 정상화 과정 I ① 회귀

[(A) 추세만 존재하는 경우] Polynomial Regression

추정한 추세를 시계열에서 제거

$$X_t - \hat{m}_t \approx Y_t$$

Stationary Error(Y_t) 만 남아

정상시계열 완성

평균이 일정하지 않은 경우의 정상화 과정 I ① 회귀

[(B) 계절성만 존재하는 경우] Harmonic Regression

주기가 d인 계절성만 존재하도록 가정!

시계열 가정

$$X_t = s_t + Y_t,\ E(Y_t) = 0 \ \ where \ \ s_{t+d} = s_t = s_{t-d}$$

계절 성분을 t에 대한 회귀식으로 표현

$$s_t = a_0 + \sum_{j=1}^{k} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

평균이 일정하지 않은 경우의 정상화 과정 I ① 회귀

[(B) 계절성만 존재하는 경우] Harmonic Regression

적절한 λ_t와 κ 선택 후, OLS를 통하여 a_j와 b_j를 추정

$$s_t = a_0 + \sum_{j=1}^{k} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

λ_t : 주기가 2π인 함수의 주기와 데이터의 주기를 맞춰 주기 위한 값

- 주기 반복 횟수 $f_1 = n/d$, $f_j = j \times f_1$
- $\lambda_t = f_j \times (\frac{2\pi}{n})$
- k는 주로 1~4 사이의 값을 사용

Ex) $n = 72,\ d = 12$

$\rightarrow f_1 = \frac{72}{12} = 6,$

$\lambda_j = j \times 6 \times \frac{2\pi}{72}$

평균이 일정하지 않은 경우의 정상화 과정 I ① 회귀

[(B) 계절성만 존재하는 경우] Harmonic Regression

추정한 계절성을 시계열에서 제거

$$X_t - \hat{s}_t \approx Y_t$$

Stationary Error(Y_t) 만 남아

정상시계열 완성

평균이 일정하지 않은 경우의 정상화 과정 I ① 회귀

[(C) 추세&계절성이 모두 존재하는 경우]

추세와 계절성 모두 존재하도록 가정!

시계열 가정

$$X_t = m_t + s_t + Y_t,\ E(Y_t) = 0$$

(A), (B) 과정 차례로 진행

- 추세에 대한 Polynomial Regression
- 계절성에 대한 Harmonic Regression

진행 후에도 추세가 남아 있다면, 같은 과정 반복

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 | ① 회귀

[(C) 추세&계절성이 모두 존재하는 경우]

회귀 방법의 문제점

OLS(최소자승법)는 **오차항의 독립성**을 가정하지만,

시계열의 오차항은 **독립성을 가정하지 않음**

(A), (B) 과정 차례로 진행

- 추세에 대한 Polynomial Regression
- 계절성에 대한 Harmonic Regression

추정의 정확성을 보장할 수 없음

진행 후에도 추세가 남아 있다면, 같은 과정 반복

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 | ① 회귀

[(C) 추세&계절성이 모두 존재하는 경우]

회귀 방법의 문제점

추세와 계절성 모두 존재하도록 가정!

특히 분산을 계산할 때 오차항이

연관되어(correlated) 있음에도 독립을 가정한 상태로 계산됨

(A), (B) 과정 차례로 진행

- 추세에 대한 Polynomial Regression
- 계절성에 대한 Harmonic Regression

신뢰구간 계산에 오류 발생 가능

진행 후에도 추세가 남아 있다면, 같은 과정 반복

평균이 일정하지 않은 경우의 정상화 과정 | ② 평활

회귀

데이터 **전체**를 한 번에 추정

국소적 변동 파악 어려움

평활

데이터의 **일부**만을 추정

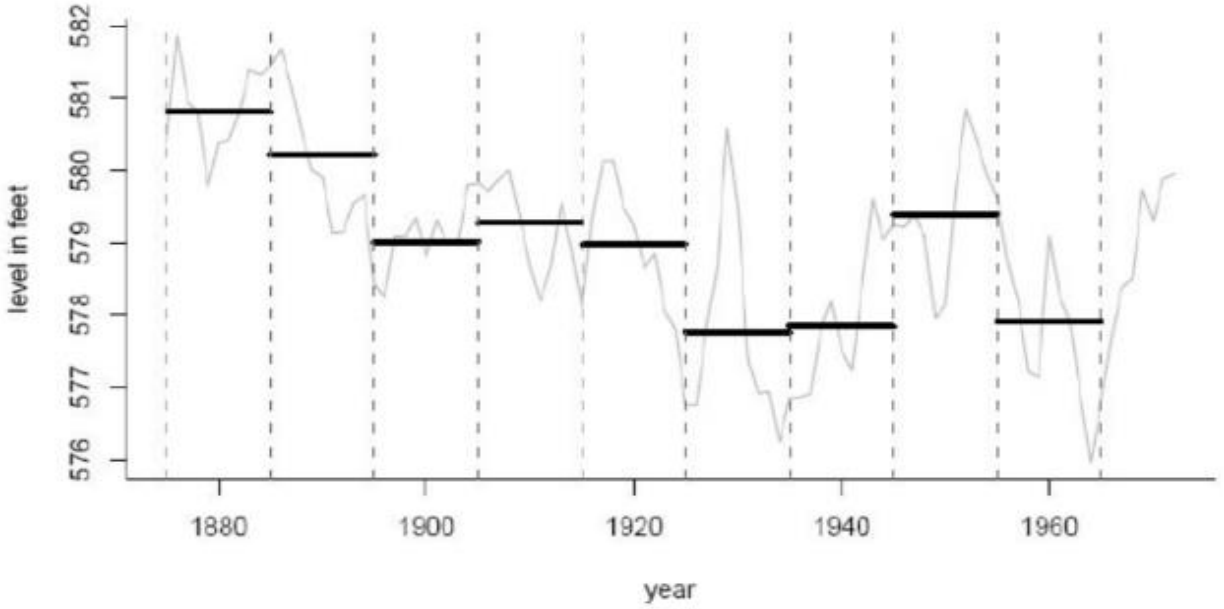
국소적 변동 파악 용이

국소적 변동에 주목해야 하는 경우, 평활 사용!

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

평활 (Smoothing)

시계열 자료를 여러 구간으로 나눈 후, 구간의 평균들로 추세를 추정하는 방법



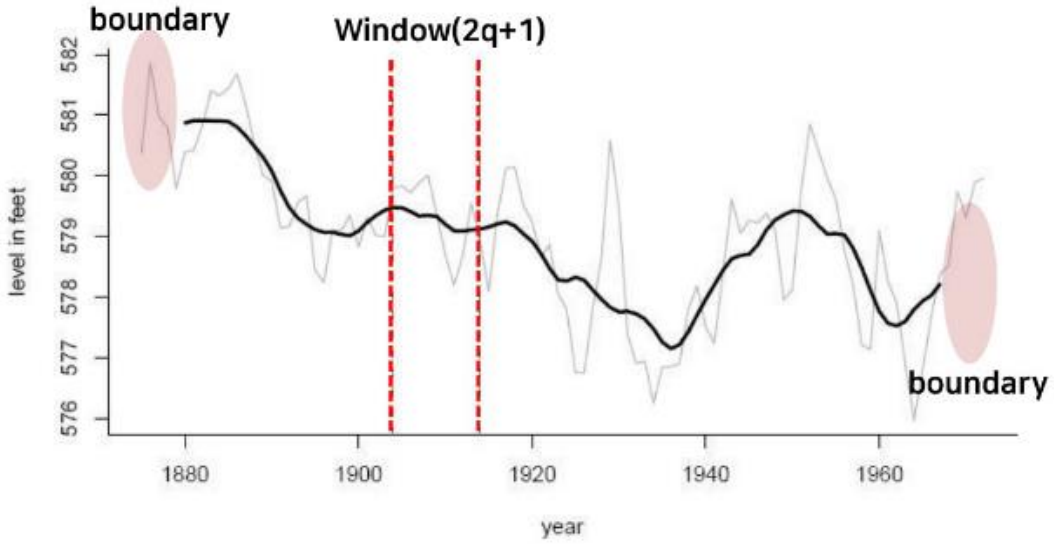
전체 시계열 자료와 구간 평균의 움직임은 비슷할 것이라는 아이디어를 이용

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A1) 추세만 존재하는 경우] **이동평균 평활법**

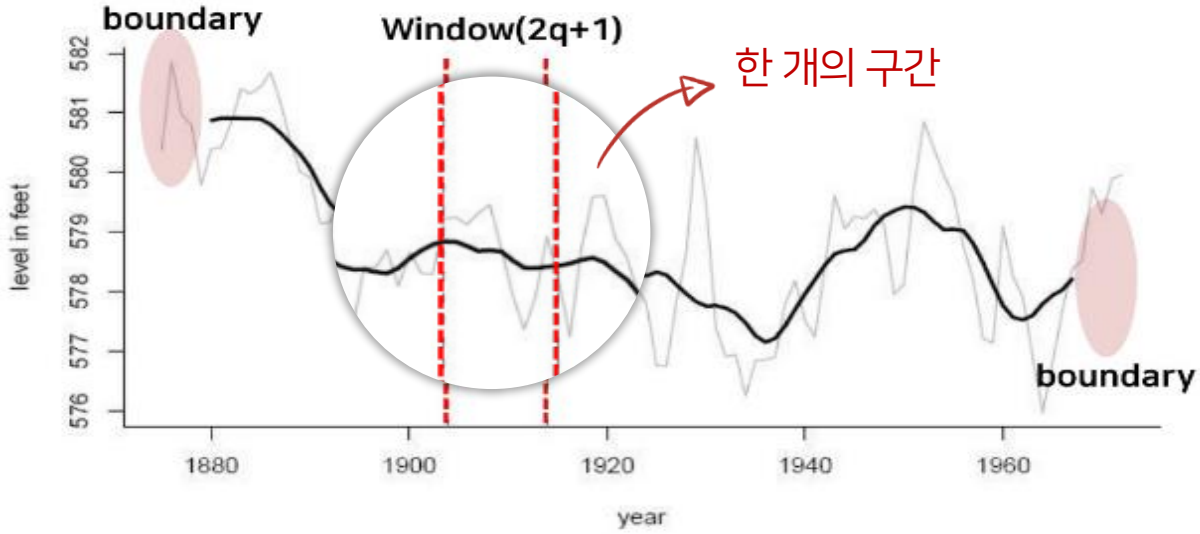
이동평균 평활법 (Moving Average Smoothing)

일정 기간마다의 평균을 계산하여 추세를 추정하는 방식



평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A1) 추세만 존재하는 경우] **이동평균 평활법**



길이가 2q +1인 구간의 평균 구하기

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{j=q} (m_{t+j} + Y_{t+j})$$

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A1) 추세만 존재하는 경우] **이동평균 평활법**

추세의 선형성 가정 시, 평균 W_t와 m_t의 관계 파악

$$m_t = c_0 + c_1 t,\ E(Y_t) = 0$$

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{j=q}(m_{t+j} + Y_{t+j})$$

$$= \frac{1}{2q+1}\sum_{j=-q}^{j=q}(m_{t+j}) + \frac{1}{2q+1}\sum_{j=-q}^{j=q}(Y_{t+j})$$

$$= c_0 + c_1 t = m_t$$

$$\frac{1}{2q+1}\sum_{j=-q}^{j=q}(Y_{t+j}) \approx E(Y_t) = 0\ (by\ WLLN)$$

$$t \in [q+1, n-q]$$

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A1) 추세만 존재하는 경우] **이동평균 평활법**

추세의 선형성 가정 시, 평균 W_t와 m_t의 관계 파악

$$m_t = c_0 + c_1 t, E(Y_t) = 0$$

평균 W_t는 근사적으로 추세 m_t와 같아짐

$$= \frac{1}{2q+1}\sum_{j=-q}^{j=q}(m_{t+j}) + \frac{1}{2q+1}\sum_{j=-q}^{j=q}(Y_{t+j})$$

$$= c_0 + c_1 = m_t$$

일정한 길이의 구간 평균이 전체 시계열의 추세를 잡을 수 있음을 의미

$$\frac{1}{2q+1}\sum_{j=-q}(Y_{t+j}) \approx E(Y_t) = 0 \ (by\ WLLN)$$

$$t \in [q+1, n-q]$$

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A1) 추세만 존재하는 경우] **이동평균 평활법**

추세 부분만 남은 W_t를 전체 데이터에서 제거

$$X_t - \hat{m}_t \approx Y_t$$

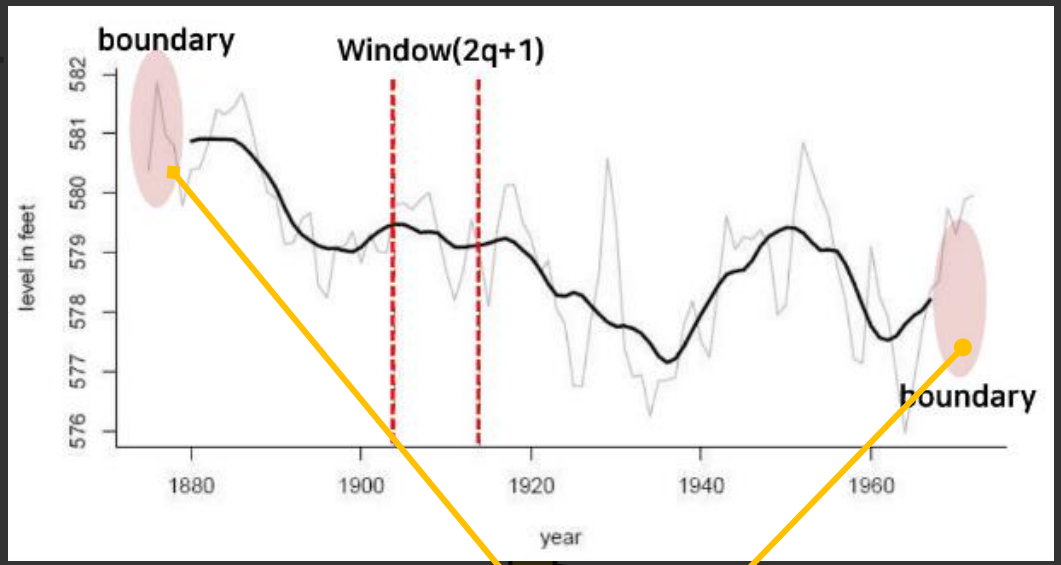
정상시계열 확보

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 | ② 평활

Moving Average Smoothing의 한계

[(A1) 추세만 존재하는 경우] 이동평균 평활법



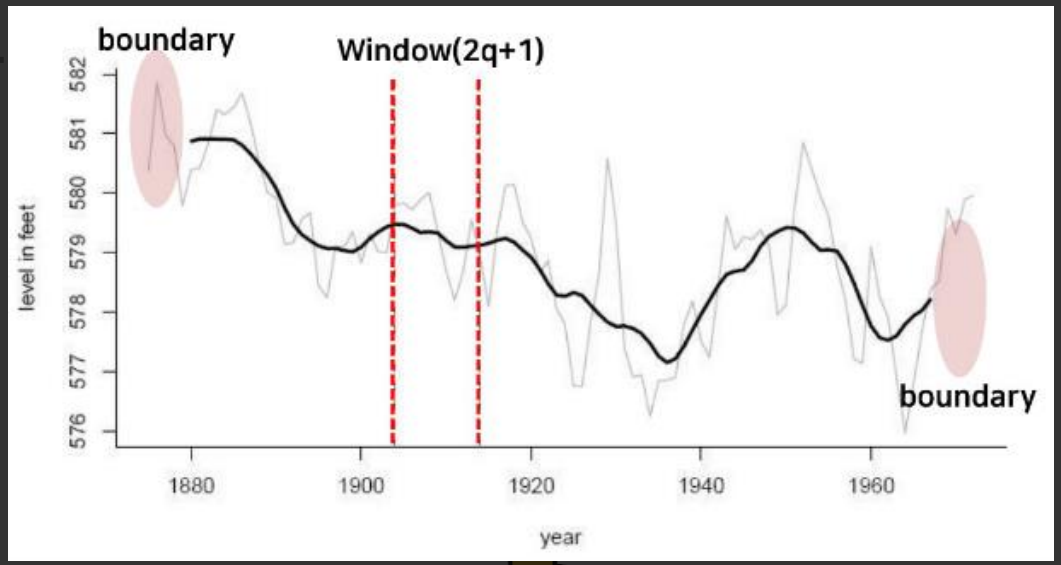
- 데이터의 맨 앞 q개와 맨 뒤 q개의 boundary에 해당하는 값 추정 못 함
- t시점 이후의 데이터는 미래 데이터로 현실에서는 활용이 거의 불가능

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 | ② 평활

[(A1) 추세만 존재하는 경우] 이동평균 평활법

Moving Average Smoothing의 한계



정상시계열 확보

과거의 데이터에만 의존하여 추세를 추정하는

지수평활법 (Exponential Smoothing) 사용!

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A2) 추세만 존재하는 경우] **지수평활법**

지수평활법 (Exponential Smoothing)

미래의 데이터를 활용하지 않고, 추세 $\mathrm{m_t}$를

t시점까지의 관측값만을 이용해 추정하고 제거하는 방법

추세 추정 방식

$$\hat{\mathrm{m}}_1 = \mathrm{X}_1$$

$$\hat{\mathrm{m}}_2 = a\mathrm{X}_2 + (1-a)\hat{m}_1 = a\mathrm{X}_2 + (1-a)\mathrm{X}_1$$

$$\vdots$$

$$\widehat{m_t} = aX_t + (1-a)\hat{m}_{t-1}$$

$$= \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

$a \in [0,1]$인 a에 대하여

현재 시점에는 a만큼,

과거의 예측값에는 $1-a$만큼의

가중치 부여

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(A2) 추세만 존재하는 경우] **지수평활법**

지수평활법 (Exponential Smoothing)

미래의 데이터를 활용하지 않고, 추세 $\mathrm{m_t}$를

t시점까지의 관측값만을 이용해 추정하고 제거하는 방법

추세 추정 방식

$$\hat{\mathrm{m}}_1 = \mathrm{X}_1$$

$$\hat{\mathrm{m}}_2 = a\mathrm{X}_2 + (1-a)\hat{m}_1 = a\mathrm{X}_2 + (1-a)\mathrm{X}_1$$

$$\vdots$$

$$\widehat{m_t} = aX_t + (1-a)\hat{m}_{t-1}$$

$$= \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

과거의 값일수록 가중치가

$a \in [0,1]$인 a에 대하여

지수적으로 줄어드는 것을 확인

현재 시점에는 a만큼,

과거의 예측값에는 $1-a$만큼의

가중치 부여

$$\mathrm{X_t} - \hat{\mathrm{m}}_\mathrm{t} \approx \mathrm{Y_t}$$

추정한 추세 제거 후 **정상시계열** 도출

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 | ② 평활

평활법에서 q와 a의 선택

평활법은 추세 외에도 tuning parameter q와 a 에 대한 추정 필요

q가 작은 경우 작은 변화들도 잘 잡아내지만, 변동성이 심해짐

q가 큰 경우 변동성은 줄어들지만 작은 변화를 잡아내지 못함

= Bias-Variance Trade Off 발생

Cross-validation(cv)을 통해 MSE를 계산하여

최적의 파라미터 선정 필요

과거의 값일수록 가중치가 지수적으로 줄어드는 것을 확인

$$X_t - \hat{m}_t \approx Y_t$$

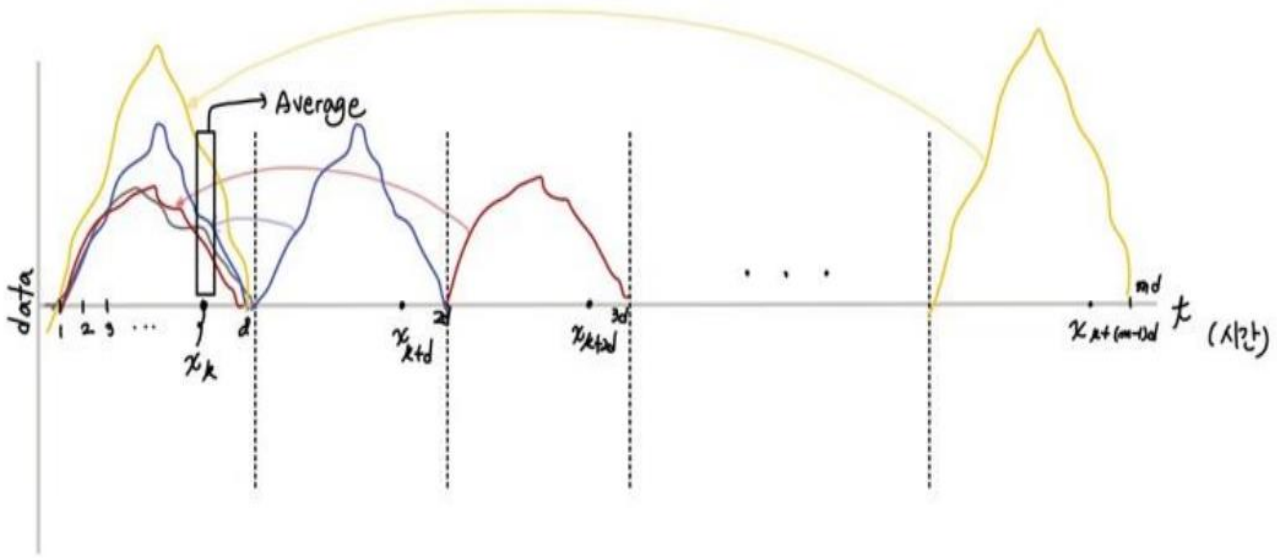
추정한 추세 제거 후 정상시계열 도출

데마팀 1주차 클린업 참고

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(B) 계절성만 존재하는 경우] Seasonal Smoothing

주기가 d인 관측치를 한 주기 안에 모두 겹친 후 **평균**을 확인



평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(B) 계절성만 존재하는 경우] Seasonal Smoothing

계절성분 $\hat{s}_t$ 추정

$$\hat{s}_k = \frac{1}{m}(x_k + x_{k+d} + \cdots + x_{k+(m-1)d}) = \frac{1}{m}\sum_{j=0}^{m-1} x_{k+jd}$$

(m =# of obs. In kth seasonal component)

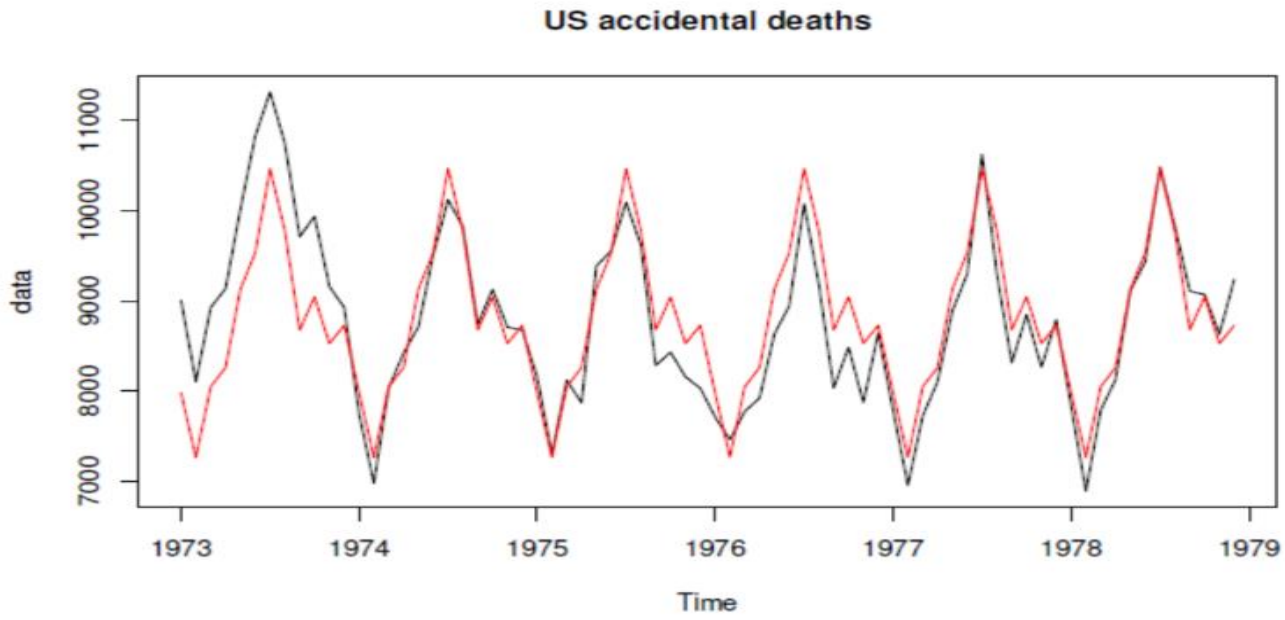
추정된 $\hat{s}_t$를 다른 주기에 적용해 **전체 계절성분** 추정

→ 시계열 자료에서 전체 계절성분 제거

$$X_t - \hat{s}_t \approx Y_t$$

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(B) 계절성만 존재하는 경우] Seasonal Smoothing



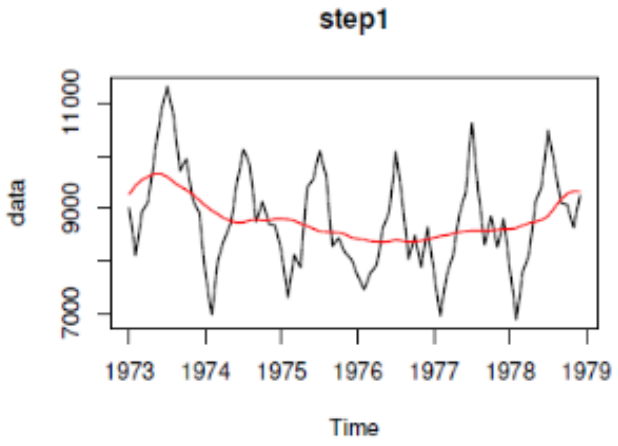
빨간 선으로 나타나는 **계절성분**은 모든 주기에서 **동일하게 반복**되고 있음

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

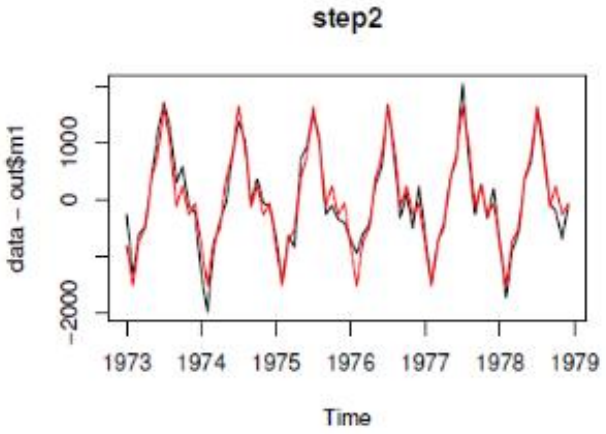
[(C1) 추세&계절성이 모두 존재하는 경우] Classical Decomposition Algorithm

시계열 가정

$$X_t = m_t + s_t + \mathrm{Y_t}\,,\ E(Y_t) = 0$$



MA filter 이용해 **추세 예측 및 제거**

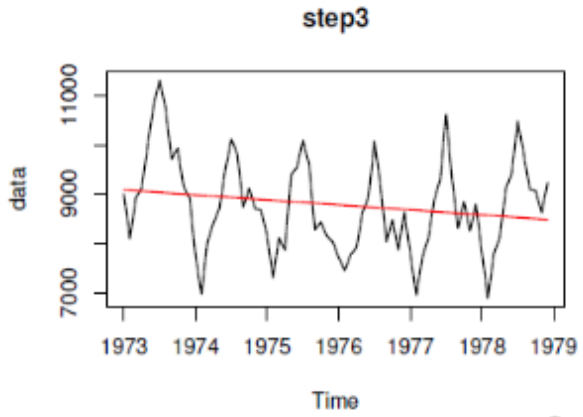


Seasonal smoothing으로

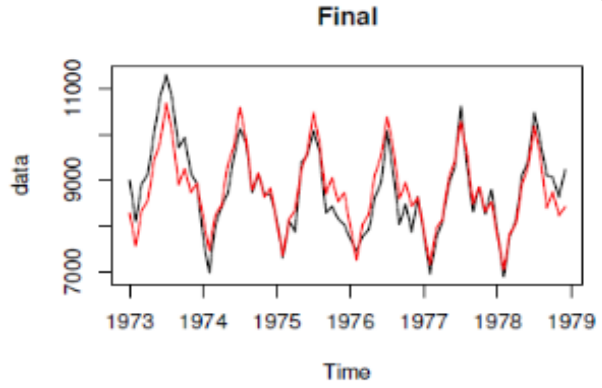
계절성분 추정

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C1) 추세&계절성이 모두 존재하는 경우] Classical Decomposition Algorithm



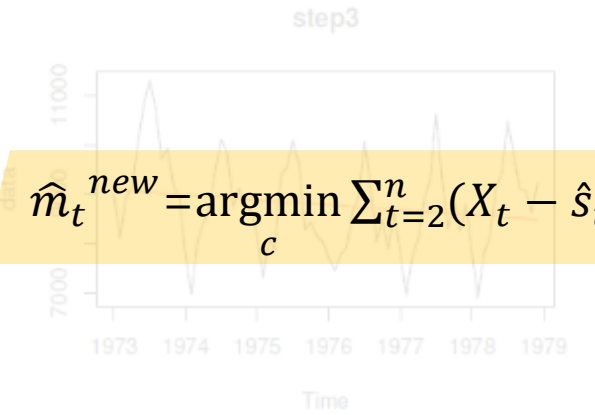
계절성이 제거된 시계열에서
OLS를 통해 다시 **추세 추정**



새롭게 추정한 추세와 계절성
제거 후 오차 추정

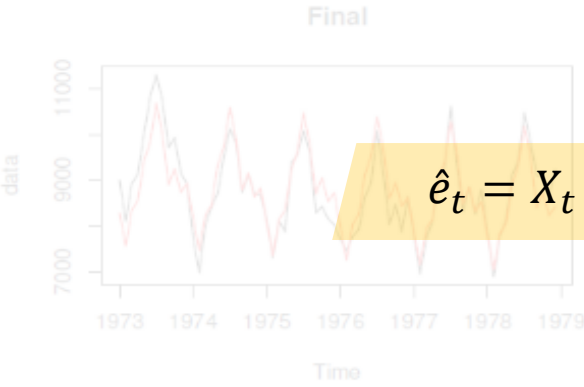
평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C1) 추세&계절성이 모두 존재하는 경우] Classical Decomposition Algorithm



$$\hat{m}_t^{new} = \underset{c}{\operatorname{argmin}} \sum_{t=2}^{n} (X_t - \hat{s}_t - c_0 - c_1 t - c_2 t^2 - \cdots - c_p t^p)^2$$

계절성이 제거된 시계열에서
OLS를 통해 다시 추세 추정

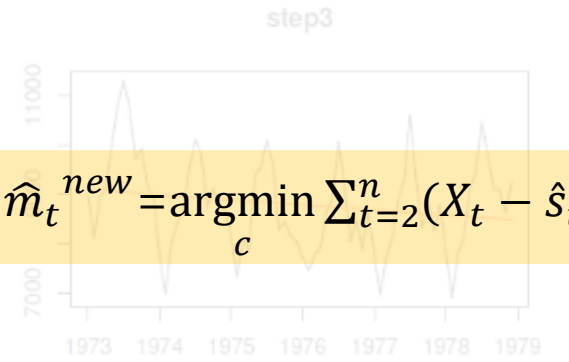


$$\hat{e}_t = X_t - \hat{m}_t - \hat{s}_t$$

새롭게 추정한 추세와 계절성
제거 후 오차 추정

평균이 일정하지 않은 경우의 정상화 과정 | ② 평활

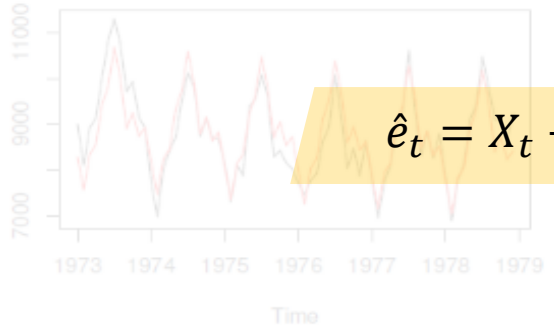
[(C1) 추세&계절성이 모두 존재하는 경우] Classical Decomposition Algorithm



계절성이 제거된 시계열에서
OLS를 통해 다시 추세 추정

$$\hat{m}_t^{new} = \underset{c}{\operatorname{argmin}} \sum_{t=2}^{n} (X_t - \hat{s}_t - c_0 - c_1 t - c_2 t^2 - \cdots - c_p t^p)^2$$

수행 결과 아직 추세가 존재한다면, 위 과정 다시 **반복**



새롭게 추정한 추세와 계절성
제거 후 오차 추정

$$\hat{e}_t = X_t - \hat{m}_t - \hat{s}_t$$

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 : 차분

Classical Decomposition의 한계

[(C1) 추세&계절성이 모두 존재하는 경우] Classical Decomposition Algorithm

초기와 마지막 일부 데이터에 대한 추세 추정 불가

$$\hat{m}_t^{new} = \text{argmin} \sum_{t=2}^{n} (X_t - \hat{s}_t - c_0 - c_1 t - c_2 t^2 - \cdots - c_p t^p)^2$$

데이터의 급격한 증가나 감소가 발생하는 부분에서

너무 강한 smoothing 발생

수행 결과 아직 추세가 존재한다면, 위 과정 다시 반복

$$\hat{e}_t = X_t - \hat{m}_t - \hat{s}_t$$

STL 분해 사용

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition

STL Decomposition

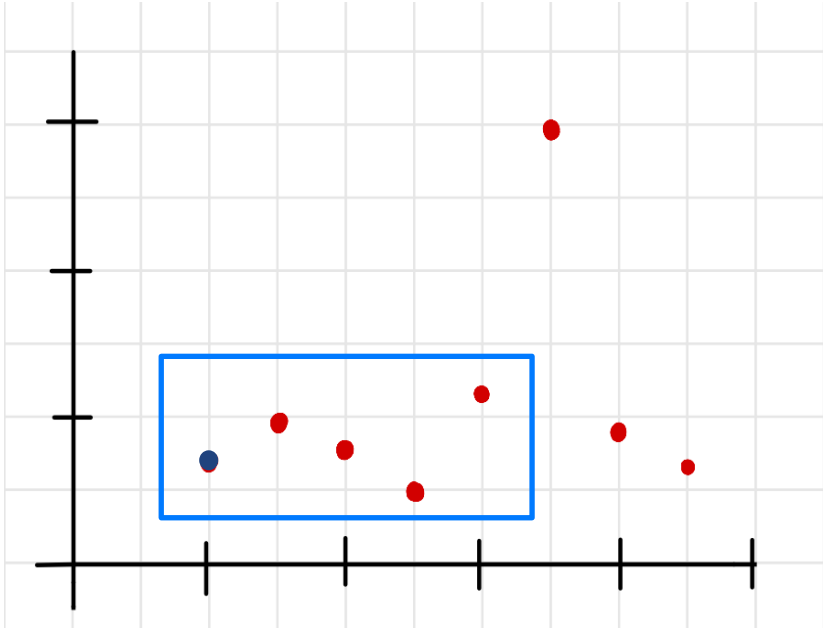
Loess 기법을 중심으로 추세와 계절성분을 구하는 방법

Loess 기법

국소적으로 데이터에 가중치를 부여하여 곡선을 fitting 하는 방법

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition



다음과 같은 데이터에 대해,

focal point를 상정한 뒤

x축 상에서 가장 가까운 데이터 포인트로

window 구성 (window size = 5)

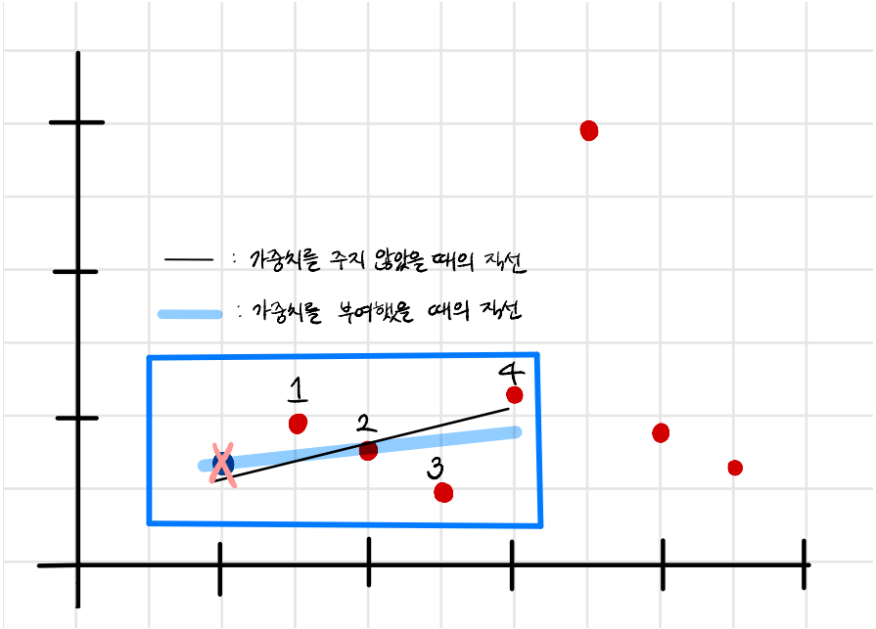
window 내의 데이터로만

fitting한 직선 도출!

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition



다음과 같은 데이터에 대해,

focal point를 상정한 뒤

x축 상에서 가장 가까운 데이터 포인트로

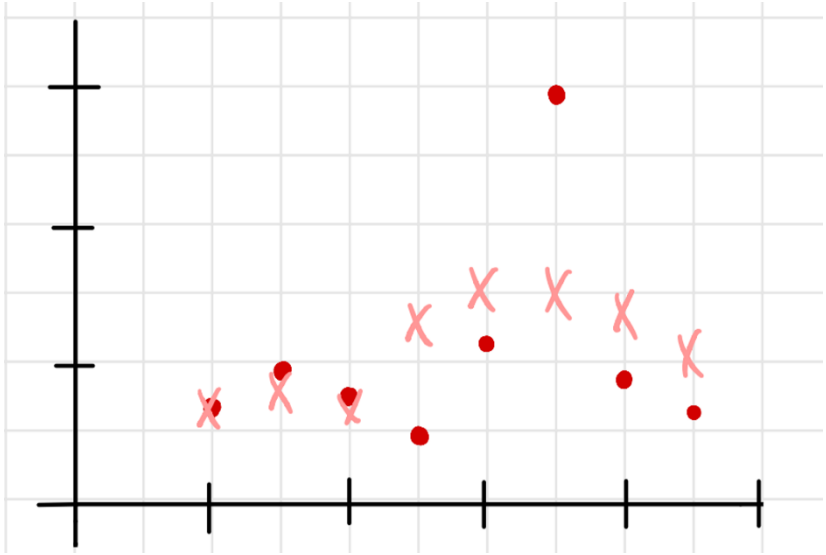
window 구성 (window size = 5)

가중치(x축 상의 거리)를 반영했을 때

곡선이 더 잘 적합되는 것을 확인할 수 있음

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition

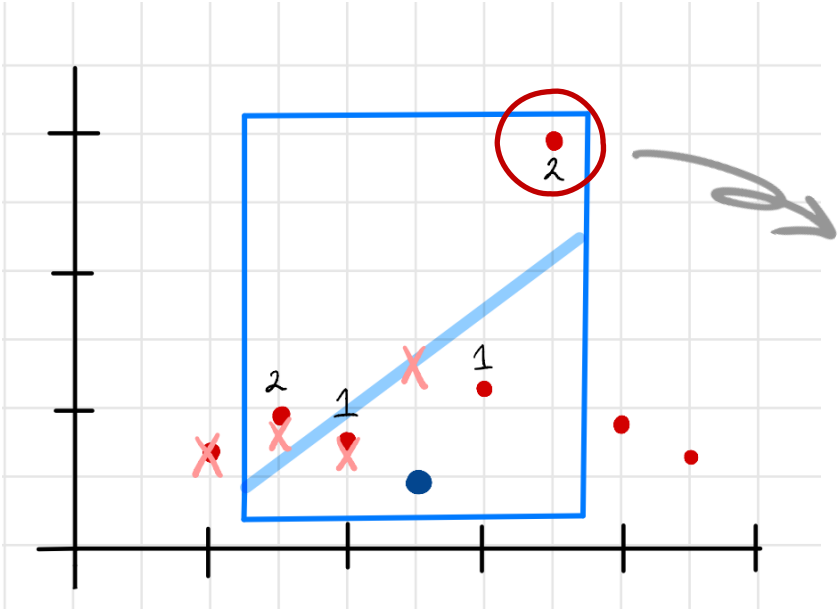


모든 데이터 포인트에 대하여
focal point의 fitting 값($\hat{y}$)을 구함

: 분홍 X로 표현

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition

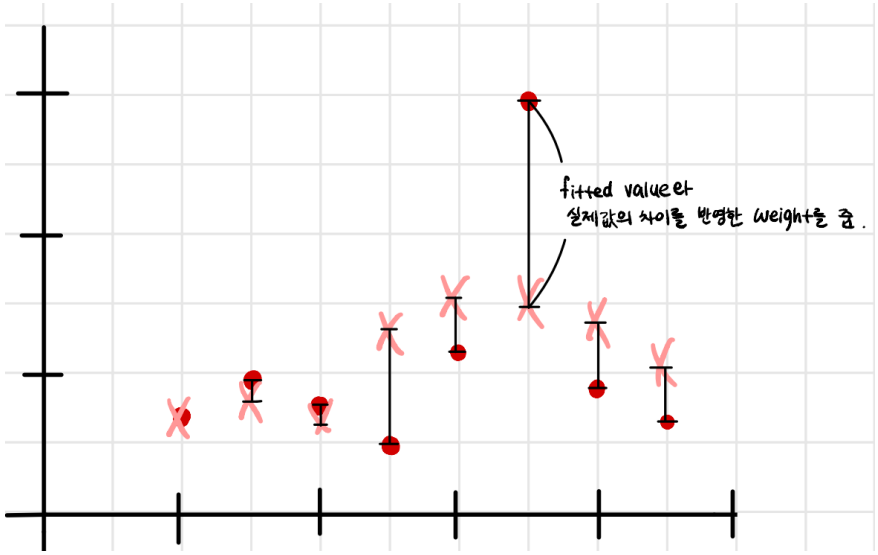


Outlier가 포함된 window의 경우
fitting이 제대로 이루어지지 않을 수 있음

두 번째 가중치 부여!

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition



Fitting된 값과 Original 데이터 포인트 간의 **y값의 차이가 클수록**

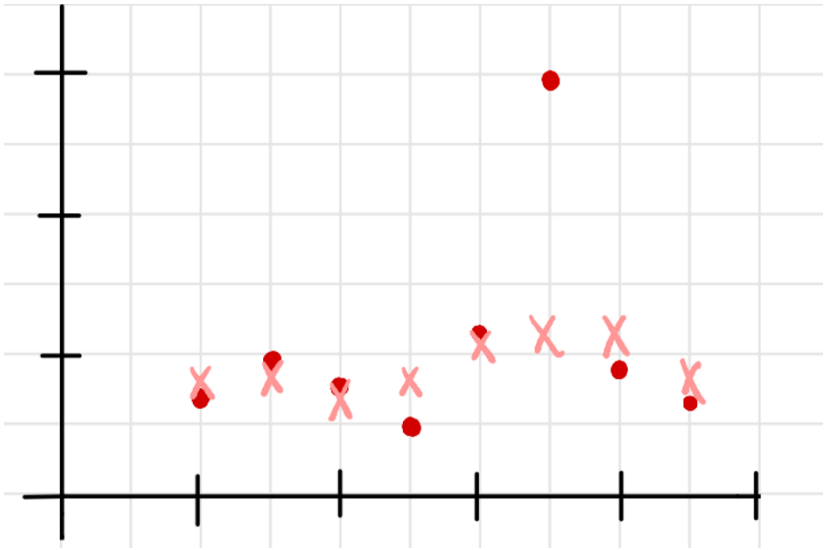
잘 추정하지 못했다는 것을 의미

⇩

더 낮은 가중치 부여

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition

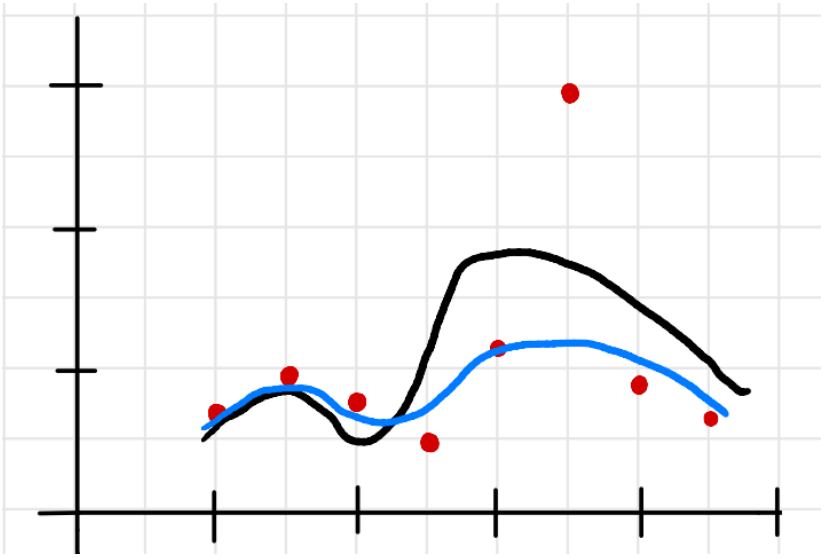


이 과정을 모든 데이터 포인트에
대해 다시 진행

이상치의 영향을 덜 받는
더 부드러운 곡선 생성

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition



두 단계의 가중치를 부여함으로써
Original data를 smooth하게 적합 可

데이터의 **비선형적인 패턴**을 포착할 수 있음

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition

시계열 가정

$$X_t = m_t + s_t + \mathrm{Y_t}\,,\ E(Y_t) = 0$$

계절 주기의 반복 평균을 이용해 초기 계절성분 추정

$$\hat{S}_t = \frac{1}{K}\sum_{k=1}^{K} X_{t+kP}$$

초기 계절성분을
더 부드럽게 만들기 위해!

Loess 평활화 기법 적용

$$\hat{S}_t^{(smoothed)} = \mathrm{Loess}(\hat{S}_t)$$

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition

계절성분 제거 후 Loess로 추세성분 추정

$$Y_t = X_t - \hat{s}_t$$

$$\hat{T}_t = \text{Loess}(R_t)$$

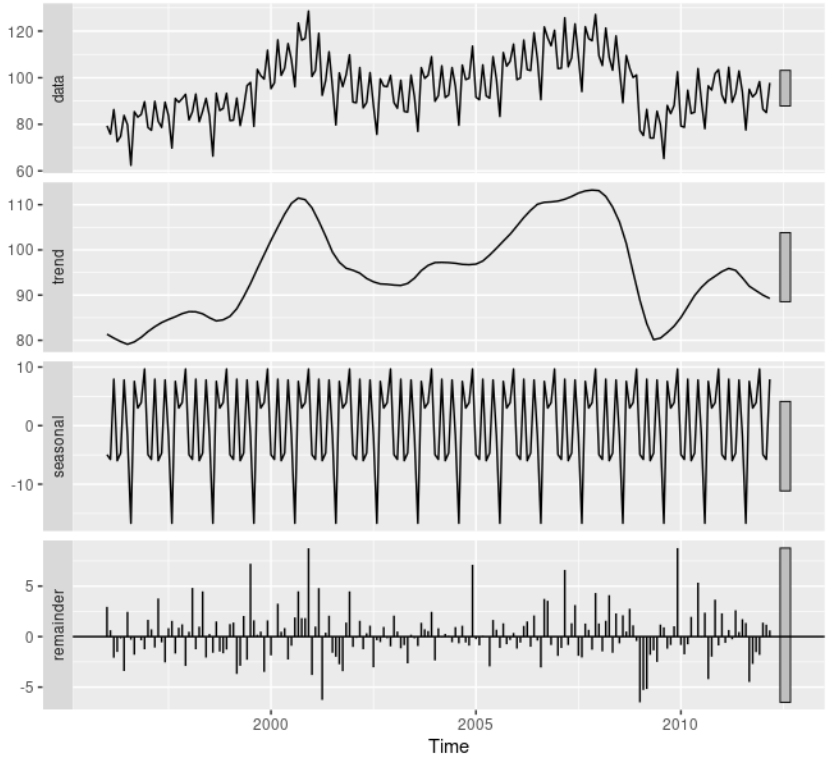
추정된 계절성분과 추세성분을 제거하여 잔차성분 도출

$$\hat{Y}_t = X_t - \hat{T}_t - \hat{s}_t$$

정상화 완료!

평균이 일정하지 않은 경우의 정상화 과정 I ② 평활

[(C2) 추세&계절성이 모두 존재하는 경우] STL Decomposition



계절성분과 추세성분의 추정을 반복 수행하여 업데이트

STL 분해는 **비선형 추세**와 **비선형 계절성**을 포함하는 **복잡한 시계열 데이터 분석**에 유용

평균이 일정하지 않은 경우의 정상화 과정 | ③ 차분

후향연산자 (Backshift Operator)

한 시점 전으로 돌려주는 작용을 하는 연산자

$$BX_t = X_{t-1}$$

후향연산자를 이용하여 차분을 표현할 수 있음

평균이 일정하지 않은 경우의 정상화 과정 | ③ 차분

차분 (Differencing)

관측값들의 차이를 구하는 것으로 차이를 통해 추세와 계절성을 제거하는 방법

후향연산자를 이용해 표현

1차 차분

$$\nabla X_t = X_t - X_{t-1}$$
$$= (1 - B)X_t$$

2차 차분

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1})$$
$$= X_t - 2X_{t-1} - X_{t-2}$$
$$= (1 - B)^2 X_t$$

평균이 일정하지 않은 경우의 정상화 과정 I ③ 차분

[(A) 추세만 존재하는 경우] Differencing

추세를 선형이라고 가정

$$m_t = (c_0 + c_1 t)$$

1차 차분을 통해 추세 제거

$$m_t = m_t - m_{t-1} = (c_0 + c_1 t) - \big(c_0 + c_1(t-1)\big) = c_1$$

시간 t 에 영향을 받지 않는 상수만 남음

⇢ 추세 제거 완료

평균이 일정하지 않은 경우의 정상화 과정 | ③ 차분

[(B) 계절성만 존재하는 경우] Seasonal Differencing

주기가 d인 계절성 가정

$$s_t = s_{t+d}$$

lag-d 차분 연산자

lag-d 차분을 통해 계절성 제거

$$\nabla_d X_t = (1 - B^d) X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + (Error)$$

lag-d 차분 적용 시 **오차항**만 남음

⇢ **계절성 제거 완료**

3 정상화

평균이 일정하지 않은 경우의 정상화 과정 | ③ 차분

[(B) 계절성만 존재하는 경우] Seasonal Differencing

d차 차분과 lag-d 차분의 차이점

주기가 d인 계절성 가정

$$s_t = s_{t+d}$$

d차 차분	lag-d 차분
$\nabla^d = (1-B)^d$	$\nabla_d = X_t - X_{t-d}$
$\nabla_d X_t = (1-B^d)X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + (Error)$	
d번 차분	간격을 d로 두고 차분

lag-d 차분을 통해 계절성 제거

lag-d 차분 적용 시 오차항만 남음

⇢ 계절성 제거 완료

4

정상성 검정

4 정상성 검정

자기공분산함수(ACVF)와 자기상관함수(ACF)

정상성 검정

시계열 자료의 추세와 계절성을 제거하고 남는

오차항이 정상성을 만족하는지를 검정하는 과정

자기공분산함수(ACVF), 자기상관함수(ACF)를

이용하여 오차의 정상성 여부 검정 가능

자기공분산함수(ACVF)와 자기상관함수(ACF)

자기공분산함수

$$\gamma_k = cov(X_t, X_{t+k})$$

자기상관함수

$$\rho_X(\mathrm{h}) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

표본자기공분산함수

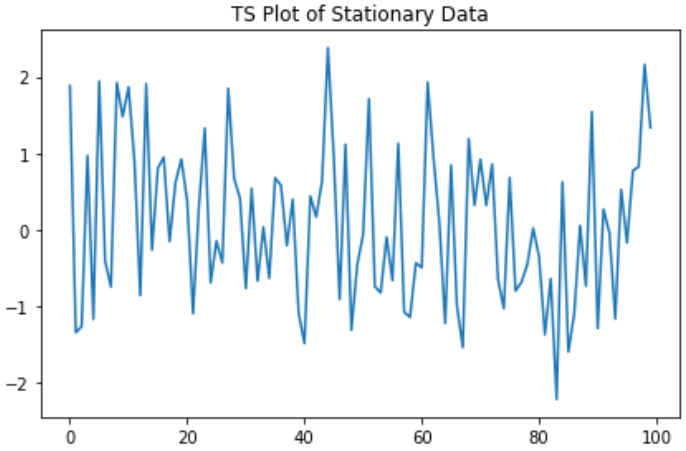
$$\hat{\gamma}_X(h) = \frac{1}{n}\sum_{j=1}^{n-h}(X_j - \bar{X})(X_{j+h} - \bar{X})$$

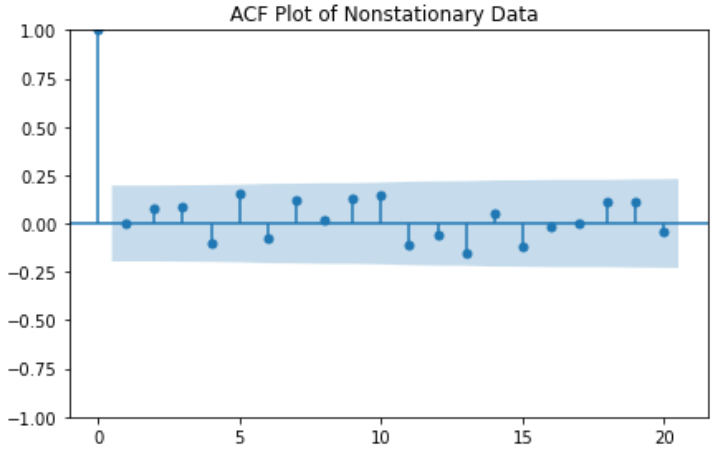
표본자기상관함수

$$\hat{\rho}_X(\mathrm{h}) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)},\ \hat{\rho}(0) = 1$$

4 정상성 검정

자기공분산함수(ACVF)와 자기상관함수(ACF)

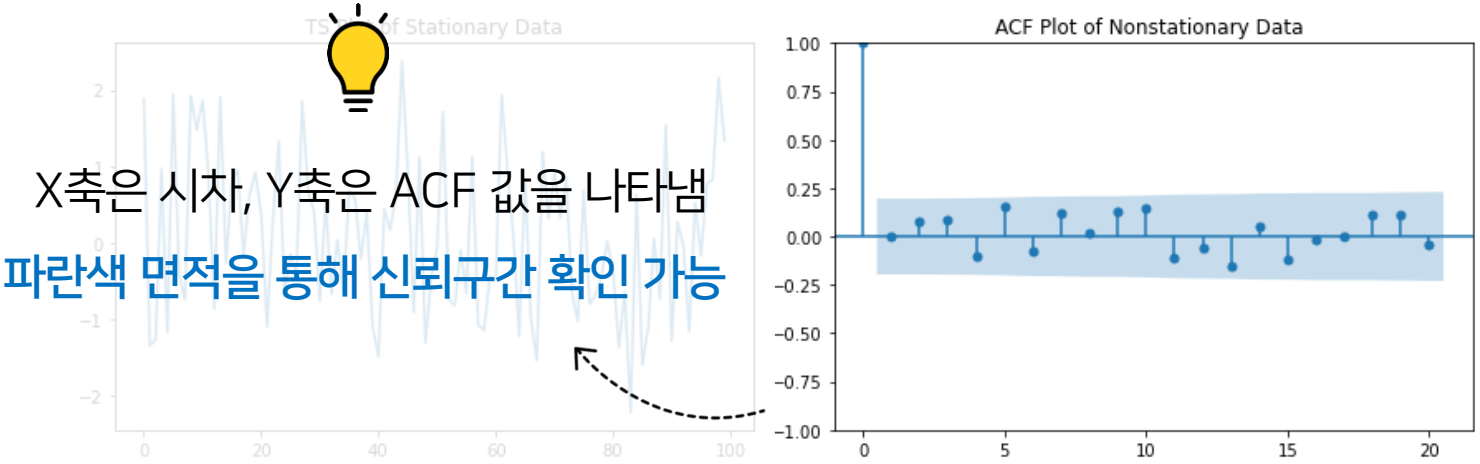




정상성 조건을 만족할 경우 **lag-1 이후 자기상관이 급격히 감소함**

4 정상성 검정

자기공분산함수(ACVF)와 자기상관함수(ACF)



X축은 시차, Y축은 ACF 값을 나타냄

파란색 면적을 통해 신뢰구간 확인 가능

정상성 조건을 만족할 경우 **lag-1 부터 자기상관이 급격히 감소함**

4 정상성 검정

백색잡음

백색잡음 White Noise

자기상관이 존재하지 않는 시계열 자료

$$X_t \sim WN(0, \sigma^2)$$

백색잡음의 세 가지 조건

- 상관관계가 존재하지 않음
- 평균이 0
- 분산 $\sigma^2 < \infty$

4 정상성 검정

백색잡음

백색잡음과 IID의 관계

백색잡음 White Noise

자기상관이 존재하지 않는 시계열 자료

$\mathrm{IID}(0, \sigma^2)$는 백색잡음이지만

백색잡음이라고 $\mathrm{IID}(0, \sigma^2)$ 가 되는 것은 아님!

$$X_t \sim WN(0, \sigma^2)$$

백색잡음의 세 가지 조건

- 상관관계가 존재하지 않음
- 평균이 0
- 분산 $\sigma^2 < \infty$

백색잡음에 Gaussian 가정이 추가된 경우에만 항상 IID 만족

백색잡음 검정

비정상 시계열의 추세와 계절성을 제거했다면,

남아있는 오차항은 WN 조건 혹은 IID 조건을 만족해야 함

이때, σ^2 을 구하기 위해 $\gamma_X(0)$의 추정만 수행하면 됨

백색잡음 검정

비정상 시계열의 추세와 계절성을 제거했다면,

남아있는 오차항은 WN 조건 혹은 IID 조건을 만족해야 함

백색잡음 검정 방법을 알아보자

이때, σ^2 을 구하기 위해 $\gamma_X(0)$의 추정만 수행하면 됨

백색잡음 검정 | ① 자기상관 검정

오차가 백색잡음 $WN(0,1)$을 따른다고 가정하면,

표본자기상관함수 $\hat{\rho}(h)$ 는 $N(0,\frac{1}{n})$에 근사

가설 검정

$$H_0: \rho(\mathrm{h}) = 0 \quad vs \quad H_1: \rho(h) \neq 0$$

귀무가설 : 자기상관 X　　대립가설 : 자기상관 O

4 정상성 검정

백색잡음 검정 | ① 자기상관 검정

오차가 백색잡음 $WN(0,1)$을 따른다고 가정하면,

$|\hat{\rho}(h)| < \frac{1.96}{\sqrt{n}}$이면 귀무가설 H_0 를 기각할 수 없음

가설 검정

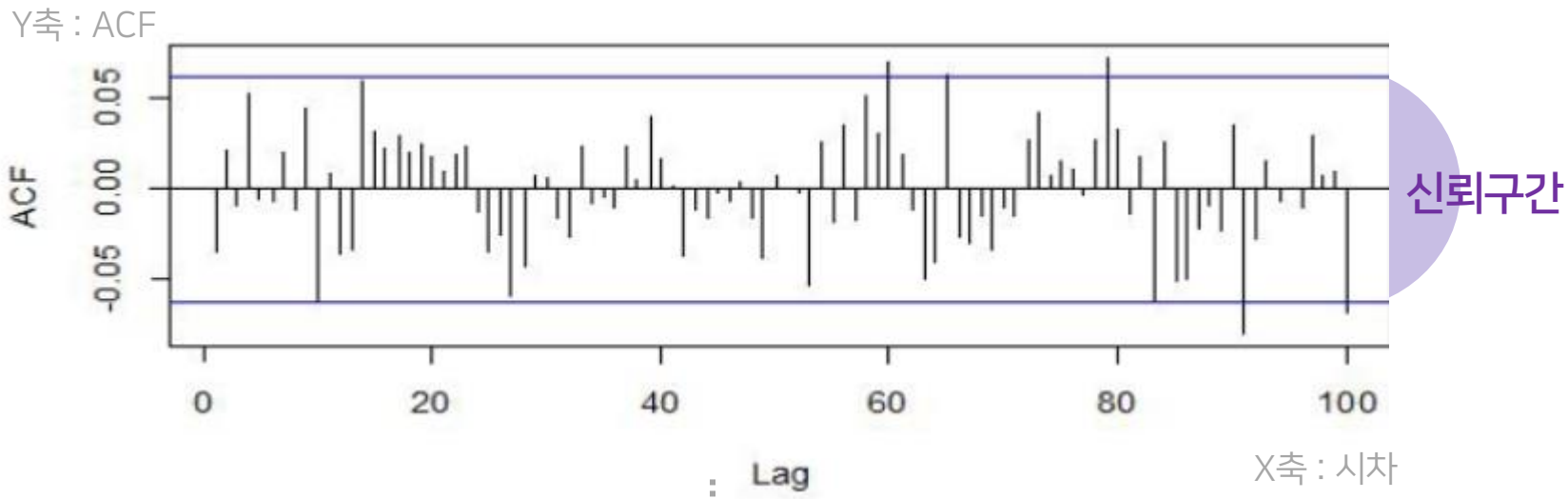
오차항에 자기상관이 없다고 판단

$H_0: \rho(h) = 0$ vs $H_1: \rho(h) \neq 0$

귀무가설 : 자기상관 X　　대립가설 : 자기상관 O

4 정상성 검정

백색잡음 검정 | ① 자기상관 검정



가설검정

ACF plot(correlogram)을 통해 시각적으로 확인 가능

위 그래프에서 대부분의 ACF 값이 신뢰구간을 벗어나지 않으므로

오차항에 자기상관이 존재하지 않는다고 판단 가능

백색잡음 검정 | ② 정규성 검정

가설 설정

H_0: 정규성이 존재함 vs H_1: 정규성이 존재하지 않음

정규성 검정 방법

QQ Plot	KS Plot	Jarque-Bera test
시각적으로 확인할 수 있는 방법	표본과 모집단의 누적확률분포의 유사도를 비교하는 방법	왜도와 첨도를 통해 검정하는 방법

백색잡음 검정 | ③ 정상성 검정

가설 설정

H_0: 정상시계열임 vs H_1: 정상시계열이 아님

정상성 검정 방법

단위근 검정 방법

KPSS test

ADF test

이분산이 존재할 때의 검정 방법

PP test

5

1주차 정리

정리 | 시계열자료

시계열 자료

관측치들 간 **종속성(dependency)**가 존재하는 데이터

규칙요소

추세 / 순환 / 계절성

불규칙요소

우연 변동

덧셈 분해

$$X_t = m_t + s_t + Y_t$$

덧셈 분해는 시계열 분석의 가장 기본적인 프레임워크!

정리 | 정상성

정상성

시계열 자료의 확률적 성질이 **시차에만 의존**

약정상성 조건

$E[X_t] = m, \forall t \in Z$	평균값이 상수로 시점 t에 무관하게 일정
$E[\lvert X_t \rvert]^2 < \infty,\ \forall t\ \in Z$	2nd-moment가 존재하며, 시점 t에 무관하게 일정
$\gamma_X(r,s)=\gamma_X(r+h,s+h),$ $\forall\, r,s,h \in Z$	공분산은 시차 h에 의존, 시점 t와 무관

실제 시계열 분석에서는 약정상성 개념 적용

정리 | 정상화

정상화

비정상 시계열을 **정상 시계열로 변환**

분산 일정 X

로그(Log) 변환

제곱근 (Square Root) 변환

Box-Cox 변환

VS

평균 일정 X

회귀 (Regression)

평활 (Smoothing)

차분 (Differencing)

정리 | 정상성 검정

자기공분산함수 Autocovariance Function

$$\gamma_k = cov(X_t, X_{t+k})$$

자기상관함수 Autocorrelation Function

$$\rho_X(\mathrm{h}) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

ACF Plot을 통해 시계열자료의 정상성 여부 파악 가능

정리 | 정상성 검정

백색잡음 $WN(0, \sigma^2)$

자기 상관이 존재하지 않는 시계열

자기상관 검정	ACF plot
정규성 검정	QQ plot / KS Test / Jarque-Bera Test
정상성 검정	KPSS Test / ADF Test / PP Test

다음 주 예고

1. AR, MA, ARMA
2. ARIMA, SARIMA
3. ARCH, GARCH
4. 추석 ><

우리는~ 시결핑~~

어?시계열이다

다소곳맨 철석

시결팀원들 다들 소식좌 ㅜㅜ