시계열자료분석팀

5팀

장다연

심현구

윤세인

이동기

천예원

INDEX

1

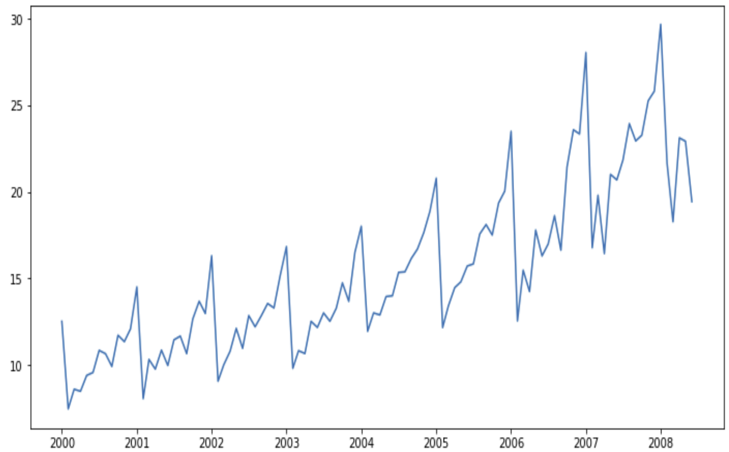
시계열 자료 분석

1 시계열 자료 분석

시계열 자료란?

시계열 자료 (Time Series)

시간 순서에 따라 관측된 자료의 집합



$$\{X_t, t = 1,2,3, \ldots\}$$

t : 시점, t의 종류에 따라 연속형, 이산형 자료로 분류됨.

시계열 자료의 특징

연관성

(dependency)

시간의 흐름이 반영되어

관측치들 사이에 연관성이 존재함

결합분포

(joint distribution)

특정 시점 X

전체 시점에서의 관측치 집합 모두 고려

1 시계열 자료 분석

시계열 자료의 특징

연관성

(dependency)

시간의 흐름이 반영되어

관측치들 사이에 연관성이 존재함

결합분포

(joint distribution)

특정 시점에서의 관측치 집합만을 고려 X

전체 시점에서의 관측치 집합 모두 고려

독립성 조건을 만족하지 않는 특징으로 인해

일반적인 선형회귀 방법 사용 X

데이터의 특성을 반영할 수 있는 특별한 분석법,

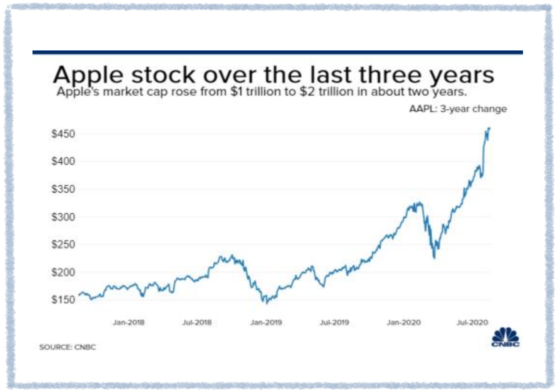
시계열 분석 필요

시계열 자료의 구성 요소

추세 (trend)

시간에 따라 증감하는 변동

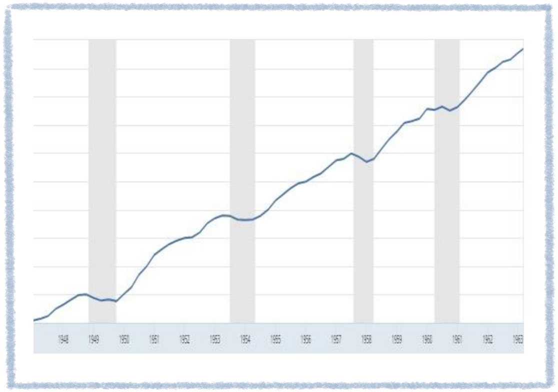
특별한 충격이 없는 한 지속됨.



순환 변동 (cycle)

일정 주기를 가지는 변동

주기가 규칙적이진 않음.



1 시계열 자료 분석

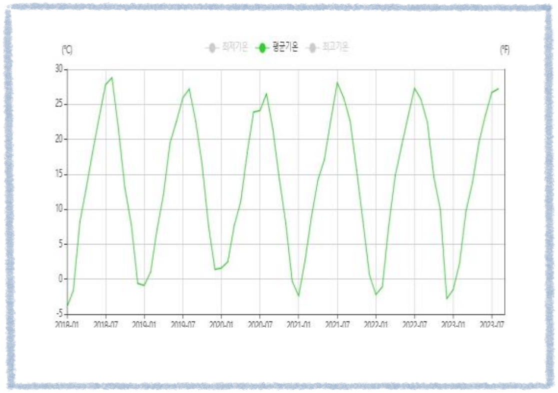
시계열 자료의 구성 요소

계절성 (seasonality)

규칙적인 주기를 가지는 변동

주별, 월별, 계절별과 같이

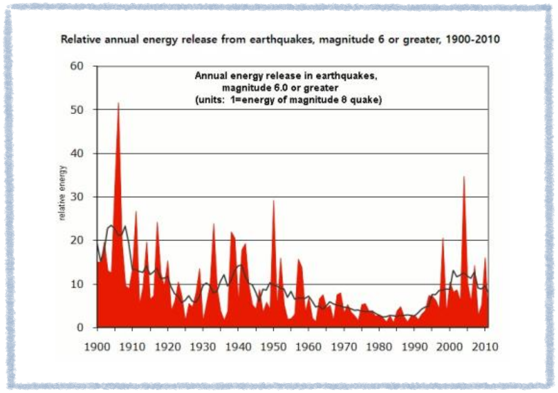
특정 시간 간격을 가짐



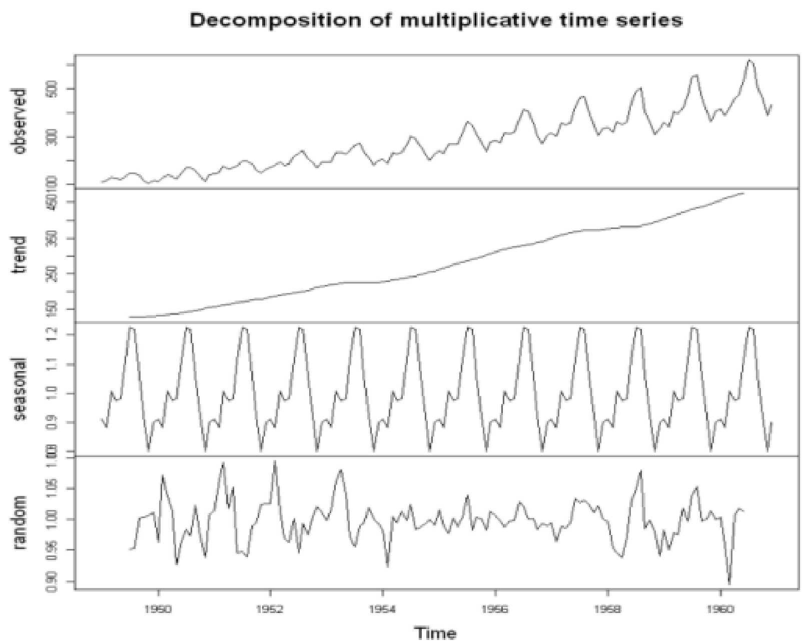
우연변동 (Random fluctuation)

무작위적임

일정한 규칙성을 보이지 않음



시계열 분해



원 시계열 데이터

분해된 시계열 데이터

시계열 분석에서는 위 그림과 같이 4가지 구성요소를 분해

시계열 분해

시계열 분해 (Time Series Decomposition)

시계열 자료를 비정상 부분(non-stationary part)과
정상 부분(stationary part)으로 분해하는 작업

추세(m_t)와 계절성(s_t)은 비정상 부분, 오차(Y_t)는 정상 부분

덧셈 분해

$$X_t = m_t + s_t + Y_t$$

곱셈 분해

$$X_t = m_t * s_t * Y_t$$

데이터에 0이 포함되어 있는지 확인 必

추세와 계절성을 제거하여 오차를 이용해 예측 모델링 진행함.

시계열 분해

시계열 분해(Time Series Decomposition)

정상 부분(stationary part)으로 분해하는 작업

덧셈 분해 vs 곱셈 분해

덧셈 분해는 추세와 계절성을 별개의 구성 요소로 보지만,

곱셈 분해는 **추세에 따라 계절성이 변화함**을 가정함

클린업에서는 주로 **덧셈 분해**에 대해 다룰 예정임

덧셈 분해

$X_t = m_t + s_t + Y_t$

곱셈 분해

$X_t = m_t * s_t * Y_t$

추세와 계절성을 제거하여 오차를 이용해 예측 모델링 진행함.

2

정상성

2 정상성

정상성이란?

정상성 (Stationarity)

시계열 자료의 **확률적 성질이** 시점 t에 의존하지 않고 **시차 lag에만 의존**하는 특성

강정상성

(Strict Stationarity)

약정상성

(Weak Stationarity)

2 정상성

강정상성

강정상성 (Strict Stationarity)

$$\left(X_{t_1}, \dots, X_{t_n}\right) \overset{d}{=} \left(X_{t_1+h}, \dots, X_{t_n+h}\right)$$

즉, 일정한 시차 간격을 가지는 관측치 집합들이

모두 같은 분포를 따른다는 것!

이렇게 엄격한 조건을 만족하는 시계열 데이터는 많지 않음

강정상성

강정상성에 정규성(Gaussianity)를 가정한다면?

$$\left(X_{t_1}, \dots, X_{t_n}\right) \sim MVN(\mu, \Sigma)$$

평균 벡터인 μ와 공분산 행렬 Σ만 추정해서 전체 분포를 구할 수 있음

1. 확률 변수의 기댓값은 상수

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

2. 공분산은 시차에 의존

$$Cov(X_r, X_s)$$

$$= Cov(X_{r+h}, X_{s+h}), \forall r, s, h \in \mathbb{Z}$$

강정상성

강정상성에 정규성(Gaussianity)를 가정한다면?

$$\left(X_{t_1}, \dots, X_{t_n}\right) \sim MVN(\mu, \Sigma)$$

정규성을 가정하여 시차에만 의존하던 분포를 공분산으로 완화하였지만,

여전히 엄격한 가정이기에 **현실에서는 약정상성을 이용**

평균 벡터인 μ와 공분산 행렬 Σ만 추정해서 전체 분포를 구할 수 있음

1. 확률 변수의 기댓값은 상수

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

2. 공분산은 시차에 의존

$$Cov(X_r, X_s)$$
$$= Cov(X_{r+h}, X_{s+h}), \forall r, s, h \in \mathbb{Z}$$

약정상성

약정상성 조건

1) $E[|X_t|]^2 < \infty, \quad \forall t \in \mathbb{Z}$

2차 적률이 존재하고 시점 t에 관계없이 일정함.

2) $E[|X_t|]^2 = \mathrm{m}, \quad \forall t \in \mathbb{Z}$

평균이 상수로 시점 t에 관계없이 일정함.

3) $\gamma_X(r,s) = \gamma_X(r+h, s+h),$

$\forall r, s, h \in \mathbb{Z}, \qquad (\gamma_X(r,s) := Cov(X_r, X_s))$

공분산은 시차 h에 의존하며 시점 t와 무관함.

클린업에서 다루는 모든 시계열 자료는 약정상성을 만족하는 시계열임.

3

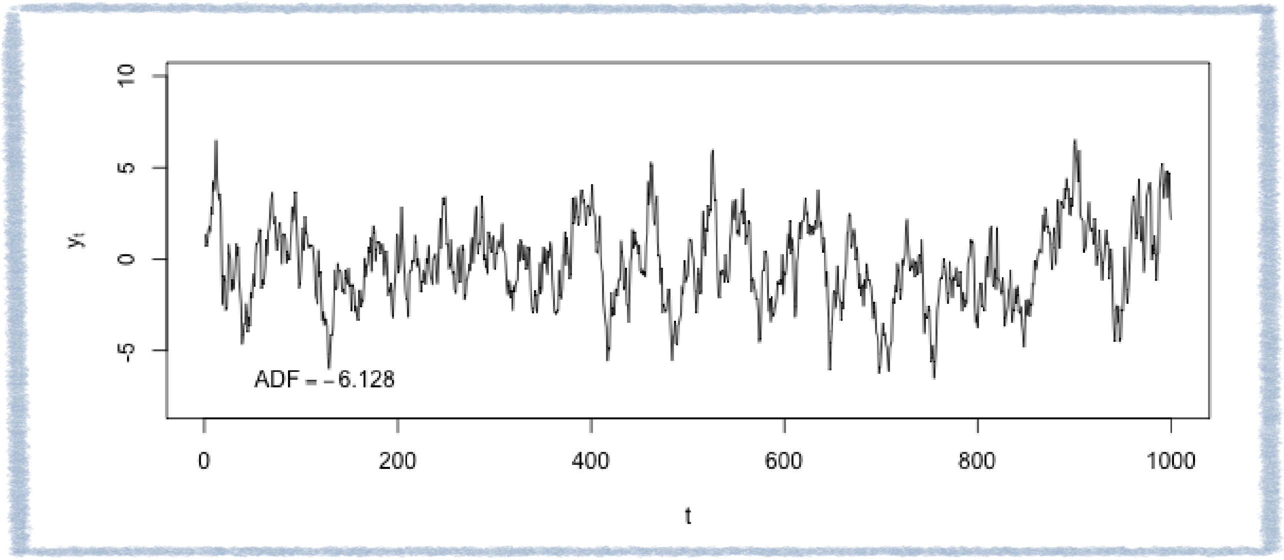
정상화

3 정상화

정상 시계열과 비정상 시계열

✔ 시계열 플랏(TS plot)을 통해 시각적으로 확인 가능

정상 시계열

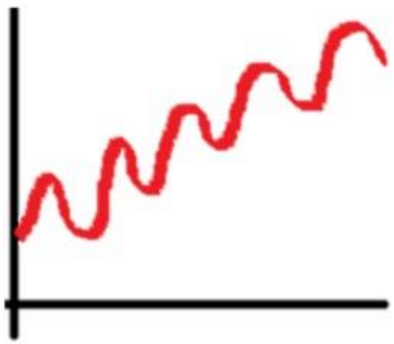


특별한 추세나 계절성이 보이지 않으며, 평균과 분산 역시 일정함.

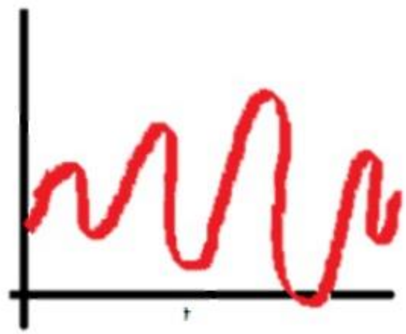
정상 시계열과 비정상 시계열

- 시계열 플랏(TS plot)을 통해 시각적으로 확인 가능

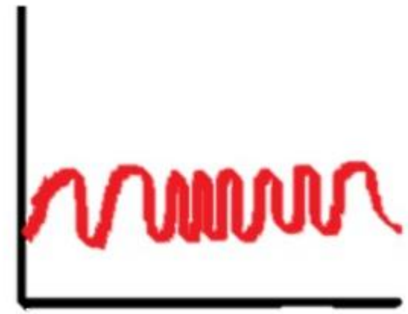
비정상 시계열



평균이 일정하지 않음



분산이 일정하지 않음



공분산이 시점에 의존함

정상 시계열과 비정상 시계열

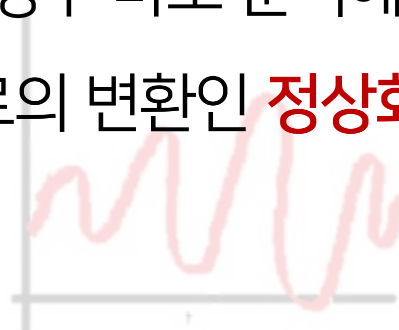
✔ 시계열 플랏(TS plot)을 통해 시각적으로 확인 가능

비정상 시계열

비정상 시계열의 경우 바로 분석에 사용할 수 없으므로

정상 시계열로의 변환인 **정상화 과정**이 필요함.

평균이 일정하지 않음



분산이 일정하지 않음

공분 … 함

3 정상화

정상화가 필요한 이유

독립성 조건의 붕괴

시계열 자료는 오차의 독립성 조건을 만족하지 않기 때문에
오차의 독립성을 가정하는 다양한 통계분석기법을 활용할 수 없음

시계열의 분해와 정상화를 통해 상관관계가 존재하지 않는 오차만 남긴 후 모델링을 진행

안정적이고 정확한 예측

정상성을 만족하지 않는 데이터를 사용하면
데이터를 설명하는 모델의 정확도가 시점에 따라 달라질 수 있음

3 정상화

경우에 따른 정상화

분산이 일정하지 않은 경우

분산이 시점에 의존하지 않고 일정하다는

약정상성 조건에 위배되는 상황

분산 안정화 변환

- 로그 (log) 변환
- 루트 (square root) 변환
- Box-Cox 변환

평균이 일정하지 않은 경우

추세 또는 계절성이 존재하여

비정상 부분을 제거해야 하는 상황

- 회귀 (Regression)
- 평활 (Smoothing)
- 차분 (Differencing)

3 정상화

경우에 따른 정상화

분산이 일정하지 않은 경우

분산이 시점에 의존하지 않고 일정하다는
약정상성 조건에 위배되는 상황

↓

분산 안정화 변환

- 로그 (log) 변환
- 루트 (square root) 변환
- Box-Cox 변환

분산 안정화 변환

Variance Stabilizing Transformation(VST)

로그 변환

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

제곱근 변환

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

Box-Cox 변환

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

3 정상화

경우에 따른 정상화

분산이 일정하지 않은 경우

분산이 시점에 의존하지 않고 일정하다는
약정상성 조건에 위배되는 상황

- 로그 (log) 변환
- 루트 (square root) 변환
- Box-Cox 변환

분산 안정화 변환

평균이 일정하지 않은 경우

추세 또는 계절성이 존재하는
비정상 부분을 제거해야 하는 상황

- 회귀 (Regression)
- 평활 (Smoothing)
- 차분 (Differencing)

평균이 일정하지 않은 경우의 정상화 - ① 회귀

a. 추세만 존재하는 경우 : Polynomial Regression

기본 가정: 시계열이 추세만 가진다

$$X_t = m_t + Y_t$$

m_t : 추세, Y_t : 정상성을 만족하는 오차

STEP 1) 추세 성분을 시간에 대한 **선형회귀식**으로 나타냄

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

m_t : 추세, t : 시간

평균이 일정하지 않은 경우의 정상화 - ① 회귀

a. 추세만 존재하는 경우 : Polynomial Regression

STEP 2) 선형회귀식의 계수를 **최소제곱법(OLS)**을 통해 추정

$$\left(\widehat{c_o}, \dots, \widehat{c_p}\right) = \underset{c}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - m_t)^2$$

STEP 3) 추정한 추세를 비정상 시계열 (Xt)에서 제거

$$X_t - \widehat{m_t} \approx Y_t$$

3 정상화

평균이 일정하지 않은 경우의 정상화 - ① 회귀

b. 계절성만 존재하는 경우 : Harmonic Regression

영감… 미안해요.. 다시는 안뽑기로 했는데…
울어라, 지옥참마도!

할모닉 ㅎ

기본 가정: 시계열이 주기가 d인 계절성만을 가진다

$$X_t = s_t + Y_t$$

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

Step 1) 계절 성분 s_t를 시간 t에 대한 회귀식으로 표현

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

s_t, λ_t : constant

평균이 일정하지 않은 경우의 정상화 - ① 회귀

b. 계절성만 존재하는 경우 : Harmonic Regression

STEP 2) 적절한 λj 와 k 선택 후, OLS를 통해 a_j 와 b_j 추정

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

STEP 3) 추정한 계절성을 시계열에서 제거

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

평균이 일정하지 않은 경우의 정상화 - ② 회귀

b. 계절성만 존재하는 경우 : Harmonic Regression

적절한 λ_j 와 k의 선택

STEP 2) 적절한 λ_j 와 k 선택 후, OLS를 통해 a_j 와 b_j 추정

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

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

① 주기 반복 횟수 $f_1 = \left[\frac{n}{d}\right] \rightarrow f_j = j * f_1$

② $\lambda_j = f_j * (2\pi/n)$

STEP 3) 추정한 계절성을 시계열에서 제거

k: 주로 1~4 사이의 값을 사용

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

n: 데이터 개수, d: 주기

평균이 일정하지 않은 경우의 정상화 - ① 회귀

b. 계절성만 존재하는 경우 : Harmonic Regression

STEP 2) 적절한 λj 와 k 선택 후, OLS를 통해 a_j 와 b_j 추정

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

STEP 3) 추정한 계절성을 시계열에서 제거

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

평균이 일정하지 않은 경우의 정상화 - ① 회귀

c. 추세와 계절성 모두 존재하는 경우

Polynomial Regression과 Harmonic Regression을 차례대로 진행

↓

남아 있는 추세가 보인다면 동일한 과정을 반복하여 제거

평균이 일정하지 않은 경우의 정상화 - ② 평활

회귀 방법은 전체 데이터를 한번에 처리

따라서 국소적 변동이 존재하는 경우 부적합

평활법 (smoothing)

시계열 자료를 여러 구간으로 나눈 뒤

구간의 평균들로 추세를 추정하는 방법

3 정상화

평균이 일정하지 않은 경우의 정상화 - ② 평활

a. 추세만 존재하는 경우

이동평균 평활법

일정 구간 마다 평균을 계산해
추세를 추정, 제거 하는 방법

지수 평활법

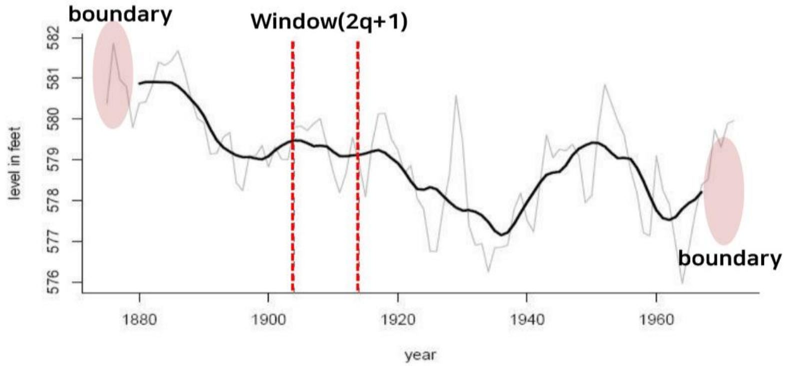
t 시점까지의 관측값을 이용해
추세를 추정, 제거하는 방법

평균이 일정하지 않은 경우의 정상화 - ② 평활

a. 추세만 존재하는 경우 : 이동평균 평활법

Step 1) 길이가 2q+1인 구간의 평균 구하기

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



평균이 일정하지 않은 경우의 정상화 - ② 평활

a. 추세만 존재하는 경우 : 이동평균 평활법

Step 2) 추세 성분 m_t를 대입하여 구간의 평균 W_t를 근사적으로 추세 m_t 와 같게 만들기

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

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

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

Step 3) 추세부분만 남은 W_t를 X_t에서 제거

이동평균 평활법의 한계

데이터의 맨 앞 q개와 맨 뒤 q개의 boundary 추정 불가

현실에서 미래 관측값 사용 불가

과거 데이터만을 활용하는 **지수 평활법** 활용

평균이 일정하지 않은 경우의 정상화 - ② 평활

a. 추세만 존재하는 경우 : 이동평균 평활법

Step 2) 추세 성분 m_t를 대입하여 구간의 평균 W_t를 근사적으로 추세 m_t와 같게 만들기

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

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

Step 3) 추세부분만 남은 W_t를 X_t에서 제거

평균이 일정하지 않은 경우의 정상화 - ② 평활

a. 추세만 존재하는 경우 : 지수 평활법

Step 1) 추세 추정

$$\hat{m}_1 = X_1$$

$$\hat{m}_2 = aX_2 + (1-a)\hat{m}_1 = aX_2 + (1-a)X_1$$

$$\hat{m}_3 = aX_3 + (1-a)\hat{m}_2 = aX_3 + a(1-a)X_2 + (1-a)^2X_1$$

$$\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1} = \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1}X_1$$

과거의 관측치일수록

가중치의 값이 지수적으로 감소

$a \in [0,1]$, 과거 관측치에 대한 가중치

Step 2) 추세 제거

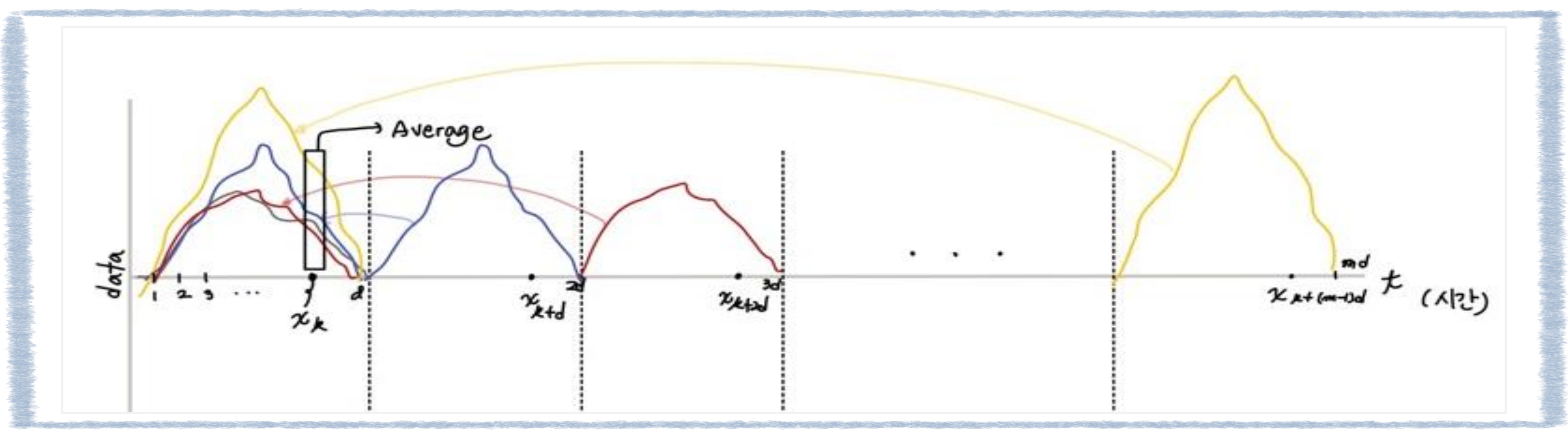
평균이 일정하지 않은 경우의 정상화 - ② 평활

b. 계절성만 존재하는 경우 : Seasonal Smoothing

Seasonal Smoothing

주기가 d인 시계열 자료에서 주기만큼의 데이터를 모두 겹친 후,

겹친 값들의 평균으로 계절성을 추정하는 방법



평균이 일정하지 않은 경우의 정상화 - ② 평활

b. 계절성만 존재하는 경우 : Seasonal Smoothing

STEP 1) 계절성분($\hat{s}_k$) 추정

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

STEP 2) 추정된 계절성분을 다른 주기에 적용, 전체 계절성 추정

STEP 3) 계절성을 시계열 자료에서 제거

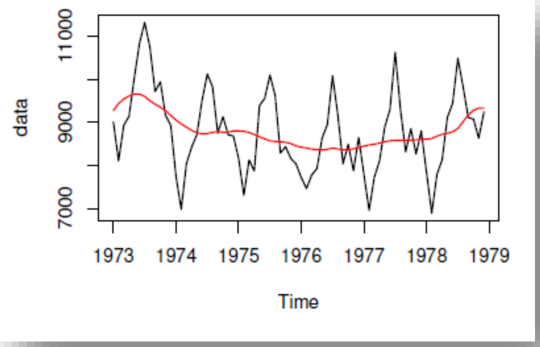
평균이 일정하지 않은 경우의 정상화 - ② 평활

c. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 1) MA filter를 사용해 추세 추정

$$if\ d = 2q\ (even), \qquad \hat{m}_t = \frac{0.5X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + 0.5X_{t+q}}{2q}$$

$$if\ d = 2q + 1\ (odds), \qquad \hat{m}_t = \frac{X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + X_{t+q}}{2q + 1}$$

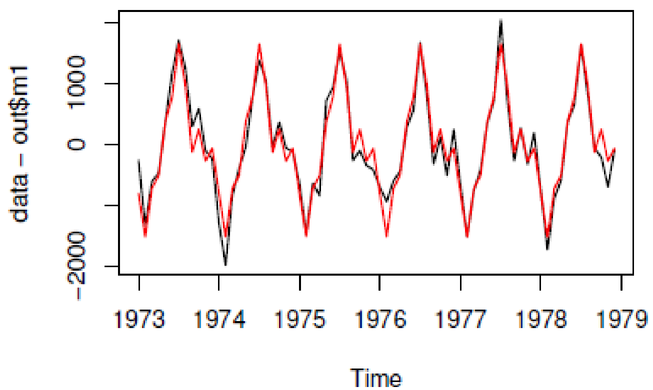


평균이 일정하지 않은 경우의 정상화 - ② 평활

c. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 2) 추세 제거 후,

Seasonal Smoothing을 통해 계절성 추정



Step 3) 계절성 제거 후,

OLS를 활용해 추세 재추정 후 제거

OLS 대신 Smoothing 사용 가능,

일반적인 경우 OLS 사용

평균이 일정하지 않은 경우의 정상화 - ② 평활

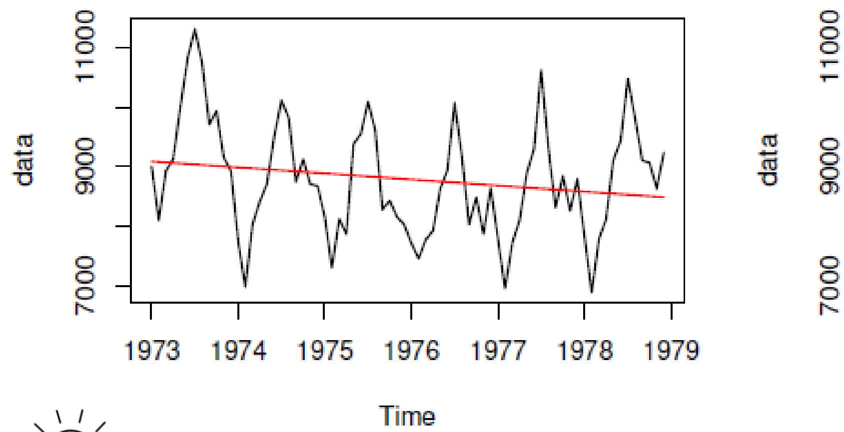
c. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 2) 추세 제거 후,

Seasonal Smoothing을 통해 계절성 추정

Step 3) 계절성 제거 후,

OLS를 활용해 추세 재추정 후 제거



OLS 대신 Smoothing 사용 가능,

일반적인 경우 OLS 사용

3 정상화

평균이 일정하지 않은 경우의 정상화 - ② 평활

c. 추세와 계절성 모두 존재하는 경우 : Classical Decomposition Algorithm

Step 2) 추세 제거 후,

Seasonal Smoothing을 통해 계절성 추정

Step 3) 계절성 제거 후,

OLS를 활용해 추세 재추정 후 제거

잔여 추세가 존재 한다면

[1]~[3] 과정 반복해 추세 제거

OLS 대신 Smoothing 사용 가능,

일반적인 경우 OLS 사용

평균이 일정하지 않은 경우의 정상화 - ③ 차분

이 '차이'를 이용하여 추세와 계절성을 제거

차분 (Differencing)

관측값들의 차이를 구하는 것

후향 연산자를 사용하여 차분을 진행

후향 연산자

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

관측값을 한 시점 전으로

돌려주는 역할을 하는 연산자

평균이 일정하지 않은 경우의 정상화 - ③ 차분

이 '차이'를 이용하여 추세와 계절성을 제거

차분 (Differencing)

관측값들의 차이를 구하는 것

후향 연산자를 사용하여 차분을 진행

1차 차분

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

2차 차분

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

후향 연산자

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

관측값을 한 시점 전으로

돌려주는 역할을 하는 연산자

평균이 일정하지 않은 경우의 정상화 - ③ 차분

a. 추세만 존재하는 경우 – 차분(Differecing)

$m_t = c_0 + c_1 t$ 를 가정한 후 1차 차분 적용

$$\nabla m_t = (c_0 + c_1 t) - \big(c_0 + c_1 (t-1)\big) = c_1$$

t에 영향을 받지 않는 상수

k차 차분하면 k차 추세가 제거됨

$$\nabla^k X_t = k!\, c_k + \nabla^k Y_t = const, +error$$

평균이 일정하지 않은 경우의 정상화 - ③ 차분

a. 추세만 존재하는 경우 – 차분(Differecing)

$m_t = c_0 + c_1 t$ 를 가정한 후 1차 차분 적용

$$\nabla m_t = (c_0 + c_1 t) - \big(c_0 + c_1(t-1)\big) = c_1$$

t에 영향을 받지 않는 상수

k차 차분하면 k차 추세가 제거됨

$$\nabla^k X_t = k!\, c_k + \nabla^k Y_t = const, +error$$

$$\nabla^k X_t = k!\, c_k + \nabla^k Y_t = const, +\ error$$

차분을 통해 추세를 제거하는 방법은 직관적이지만,

위 식과 같이 오차까지 차분되어 식이 복잡해질 수 있다는 단점!

3 정상화

평균이 일정하지 않은 경우의 정상화 - ③ 차분

b. 계절성만 존재하는 경우 – Seasonal Differencing

계절성만 존재하는 경우 lag-d differencing을 통해 계절성을 제거함

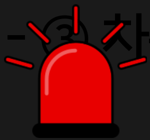
$$\nabla_d X_t = (1 - B^d) X_t , \qquad t = 1, \dots, n$$

$s_t = s_{t+d}$를 가정하고 lag-d 차분을 적용하면…

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

오차항만 남아 계절성이 제거됨!

평균이 일정하지 않은 경우의 정상화 - ③ 차분



b. 계절성만 존재하는 경우 – Seasonal Differencing

차분의 표현법 차이

계절성만 존재하는 경우 lag-d differencing을 통해 계절성을 제거함

$$\nabla_d X_t = (1 - B^d)X_t , \qquad t = 1, \dots, n$$

d차 차분

$$\boxed{\nabla^d} = (1 - \boxed{B)^d}$$

lag-d 차분

$$\boxed{\nabla_d} = (1 - \boxed{B^d})$$

$s_t = s_{t+d}$를 가정하고 lag-d 차분을 적용하면 :

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

오차항만 남아 계절성이 제거됨!

평균이 일정하지 않은 경우의 정상화 - ③ 차분

c. 추세와 계절성 모두 존재하는 경우 : Lag-d 차분 + p차 차분

STEP 1) 계절차분

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

계절성은 사라지고 추세만 남게 됨

평균이 일정하지 않은 경우의 정상화 - ③ 차분

c. 추세와 계절성 모두 존재하는 경우 : Lag-d 차분 + p차 차분

STEP 2) 남아있는 추세 제거를 위한 차분 진행

$$\nabla_d = (1 - B^d) = (1 - B)(1 + B + \cdots + B^{d-1})$$

계절 차분에 $(1 - B)$가 포함

→ (p-1)차 차분을 진행

평균이 일정하지 않은 경우의 정상화 - ③ 차분

c. 추세와 계절성 모두 존재하는 경우 : Lag-d 차분 + p차 차분

STEP 2) 남아있는 추세 제거를 위한 차분 진행

$$\nabla_d = \left(1 - B^d\right) = \boldsymbol{(1 - B)}(1 + B + \cdots + B^{d-1})$$

계절 차분에 $(1 - B)$가 포함

→ (p-1)차 차분을 진행

4

정상성 검정

정상성 검정

시계열 데이터에서 비정상 부분이 성공적으로 제거되었다면

정상성을 만족하는 오차 Y_t**만**이 데이터에 남아 있어야 함

① 자기공분산함수 (ACVF)

② 자기상관함수 (ACF)

를 이용하여 정상성 만족 여부를 확인

자기공분산함수(ACVF), 자기상관함수(ACF)

자기공분산함수(ACVF)

$$\gamma_x(h) = \mathrm{Cov}(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

자기상관함수(ACF)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_t, X_{t+h}) = \frac{\mathrm{Cov}(X_t , X_{t+h})}{\sqrt{\mathrm{var}(X_t)}\sqrt{\mathrm{var}(X_{t+h})}}$$

자기공분산함수(ACVF), 자기상관함수(ACF)

자기공분산함수(ACVF)

$$\gamma_x(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

같은 결합 분포 내에서 시차를 가지는 변수들의 **공분산**을 구하는 식이므로

위 식을 통해 $\gamma_x(0) = Var(X_t)$라는 것을 알 수 있음

자기상관함수(ACF)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t+h}) = \frac{Cov(X_t, X_{t+h})}{\sqrt{var(X_t)}\sqrt{var(X_{t+h})}}$$

4 정상성 검정

자기공분산함수(ACVF), 자기상관함수(ACF)

자기공분산함수(ACVF)

$$\gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

위 식을 통해 $\gamma_X(0) = Var(X_t)$라는 것을 알 수 있음

자기상관함수(ACF)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t+h}) = \frac{Cov(X_t, X_{t+h})}{\sqrt{var(X_t)}\sqrt{var(X_{t+h})}}$$

시계열데이터에서 **샘플링**된 데이터의 정상성을 검정하기 위해서는

표본자기공분산함수(SACVF)와 표본자기상관함수(SACF)를 사용함!

표본자기공분산함수(SACVF), 표본자기상관함수(SACF)

표본자기공분산함수(SACVF)

$$\hat{\gamma}_x(h) = \frac{1}{n}\Sigma_{j=1}^{n-h}\left(X_j - \overline{X}\right)\left(X_{j+h} - \overline{X}\right)$$

표본자기상관함수(SACF)

$$\widehat{\rho_x}(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}, \qquad \hat{\rho}(0) = 1$$

백색잡음 (White Noise)

White Noise Process (백색잡음)

자기상관이 존재하지 않는 시계열 데이터

1. 시계열 $\{X_t\}$의 **평균이 0, 분산이 $\sigma^2 < \infty$**인 경우

2. 시계열 $\{X_t\}$에 **상관관계가 존재하지 않는 경우**

시계열 $\{X_t\}$는 $\{X_t\} \sim WN(0, \sigma^2)$로 표현되는

백색잡음이라고 볼 수 있음

백색잡음 (White Noise)

White Noise Process (백색잡음)

자기상관이 존재하지 않는 시계열 데이터

우리가 일반적으로 사용하는 $IID(0, \sigma^2)$는 백색잡음이지만,

반대로 백색잡음은 $IID(0, \sigma^2)$가 아닐 수 있음에 유의 !!

1. 시계열 $\{X_t\}$의 **평균이 0, 분산이** $\sigma^2 < \infty$인 경우

2. 시계열 $\{X_t\}$에 **상관관계가 존재하지 않는 경우**

IID와 비교하여 독립성 조건이 완화된 것이 백색잡음!

시계열 $\{X_t\}$는 $\{X_t\} \sim WN(0, \sigma^2)$로 표현되는

백색잡음이라고 볼 수 있음

백색잡음 (White Noise) 검정

비정상 시계열 데이터로부터 추세 m_t, 계절성 s_t를 성공적으로 제거했다면,

남아있는 오차항은 WN조건 혹은 IID조건을 만족함

WN조건을 만족한다는 것은 $\{X_t\} \sim WN(0, \sigma^2)$ 와 동치

따라서 $\boldsymbol{\gamma_x(0) = Var(X_t) = \sigma^2}$ 이므로, $\boldsymbol{\sigma^2}$만을 추정하여

오차항이 WN조건을 만족하는지를 검정!

백색잡음 (White Noise) 검정

자기상관 검정

정규성 검정

정상성 검정

4 정상성 검정

백색잡음 (White Noise) 검정 - ① 자기상관 검정

오차항이 백색잡음을 따른다면 $\hat{\rho} \approx N(0,\frac{1}{n})$ 라는 점에 기반하여 아래 가설을 검정

$$H0 : p(h) = 0$$

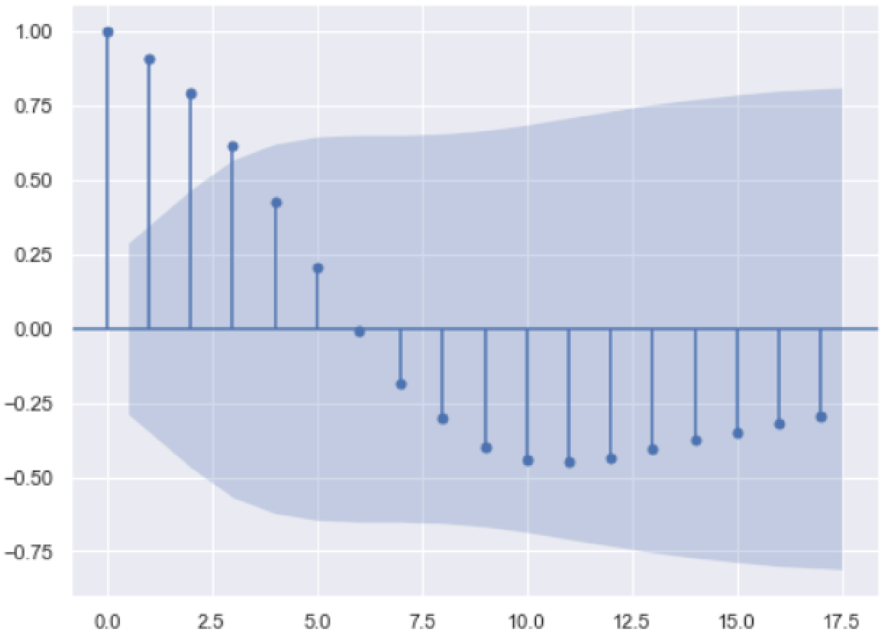
$$H1 : p(h) \neq 0$$

$|\hat{\rho}(h)| < \frac{1.96}{\sqrt{n}}$ 라면 $H0$를 기각할 수 없다.

이는 곧 **오차항에 자기상관성이 없음**을 의미한다.

백색잡음 (White Noise) 검정 - ① 자기상관 검정

ACF plot



ACF plot의 X축은 시차, y축은 acf를 의미한다.

파란색 영역은 자기상관 검정의 신뢰구간이다.

백색잡음 (White Noise) 검정 - ① 자기상관 검정

파란색 신뢰구간 밖으로 벗어난 그래프는

해당 **X축 값의 시차만큼 자기상관성이 존재**함을 의미한다.

ACF plot의 X축은 시차, y축은 acf를 의미한다.

파란색 영역은 자기상관 검정의 신뢰구간이다.

백색잡음 (White Noise) 검정 - ② 정규성 검정

$H0$: 정규성이 존재한다.

$H1$: 정규성이 존재하지 않는다.

QQ Plot	시각적으로 정규성을 확인할 수 있는 검정
KS Test	표본과 모집단의 누적확률분포가 얼마나 유사한지 비교하는 검정
Jarque-Bera Test	왜도와 첨도를 이용한 검정

4 정상성 검정

백색잡음 (White Noise) 검정 - ③ 정상성 검정

<table>
<tr><th>검정</th><th>H0</th></tr>
<tr><td>Kpss Test</td><td>정상 시계열이다.</td></tr>
<tr><td>ADF Test</td><td rowspan="2">정상 시계열이 아니다.</td></tr>
<tr><td>PP Test</td></tr>
</table>

이분산이 있는 경우에도 사용 가능한 검정 방법

다음 주 예고

1. 모형 식별
2. 선형과정
3. AR 모형
4. MA 모형
5. ARMA 모형
6. 적합 절차

감사합니다