클린업 1주차

5팀 시계열자료분석팀

염예빈
한유진
이재현
박세령
이정우

INDEX

1

시계열 자료 및 분석

시계열 자료 시계열 분석

시계열 자료란?

시계열 자료 : 연도, 계절, 월 등 **시간의 흐름에 따라 순서대로** 관측 되는 자료

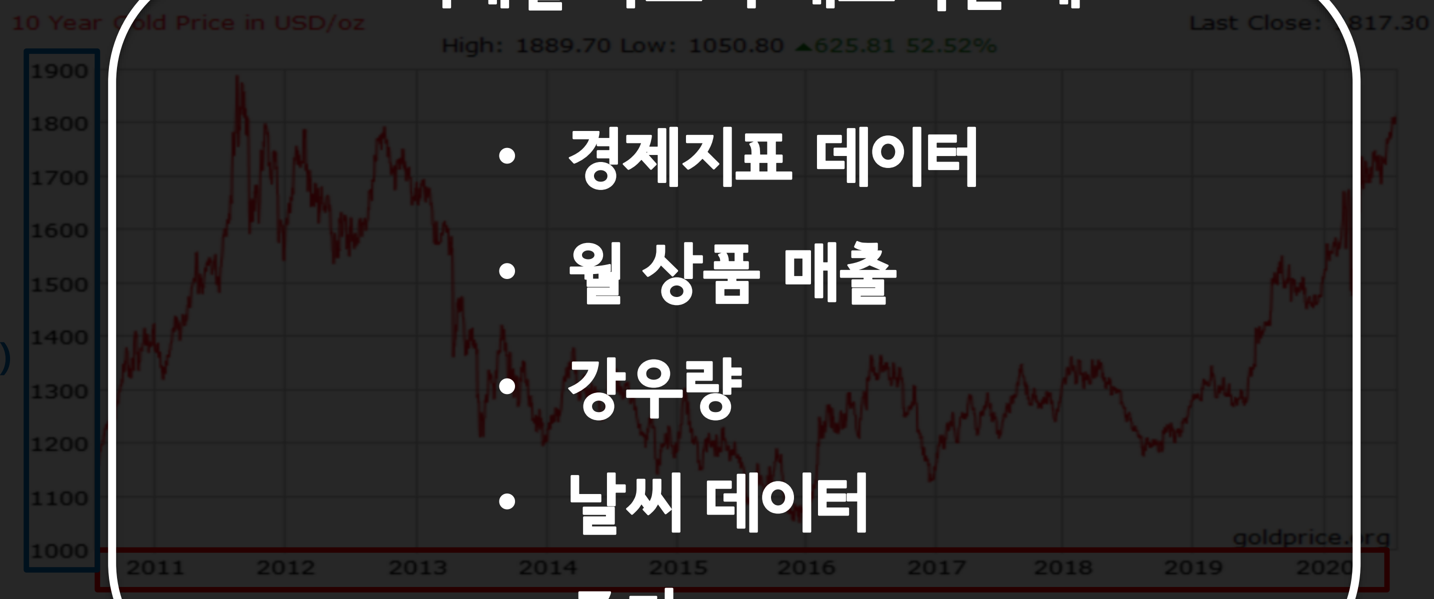
시계열 자료란?

시계열 자료 : 연도, 계절, 월 등 **시간의 흐름에 따라 순서대로** 관측 되는 자료

시계열 자료란?

시계열 자료 : 연도, 계절, 월 등 **시간의 흐름에 따라 순서대로** 관측 되는 자료

금값
(관측값)



연도
(시점)

시계열 자료의 대표적인 예

- 경제지표 데이터
- 월 상품 매출
- 강우량
- 날씨 데이터
- 주가

시계열 자료 | 시계열 분석

시계열 자료란?

[시계열 자료의 종류]

Discrete time series

이산 시계열 : 특정 시점에 측정된 관측값

→ 일반적으로 관측값 사이의 간격이 일정

Continuous time series

연속 시계열 : 시간의 모든 점에서 측정된 관측값

→ 이산 시계열에 비해 자료가 적은 편

시계열 자료 시계열 분석

시계열 자료란?

대부분의 시계열 자료

Discrete time series

이산 시계열 : 특정 시점에 측정된 관측값

→ 일반적으로 관측값 사이의 간격이 일정

Continuous time series

연속 시계열 : 시간의 모든 점에서 측정된 관측값

→ 이산 시계열에 비해 자료가 적은 편

시계열 자료 시계열 분석

시계열 자료란?

[시계열 자료의 특성]

- 추세변동 : 상승과 하락의 경향
- 순환변동 : 2~10년의 주기 안에서 상승과 하락을 반복
- 계절변동 : 일정한 기간 단위로 반복적 변동
- 불규칙변동 : 어떤 규칙없이 예측 불가능한 변동요인

시계열 자료 시계열 분석

시계열 분석의 목적

1.

시간 변수의 흐름에
따른 종속 변수의
패턴을 예측

2.

시계열 자료가
생성된 시스템을
이해하고 제어

시계열 자료 | 시계열 분석

시계열 분석의 목적

1.

시간 변수의 흐름에
따른 종속 변수의
패턴을 예측

2.

시계열 자료가
생성된 시스템을
이해하고 제어

시계열 자료 | 시계열 분석

회귀분석 vs 시계열분석

Regression
[회귀분석]

- 설명변수를 통한 예측
- 독립성을 전제
  → 순서 신경 X

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

VS

Time Series anlysis
[시계열분석]

- 과거 자료를 토대로 예측
- 자기상관성을 전제
  → 순서가 중요

$$y_t = \delta + \phi_1 y_{1-1} + \varepsilon_t$$

회귀분석 vs 시계열분석

Regression

[회귀분석]

- 설명변수를 통한 예측
- 독립성을 전제

→ 순서 시경 X

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Time Series analysis

[시계열분석]

- 과거 자료를 토대로 예측
- 자기상관성을

→ 순서가 중요

$$y_t = \delta + \phi_1 y_{1-1} + \varepsilon_t$$

VS

자기 상관성

어떤 변수에 대해 이전의 값이 이후의 값에 영향을 주는 것

즉, 이전 시점의 값이 이후 시점의 값에 영향을 주는 것

Earlier Data → Later Data

2

정상성

정상성 강정상성 약정상성

정상성이란?

정상성(Stationarity)

: 시계열의 확률적 성질들이 시간의 흐름에 따라

평균, 분산 등

변하지 않는 성질

정상성 강정상성 약정상성

정상성의 필요성

시계열에서 각 시점은 확률 변수이므로 각 시점별로 확률분포를 가짐

→ 즉, 예측을 위해서는 **무한한 시점의 결합분포를 고려**해야함 !!

정상성 강정상성 약정상성

정상성의 필요성

But..

It's too hard !!

따라서,

다수의 시점들에 대한 결합분포의 파악을 좀 더 간략하게 하기 위해 정상성이라는 가정이 필요 !!

시계열에서

→ 즉, 예측을 위해서는 무한한 시점의 결합분포를 고려해야함 !!

강정상성

$$F(X_t, \dots, X_{t+p}) = F(X_{t+k}, \dots, X_{t+k+p})$$

동일한 기간의 시계열에 대한 결합확률분포가 모든 시계열 구간에서 동일

→ **지나치게 엄격한 가정** :
이를 만족하는 시계열 자료는 거의 없다

따라서, 강정상성을 만족하기는 어려움!!

정상성

정상성 강정상성 약정상성

그래서 생각해낸

"약정상성 "

약정상성

- 분포의 동일성(강정상성의 조건)이 만족되지 않더라도 특성치의 동일성이 만족하는 경우

평균, 분산

- 시간 t에 관계없이 평균과 분산이 일정

- 공분산이 시차(time lag)에만 의존하고 시점 t와는 무관!

약정상성

1. $E(y_t) = \mu < \infty$

: 평균이 시점 t에 무관하게 일정

2. $Var(y_t) = E(y_t - \mu)^2 = \gamma_0 < \infty$

: 분산이 시점 t에 무관하게 일정

3. $\mathrm{Cov}(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) = \gamma_k < \infty \quad \forall k$

: 자기공분산(공분산이 시점이 아닌 시차에만 의존)

→ t에 무관하고 k에 의존

약정상성

1. $E(y_t) = \mu < \infty$

: 평균이 시점 t에 무관하게 일정

2. $Var(y_t) = E(y_t - \mu)^2 = \gamma_0 < \infty$

: 분산이 시점 t에 무관하게 일정

3. $Cov(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) = \gamma_k < \infty \quad \forall k$

: 자기공분산(공분산이 시점이 아닌 시차에만 의존)

→ t에 무관하고 k에 의존

이 3가지를 만족 :

정상적 시계열로 간주

또한, 앞으로 언급하는 정상성 = "약정상성"

3

정상화

비정상시계열 정상화 방법

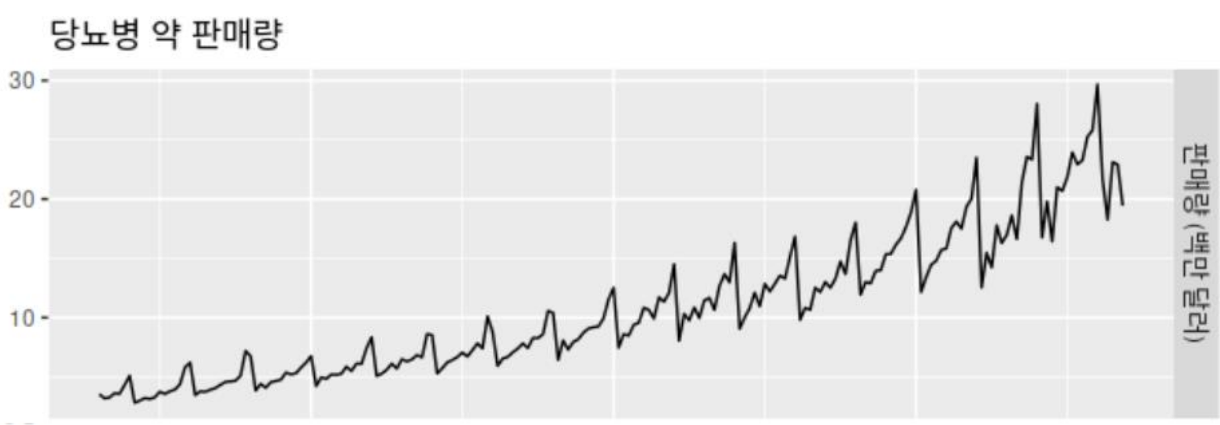
비정상시계열의 특징

1. 평균이 변하는 경우

비정상시계열 정상화 방법

비정상시계열의 특징

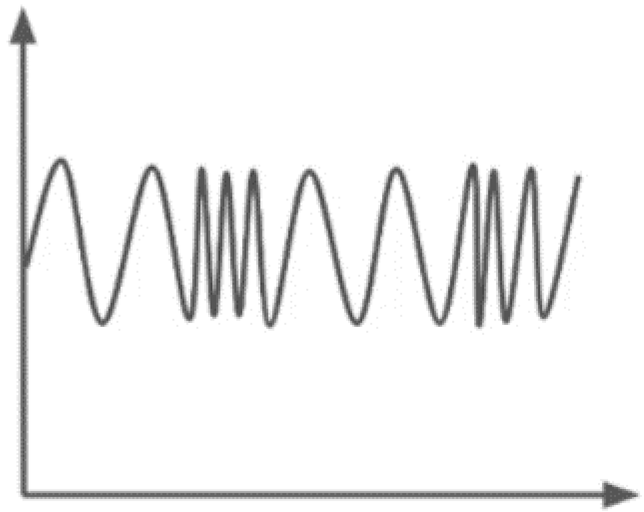
2. 분산이 시점에 영향을 받는 경우



비정상시계열 정상화 방법

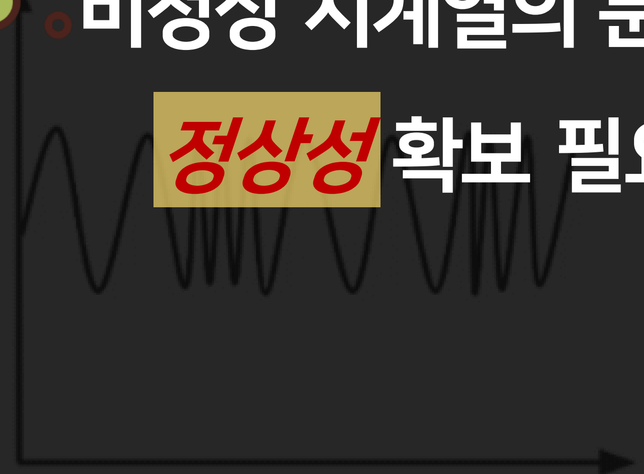
비정상시계열의 특징

3. 공분산이 시점에 영향을 받는 경우



비정상시계열의 특징

3. 공분 시점에 영향을 받는 경우



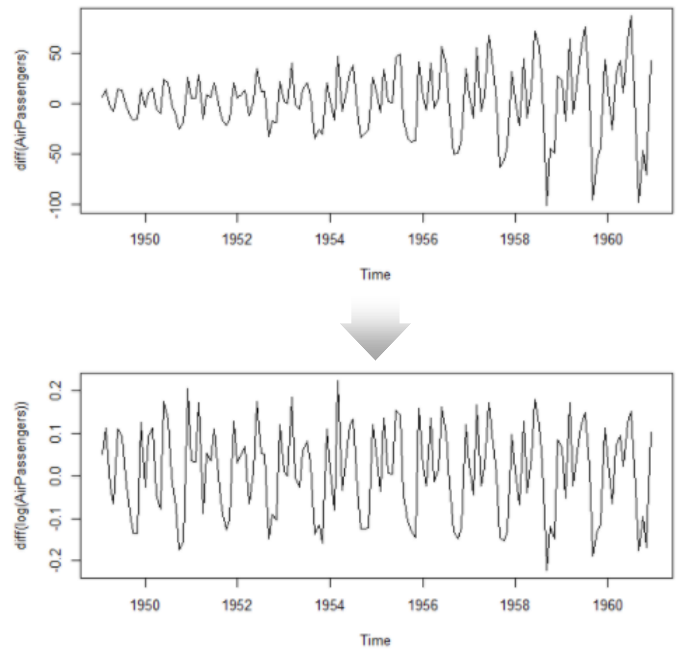
비정상 시계열의 분석?

정상성 확보 필요!

분산이 일정하지 않은 경우

HOW?

- 로그변환
- 제곱근 변환
- Box-cox 변환



비정상시계열 정상화 방법

추세/계절성을 가질 경우

시계열의 유형과 방법에 따라 과정이 달라진다.

방법	시계열 유형
회귀	추세만 있는 시계열
평활	계절성만 있는 시계열
차분	추세, 계절성 있는 시계열

회귀 – 추세만 있는 시계열

STEP 1. 추세성분과 불규칙성분만 있는 시계열 모형을 가정

$$\mathrm{Y_t} = \boxed{T_t} + \boxed{I_t}, E(I_t) = 0$$

$$Y_t = T_t + S_t + I_t \quad (E(I_t) = 0)$$

추세 계절성 오차

회귀 – 추세만 있는 시계열

EP 1. 추세성분과 불규칙성분만 있는 시계열 모형을 가정

어라? 순환변동을 나타내는 변수는 왜 없나요?

순환변동은 관측이 쉽지 않고 주기를 찾아내기 쉽지 않기 때문에 일반적으로 분해법에서 고려하지 않습니다!

비정상시계열 정상화 방법

추세/계절성을 가질 경우

시계열의 유형과 방법에 따라 과정이 달라진다.

① ***회귀*** 를 활용한

추세/계절성 제거

차분 추세, 계절성 있는 시계열

회귀 – 추세만 있는 시계열

STEP 1. 추세성분과 불규칙성분만 있는 시계열 모형을 가정

$$\mathrm{Y_t} = \boxed{T_t} + \boxed{I_t}, E(I_t) = 0$$

➡ T_t: 추세성분, I_t : 불규칙성분

비정상시계열 **정상화 방법**

회귀 – 추세만 있는 시계열

STEP 1. 추세성분과 불규칙성분만 있는 시계열 모형을 가정

$$Y_t = \boxed{T_t} + \boxed{I_t},\ E(I_t) = 0$$

T_t: 추세성분, I_t : 불규칙성분

STEP 2. 추세성분 T_t 를 시점에 대한 회귀식으로 표현

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

- 추세의 비선형 관계를 고려하기 위해 p차 다항식으로 표현
- P의 값은 플랏을 보고 분석자가 결정

회귀 – 추세만 있는 시계열

STEP 3. 최소제곱법을 통해 회귀식의 계수 추정

$$\left(\widehat{c_0}, \widehat{c_1}, \ldots, \widehat{c_p}\right) = argmin \sum (Y_t - T_t)^2$$

STEP 4. 추정한 추세를 시계열에서 제거

$$\mathrm{Y_t} - \widehat{T}_t = I_t$$

비정상시계열 정상화 방법

회귀 - 계절성만 있는 시계열

STEP 1. 계절성분과 불규칙성분이 있는 모형 가정

$$\mathrm{Y_t} = S_t + I_t, E(I_t) = 0$$

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

S_t는 주기함수를 삼각함수의 선형결합으로 표현해주는
"푸리에 변환"을 이용해 분해 → 자세한 내용은 Pass!

비정상시계열 정상화 방법

회귀 - 계절성만 있는 시계열

STEP 2. λ_j, k를 분석자가 결정한 후 최소제곱법을 통해 계수 추정

$$(a_0, a_1, \cdots, a_n, b_1, \cdots, b_n) = argmin \sum (Y_t - S_t)^2$$

STEP 3. 추정된 계절성을 시계열에서 제거

$$\mathrm{Y_t} - \widehat{S_t} = I_t$$

비정상시계열 | **정상화 방법**

회귀 - 계절성만 있는 시계열

STEP 2. … 계수 추정

$(Y_t - S_t)^2$

STEP 3. 추정된 계절성을 시계열에서 제거

$$Y_t - S_t = I_t$$

추세와 계절성 둘다 있으면 어떡하지?

➡ **추세 제거와 계절성 제거 반복!**

비정상시계열 정상화 방법

회귀 - 추세와 계절성 둘 다 있는 시계열

STEP 1. 최소제곱법으로 추세를 추정 후 추세 제거

$$Y_t = T_t + I_t, E(I_t) = 0 \xrightarrow{T_t\text{를 최소제곱법으로 추정}} Y_t - \widehat{T}_t = I_t$$

STEP 2. 최소제곱법으로 계절성을 추정 후 계절성 제거

$$Y_t = S_t + I_t, E(I_t) = 0 \xrightarrow{S_t\text{를 최소제곱법으로 추정}} Y_t - \widehat{S}_t = I_t$$

STEP 3. 추세가 다시 생길 경우 추세를 다시 추정 후 제거

비정상시계열 정상화 방법

회귀 - 추세와 계절성 둘 다 있는 시계열

$Y_t = T_t + I_t , E(I_t) = 0$ → T_t를 최소제곱법으로 추정 → $Y_t - T_t = I_t$

$Y_t = S_t + I_t$ → S_t를 최소제곱법으로 추정 → $Y_t - S_t = I_t$

STEP 3. 추세가 다시 생길 경우 STEP 1 과정 반복

하지만 한계점이 있다는 점…

➔ 1. 시계열에선 잔차의 독립성을 가정하지 않으므로 회귀의 가정을 만족하지 않을 수 있다.

➔ 2. 추세와 계절성이 독립이 아니므로 앞선 방법의 분해는 옳지 않을 수 있다.

➔ 3. 모수들이 시간에 따라 변하지 않는다는 가정을 한다.

회귀 - 추세와 계절성 둘 다 있는 시계열

STEP 1. 추세만 있는 시계열을 가정한 후 추세 제거

T_t를 최소제곱법으로 추정

Y_t = ... = I_t

Y_t = ... I_t

② ***평활*** 을 활용한

추세/계절성 제거

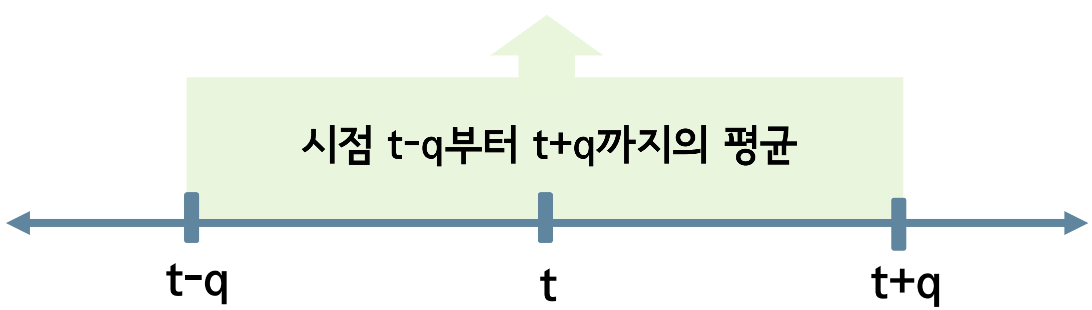
STEP 3. 추세가 다시 생길 경우 STEP 1 과정 반복

평활 – 추세만 있는 시계열

1. 이동평균 평활법 (추세는 선형이라고 가정)

✓ 일정기간별 평균을 계산하여 추세를 파악하는 방법

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



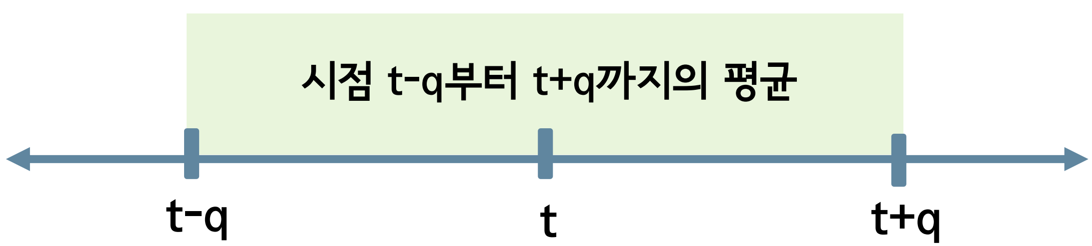
비정상시계열 정상화 방법

평활 – 추세만 있는 시계열

1. 이동평균 평활법

추세성분 T_{t+j}와 **불규칙성분** I_{t+j}으로 분해!

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \quad \Rightarrow \quad W_t = \frac{1}{2q+1}\sum_{j=-q}^{q}(T_{t+j} + I_{t+j})$$



평활 – 추세만 있는 시계열

1. 이동평균 평활법

추세성분 T_{t+j}

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

$$\frac{1}{2q+1}\sum_{j=-q}^{q} T_{t+j} = c_0 + c_1 t = T_t\,,$$

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

시점 t-q부터 t+q까지의 평균 = 시점 t에서의 추세!

t-q t t+q

비정상시계열 정상화 방법

평활 – 추세만 있는 시계열

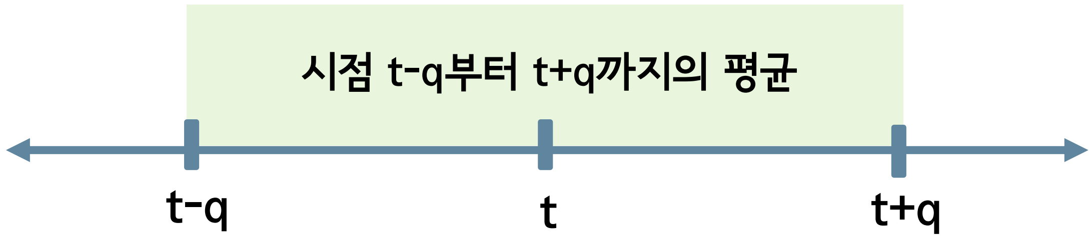
1. 이동평균 평활법

불규칙성분 I_{t+j}

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q}(T_{t+j} + \boxed{I_{t+j}})$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} I_{t+j} \approx E(I_t) = 0$$

불규칙성분 I_t 의 평균 =0



평활 – 추세만 있는 시계열

1. 이동평균 평활법

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q}(T_{t+j} + \boxed{I_{t+j}})$$

불규칙성분 I_{t+j}

$$\frac{1}{2q+1}\sum_{j=-q}^{q} I_{t+j} \approx E(I_t) = 0$$

$$W_t \approx T_t \,,$$

시점 t-q부터 t+q까지의 평균

t-q t t+q

$$Y_t - W_t = I_t$$

불규칙성분 I_t 의 평균 =0

추세 제거 성공!

평활 – 추세만 있는 시계열

1. 이동평균 평활법

이동평균평활법:

최근 자료와 과거 자료의 가중치가 동일하다는 단점!

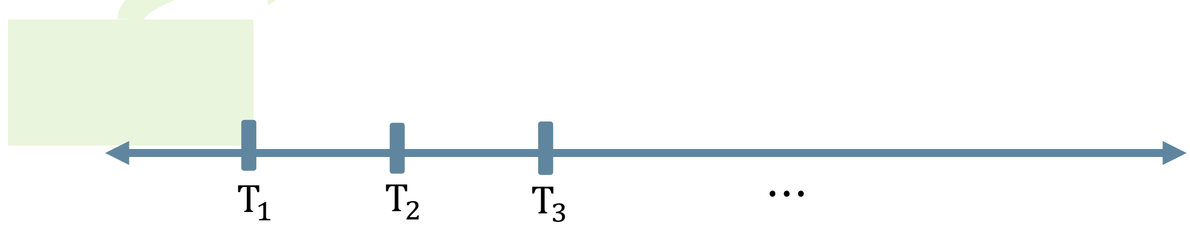
시점 t-q부터 t+q까지의 평균

평활 – 추세만 있는 시계열

2. 지수 평활법

✓ 시점별 가중평균으로 추세를 파악하는 방법

$$\hat{T}_1 = Y_1$$



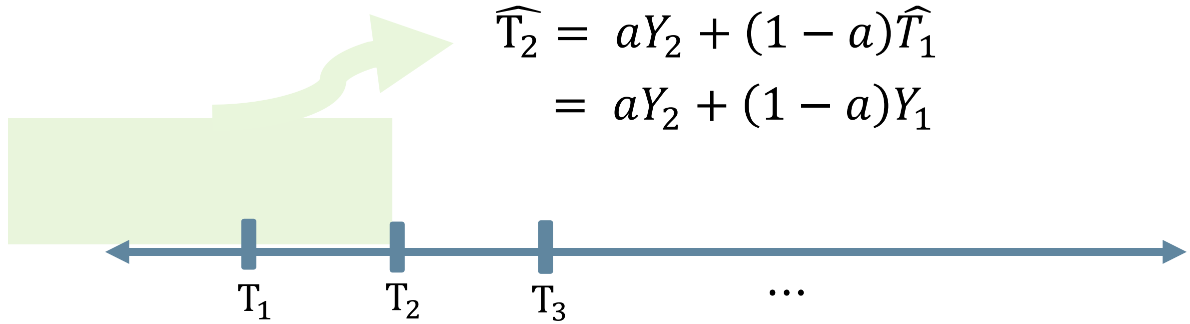
비정상시계열 정상화 방법

평활 – 추세만 있는 시계열

2. 지수 평활법

✓ 시점별 가중평균으로 추세를 파악하는 방법

$$\widehat{T_2} = aY_2 + (1-a)\widehat{T_1}$$
$$= aY_2 + (1-a)Y_1$$

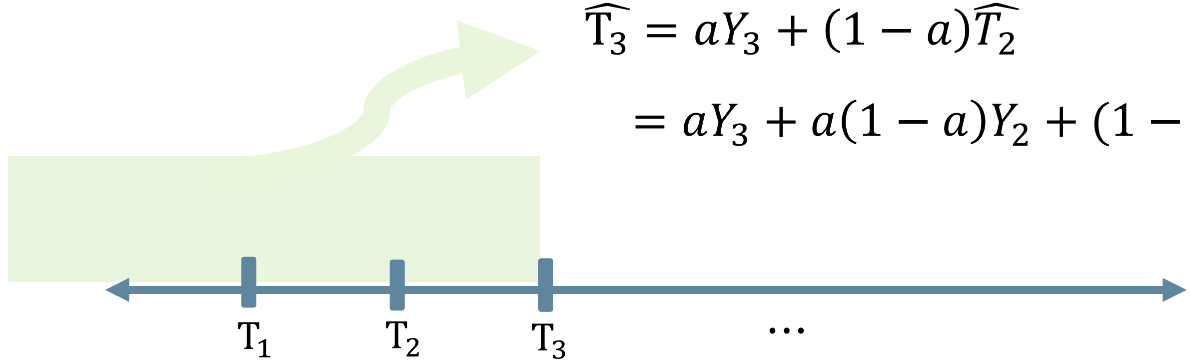


평활 – 추세만 있는 시계열

2. 지수 평활법

✓ 시점별 가중평균으로 추세를 파악하는 방법

$$\widehat{T_3} = aY_3 + (1-a)\widehat{T_2}$$
$$= aY_3 + a(1-a)Y_2 + (1-a)^2Y_1$$



비정상시계열 정상화 방법

평활 – 추세만 있는 시계열

2. 지수 평활법

시점1 추세 : $\widehat{\mathrm{T}_1} = Y_1$

시점2 추세 : $\widehat{\mathrm{T}_2} = aY_2 + (1-a)\widehat{T_1} = aY_2 + (1-a)Y_1$

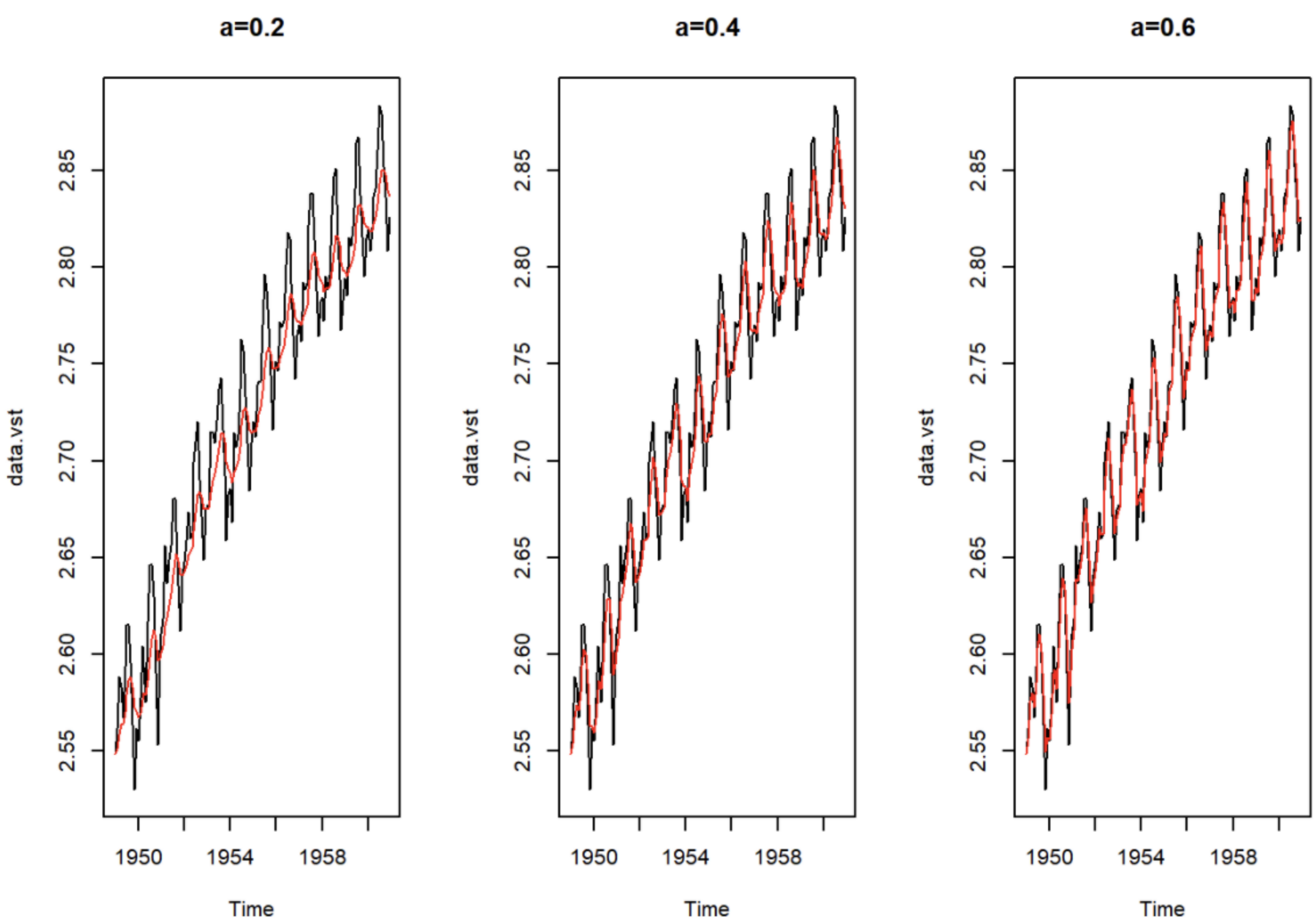
시점3 추세 : $\widehat{\mathrm{T}_3} = aY_3 + (1-a)\widehat{T_2} = aY_3 + a(1-a)Y_2 + (1-a)^2 Y_1$

…

시점t 추세 : $\widehat{\mathrm{T}_\mathrm{t}} = aY_t + (1-a)\widehat{T_{t-1}} = \sum_{j=0}^{t-2}(a(1-a)^j Y_{t-j}) + (1-a)^{t-1}Y_1$

- 평활계수 a가 클수록 시계열 변화에 따른 예측값의 변화가 크다.
- 최근자료에 더 큰 가중치를 부여함으로써 이동평균평활법의 단점 보완

정상화



최근자료에 더 큰 가중치를 부여함으로써 이동평균평활법의 단점 보완

비정상시계열 정상화 방법

평활 – 계절성만 있는 시계열

Seasonal Smoothing

✓ $\widehat{S}_t$를 같은 주기를 갖는 모든 값들의 평균값으로 대체.

(계절성 주기 : d)

$$\widehat{S_1} = \frac{1}{\mathrm{m}}\left(Y_1 + Y_{1+\mathrm{d}} + Y_{1+2\mathrm{d}} + Y_{1+3\mathrm{d}} + \cdots + Y_{1+(\mathrm{m}-1)\mathrm{d}}\right)$$

$$\widehat{S_2} = \frac{1}{m}(Y_2 + Y_{2+\mathrm{d}} + Y_{2+2\mathrm{d}} + Y_{2+3\mathrm{d}} + \cdots + Y_{2+(\mathrm{m}-1)\mathrm{d}}$$

$$\vdots$$

$$\widehat{S_{\mathrm{d}}} = \frac{1}{\mathrm{m}}(Y_{\mathrm{d}} + Y_{\mathrm{d}+\mathrm{d}} + Y_{\mathrm{d}+2\mathrm{d}} + Y_{\mathrm{d}+3\mathrm{d}} + \cdots + Y_{\mathrm{d}+(\mathrm{m}-1)\mathrm{d}}$$

평활 - 추세와 계절성 둘 다 있는 시계열

<Classical decomposition algorithm>

Step 1. 이동평균평활법을 이용해 추세 추정 (이때, $\sum_{j=1}^{d} S_j = 0$)

Step 2. 관측값에서 추정한 추세를 빼서 계절성과 불규칙성분만 남긴다.

$$\mathrm{Y_t} - \widehat{T_t} \approx S_t + I_t$$

Step 3. Seasonal Smoothing을 통해 계절성분 S_t 를 추정

Step 4. 관측값에서 추정한 계절성을 빼서 추세와 불규칙성분만 남긴다.

Step 5. (Step 4)식에서 추세성분 T_t를 회귀를 통해 추정

마무리

(Step 3)에서 추정한 계절성분 S_t와 (Step 5)에서 추정한 추세성분 T_t를 관측치에서 제거한다.

회귀 - 추세와 계절성 둘 다 있는 시계열

STEP 1. 추세만 있는 시계열을 가정한 후 추세 제거

T_t를 최소제곱법으로 추정

$Y_t =$ … $= I_t$

$Y_t =$ … I_t

③ 차분 을 활용한

추세/계절성 제거

STEP 3. 추세가 다시 생길 경우 STEP 1 과정 반복

차분의 정의

현재 시계열에서 과거 시계열을 뺀 것!

1차 차분

$$\nabla X_t = X_t - X_{t-1}$$

2차 차분

$$\nabla^2 X_t = \nabla(\nabla X_t)$$

$$= X_t - X_{t-1} - (X_{t-1} - X_{t-2})$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

차분의 정의

현재 시계열에서 과거 시계열을 뺀 것!

1차 차분

$$\nabla X_t = X_t - X_{t-1}$$

2차 차분

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

후향연산자(Backshift Operator)

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

앞으로 배울 식들을 간단하게 표현할 수 있다!!

비정상시계열 정상화 방법

차분 – 후항연산자

후항 연산자 이용

1차 차분

$$\nabla X_t = X_t - BX_t = (1 - B)X_t$$

2차 차분

$$\nabla^2 X_t = \nabla(\nabla X_t)$$

$$= X_t - X_{t-1} - (X_{t-1} - X_{t-2})$$

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

차분 - 추세만 있는 시계열

추세가 k차 다항식일 경우 k번 차분

추세가 2차식일 때 :

$$Y_{\mathrm{t}} = T_t + I_t = (c_0 + c_1 t + c_2 t^2) + I_t$$

$$\nabla^2 \mathrm{Y_t} = Y_t - 2Y_{t-1} + Y_{t-2}$$

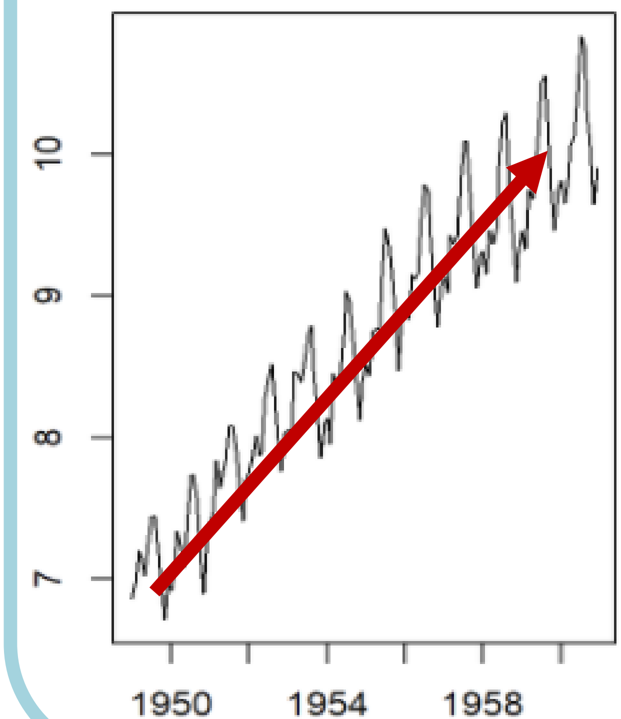
$$= 2c_2 + (I_t - 2I_{t-1} + I_{t-2})$$

∴ 상수와 오차항만 남아 추세 제거 가능!!

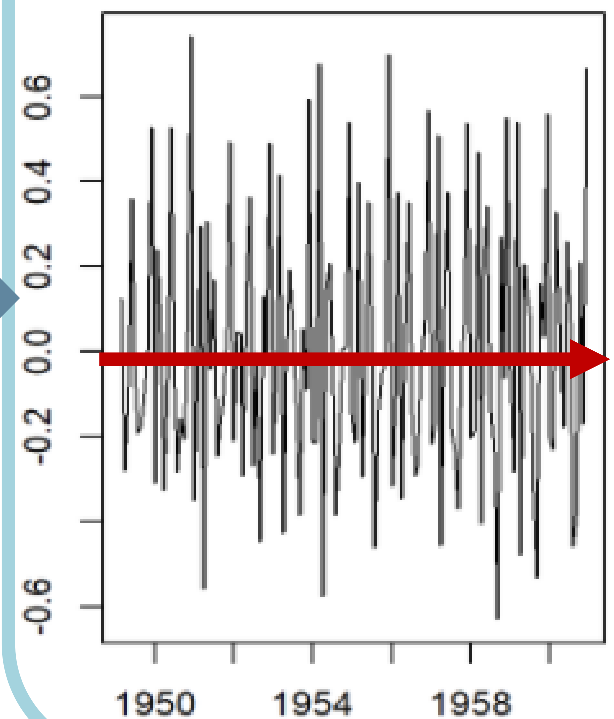
비정상시계열 정상화 방법

차분 - 추세만 있는 시계열

[차분 전]



[2차 차분 후]



비정상시계열 정상화 방법

차분 - 계절성만 있는 시계열

계절차분(Seasonal differencing)을 통해 계절성 제거 가능!

주기가 d인 계절차분 :

$$\nabla_{\mathrm{d}} X_t = (1 - B^d) X_t = X_t - X_{t-d}$$

Cf) d차 차분 : $\nabla^{\mathrm{d}} X_t = (1 - B)^d X_t$

∴ 오차항만 남아 계절성 제거 가능!!

차분 - 계절성만 있는 시계열

계절차분(Seasonal differencing)을 통해 계절성 제거 가능!

예시를 달라!

주기가 d인 계절차분 :

$$\nabla_{d} X_t = (1 - B^d) X_t = X_t - X_{t-d}$$

Cf) d차 차분 : $\nabla^{d} X_t = (1 - B)^d X_t$

∴ 오차항만 남아 계절성 제거 가능!!

Ex) 주기가 4인 시계열

$$\nabla_4 Y_t = (1 - B^4)\ Y_t$$

$$= (1 - B^4)(S_t + I_t)$$

$$= (S_t + I_t) - (S_{(t-4)} + I_{(t-4)})$$

$$= I_t - I_{(t-4)}$$

차분 – 추세와 계절성 모두 있는 시계열

STEP 1. 추세와 계절성이 있는 모형을 가정

$$\mathrm{Y_t} = T_t + S_t + I_t = (c_0 + c_1 t) + S_t + I_t$$

STEP 2. 1차 차분과 주기가 d인 계절 차분 동시에 적용

$$\nabla\nabla_{\mathrm{d}} Y_t = (1 - B)(Y_t - Y_{t-d})$$
$$= (1 - B)(c_0 + c_1 t + S_t + I_t - c_0 - c_1(t - d) - S_{t-d} - I_{t-d})$$
$$= (1 - B)(c_1 d + S_t - S_{t-d} + I_t - I_{t-d})$$
$$= I_t - I_{t-1} - I_{t-d} - I_{t-d-1}$$

차분 – 추세와 계절성 모두 있는 시계열

STEP 1. 추세와 계절성이 있는 모형을 가정

$$Y_t = T_t + S_t + I_t = (c_0 + c_1 t) + S_t + I_t$$

→ 오차항만 남아 추세와 계절성 제거 가능

STEP 2. 1차 차분과 주기가 d인 계절 차분 동시에 적용

$$\nabla\nabla_d Y_t = (1-B)(Y_t - Y_{t-d})$$

→ 계절차분과 1차차분의 순서를 바꾸어도 결과는 같음

$$= (1-B)(c_0 + c_1 t + S_t + I_t - c_0 - c_1(t-d) - S_{t-d} - I_{t-d})$$

$$= (1-B)(c_1 d + S_t - S_{t-d} + I_t - I_{t-d})$$

$$= I_t - I_{t-1} - I_{t-d} - I_{t-d-1}$$

4

정상성 검정

정상성 검정이란?

$$Y_t = \cancel{T}_t + \cancel{S}_t + I_t$$

추세와 계절성 제거 후
마지막에 남은 오차항이 정상성을 따르는지
검정하는 과정!

자기공분산함수(Auto-Covariance Function)

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

자기상관계수(Auto-Correlation Funciton)

$$\rho(k) = Corr(X_t, X_{t+k}) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+k})}} = \frac{\gamma(k)}{\gamma(0)},$$

이때 $\gamma_0 = Var(X_t) = E\left[(X_t - \mu)^2\right]$

다루는 모든 데이터는 샘플이므로 추정한 값을 쓴다!

자기공분산함수 백색잡음 백색잡음 검정

표본자기공분산함수(SACVF)

$$\widehat{\gamma(k)} = \frac{1}{T}\sum_{j=1}^{T-k}(X_j - \bar{X})(X_{j+k} - \bar{X}), k = 1, 2, 3, \dots$$

다루는 모든 데이터는 샘플이므로 추정한 값을 쓴다!

표본자기상관계수(SACF)

$$\widehat{\rho(\mathrm{k})} = \frac{\widehat{\gamma(k)}}{\widehat{\gamma(0)}}$$

∴ SACF가 크면
시차 K에 대해 상관관계 존재!!

백색잡음이란?

$$\mathbf{Y}_t \sim \boldsymbol{WN}(0, \boldsymbol{\sigma}_Y^2)$$

대표적인 **정상시계열**의 예시로, 서로 독립 & 동일한 분포를 따르는(iid) 확률변수들로 구성된 **확률과정**

① $E(Y_t) = 0$

② $Var(Y_t) = \sigma^2$

③ $Cov(Y_t, Y_{t+k}) = 0$

강한 조건 대신 다음 조건만 만족하면 백색잡음이라 하자!

백색잡음의 조건

백색잡음인 오차항

✓ "약정상성" + "공분산 = 0"

∴ 시계열 자료의 공분산 행렬 추정 불필요

⇒ **추가적인 모델링 불필요**

① $E(Y_t) = 0$

② $Var(Y_t) = \sigma^2$

③ $Cov(Y_t, Y_{t+k}) = 0$

약정상성 조건

$$\boldsymbol{E(Y_t) = 0}$$

$$\boldsymbol{Var(Y_t) = \sigma^2}$$

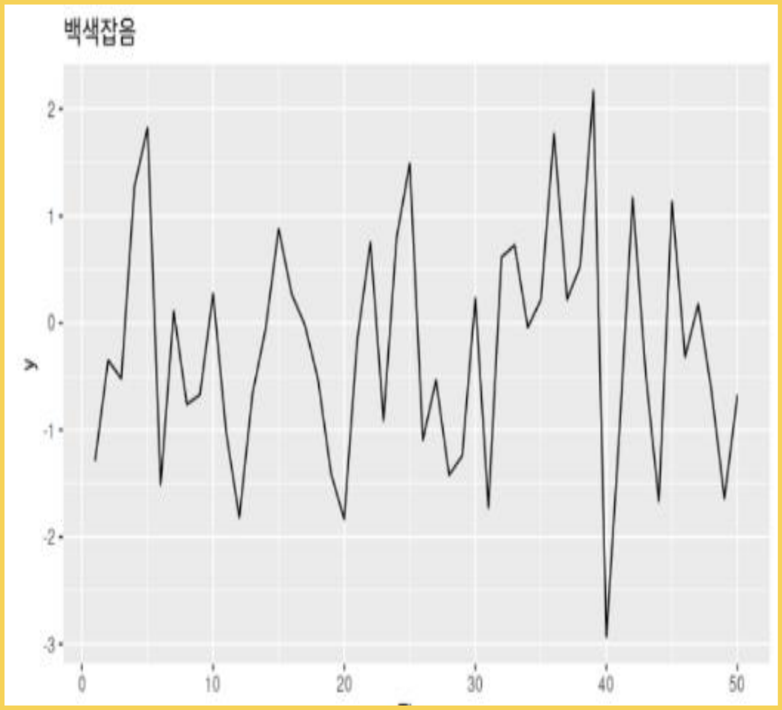
$$\boldsymbol{Cov(Y_t, Y_{t+k})}$$

$$= \boldsymbol{E(y_t - \mu)(y_{t+k} - \mu)}$$

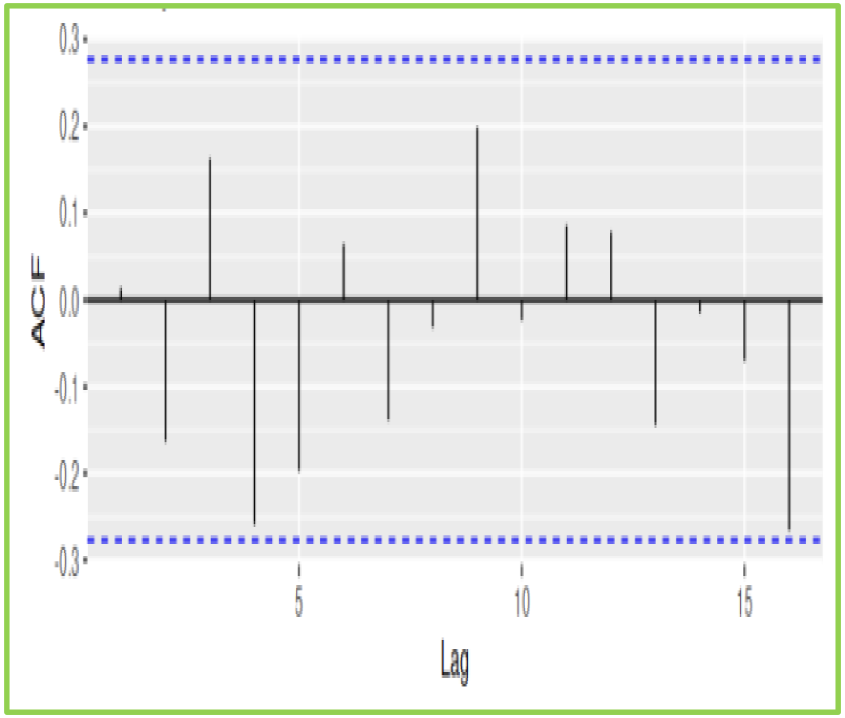
자기공분산함수 **백색잡음** 백색잡음 검정

백색잡음 그래프

[백색잡음 그래프]



[백색잡음의 ACF]



자기상관관계 유무 검정

귀무가설: $\mathrm{H}_0: \rho_1 = \rho_2 = \cdots = \rho_k = 0$

✓ **ACF그래프를 통한 검정** :

ACF, PACF그래프가 각 시점에서 신뢰구간안에 존재하고 있을 경우

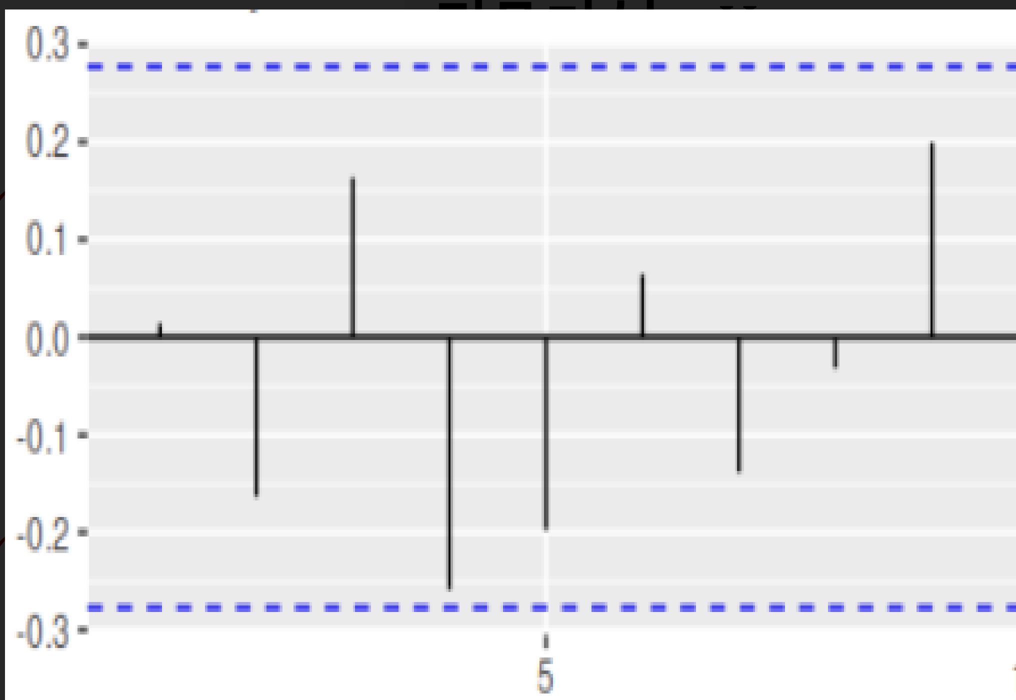
⇒ 잔차 간의 상관관계가 없다고 판단

✓ **포트맨토 검정(Portmanteau Test, Ljung-Box 검정)** :

$\epsilon_t \sim N(0, \sigma^2)$ 가정하에서 검정통계량 $Q' = T(T+2)\sum_{k=1}^{h}\frac{r_k^2}{T-k} \sim \chi(h-K)$ 를 사용하여 검정

자기상관관계 유무 검정

[백색잡음의 ACF]



ACF함수로 검정한 결과,

신뢰구간 내에 있으므로
유의미한 자기상관성을
갖지 않는다!

$= \cdots = \rho_k = 0$

재하고 있을 경우

정) :

$\epsilon_t \sim N(0, \sigma^2)$ 가정하에서 검정통계량 $Q^* = T(T+2)\sum_{k=1}^{h}\frac{r_k^2}{T-k} \sim \chi(h-K)$ 를 사용하여 검정

정규성 검정

- ✓ QQplot확인 :

  Qqplot으로 정규성 가정 만족 여부 시각적으로 확인

- ✓ Kolmogorov-Smirnov test :

  자료의 평균/표준편차와 히스토그램을 표준정규분포와 비교

  ⇒ 적합도 검정

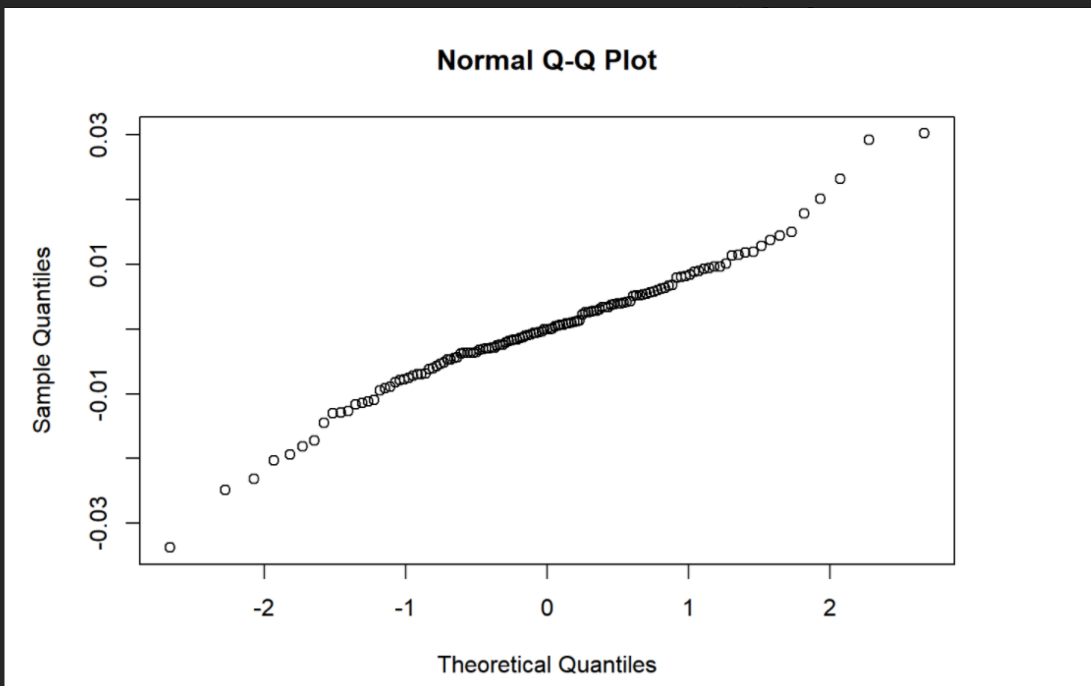
  귀무가설 : '정규성을 따른다'

- ✓ Jarque-Bera test : 왜도와 첨도를 정규분포와 비교

  ⇒ 정규성 검정

  귀무가설 : '데이터가 정규분포를 따른다'

정규성 검정



… 만족 여부 시각적으로 확인

…t :

…히스토그램을 표준정규분포와 비교

qqplot을 통해
간편하게
정규성을 확인할 수 있다!

…른다'

… 첨도를 정규분포와 비교

→ 정규성 검정

귀무가설 : '데이터가 정규분포를 따른다'

정상성 검정

각 방법의 귀무가설이
모두 다르므로 주의!!

✓ **kpss test** :

단위근(Unit-root) 검정 방법 중 하나.

귀무가설: '시계열이 정상(stationary)시계열이다.'

✓ **ADF(Augmented Dickey-Fuller)Test** :

단위근 검정방법 중 하나이며, DF검정을 일반화 한 것

귀무가설: '자료에 단위근이 존재한다.' VS 대립가설: '자료가 정상성을 만족한다.'

✓ **PP(Phillips-Perron)Test** :

이분산의 경우에도 사용 가능.

귀무가설은 '데이터가 비정상이다.' VS 대립가설: '자료가 정상성을 만족한다.'

정상성 검정

adf테스트

- 귀무가설 : 이데이터는 비정상적이다.
- 대립가설 : 이데이터는 정상적이다.

```
adf.test(dat.wn)
```

```
## Warning in adf.test(dat.wn): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  dat.wn
## Dickey-Fuller = -5.2679, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

각 방법의 귀무가설이
모두 다르므로 주의!!

Adf테스트 결과
p-value를 통해
귀무가설의 기각여부로
정상성을 판단할 수 있다!

'자료가 정상성을 만족한다.'

...료가 정상성을 만족한다.'

THANK YOU