시계열자료분석팀

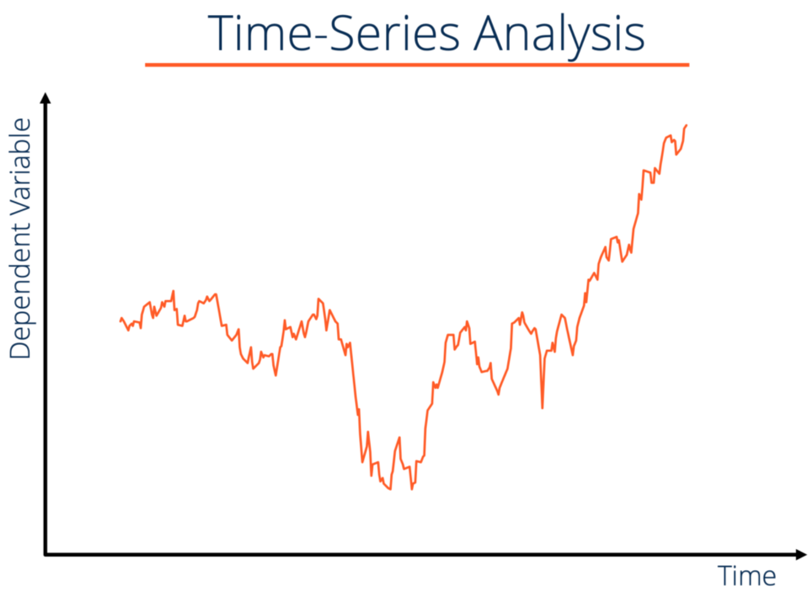
5팀

김규범
김민지
김준서
안세현
정희철

● 시계열 알아가기

시계열 자료

연도별, 월별, 일별 등 시간에 따라 관측된 자료



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

시간 t에 따라 이산형과 연속형으로 구분됨

1 시계열이란

시계열 알아가기

시계열 자료의 구성요소

불규칙 성분
시간의 영향 X

일정한 규칙이 없어 예측 불가능한 변동요인

체계적 성분
시간의 영향 O

추세변동
시간이 경과함에 따라 상승,하락이 반복되는 요소

순환변동
주기적인 변화를 가지나 계절에 의한 것이 아님
주기가 긴 경우의 변동

계절변동
주별/월별/계절별과 같은 주기적 성분에 의한 변동

정상성(Stationarity)

강정상성

모든 h와 n 〉0에 대해 시계열이 다음 조건을 만족하는 경우

$$(X_{t_1}, \dots, X_{t_n}) = (X_{t_1+h}, \dots, X_{t_n+h})$$

동일한 기간의 시계열의 결합확률분포가

시간대 (t)를 바꾸어도 동일해야 함

지나치게 엄격한 가정으로 만족하기에 현실적으로 **어려움**

정상성(Stationarity)

약정상성

시계열이 다음 세 조건을 만족하는 경우

1 $E[X_t] = \mu < \infty, \forall_t \in Z$

2 $Var[X_t] = \gamma_0 < \infty, \forall_t \in Z$

3 $Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)] = \gamma_h < \infty$

정상화 과정

1. 분산이 일정하지 않은 경우

■ 로그 변환

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

■ 제곱근 변환

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

■ Box-Cox 변환

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

분산 안정화를 통해 시간에 따라 증가하거나 감소하는 분산을
일정하게 변환

정상화 과정

2. 평균이 일정하지 않은 경우: 회귀

추세만 존재하는 경우

1

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

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

추세 성분 m_t를 다음과 같은 t에 대한 선형 회귀식으로 표현

정상화 과정

2. 평균이 일정하지 않은 경우: 회귀

추세만 존재하는 경우

2

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

최소제곱법(OLS)를 통해 선형회귀식 계수를 추정
추정한 추세를 시계열에서 제거

⇩

정상 시계열

정상화 과정

2. 평균이 일정하지 않은 경우: 회귀

계절성만 존재하는 경우

1

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

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

λ_j는 2π인 함수의 주기와 데이터 주기를 맞추기 위한 값

계절 성분 s_t를 다음과 같이 시간 t에 대한 회귀식으로 표현

적절한 λ_j와 k를 선택하여 OLS로 a_j와 b_j를 추정

● 정상화 과정

2. 평균이 일정하지 않은 경우: 회귀

추세와 계절성이 동시에 존재하는 경우

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

1. 추세만 존재하는 경우의 회귀 방식과
   계절성만 존재하는 경우의 회귀 방식을 차례로 적용

2. 여전히 추세가 존재 -> 추가적인 추세를 제거

정상화 과정

2. 평균이 일정하지 않은 경우: 평활

추세만 존재하는 경우 – 이동평균 평활법

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

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

특정 시점 t에 대해 **±q시점의** 평균을 구함

-〉 길이가 $2q+1$인 구간의 평균

정상화 과정

2. 평균이 일정하지 않은 경우: 평활

추세만 존재하는 경우 – 지수평활법

시점 1일 때의 추세 추정값 : $\hat{m}_1 = X_1$

시점 2일 때의 추세 추정값 : $\hat{m}_2 = \alpha X_2 + (1-\alpha)\hat{m}_1 = \alpha X_2 + (1-\alpha)\, X_1$

⋮

시점 t일 때의 추세 추정값 : $\hat{m}_t = aXt + (1-a)\hat{m}_{t-1}$

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

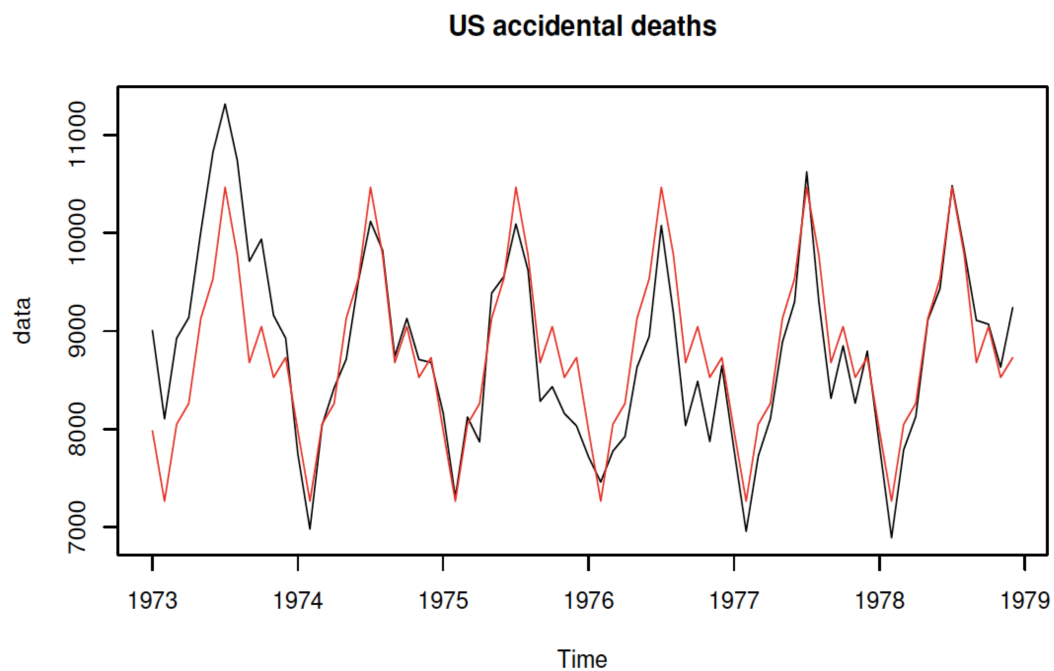
0< α <1이므로 과거 시점일수록 더 작은 가중치 부여

지수평활법은 추세 m_t를 t시점까지의 관찰값을 이용해 추정하는 방법

정상화 과정

2. 평균이 일정하지 않은 경우: 평활

계절성만 존재하는 경우 – Seasonal Smoothing



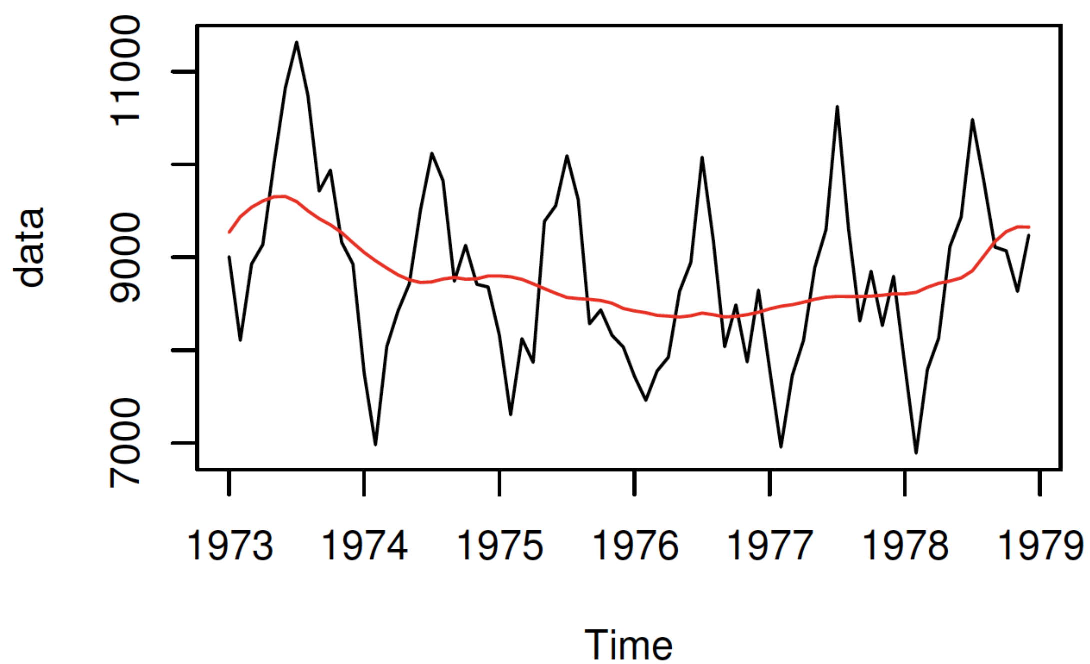
각 주기의 k번 째 Observations 평균값 계산

해당 평균값을 각 주기의 k번 째 Observations들에 공통적으로 **차감**

● 정상화 과정

2. 평균이 일정하지 않은 경우: 평활

추세와 계절성이 동시에 존재 – Classical Decomposition



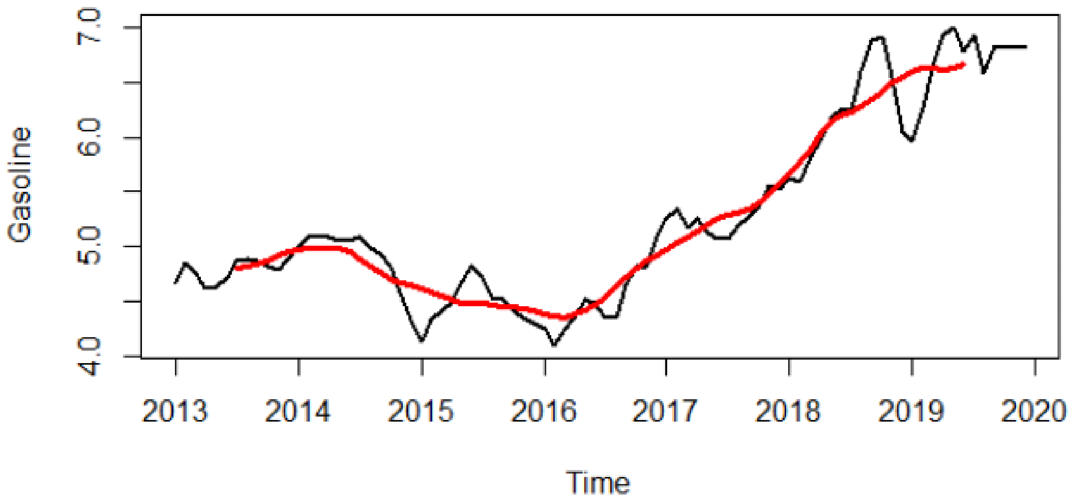
앞서 설명한 추세만 있을 경우와 계절성만 있을 경우

두 가지 가정과 방법을 독립적으로 진행

● 정상화 과정

2. 평균이 일정하지 않은 경우: 차분

차분이란?



관측값들의 차이를 구해 추세 또는 계절성을 제거하는 방법

K차 차분 -> 데이터에 존재하는 K차 추세 제거 가능

정상화 과정

2. 평균이 일정하지 않은 경우: 차분

차분이란?

[**후향 연산자**(Backshift Operator)]

차분 식을 간단하게 하기 위해 등장하는 새로운 연산자

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

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

관측값들의 차이를 구해 추세 또는 계절성을 제거하는 방법

K차 차분 -〉 데이터에 존재하는 K차 추세 제거 가능

정상화 과정

2. 평균이 일정하지 않은 경우: 차분

추세와 계절성이 동시에 존재

가정 : $X_t = s_t + m_t + Y_t$

1 계절차분을 먼저 진행

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

2 추세차분을 진행

$$\nabla^{p-1} \nabla_d X_t$$

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

백색잡음

자기상관이 없는 시계열로 대표적인 정상 시계열

1. $E[X_t] = 0$
2. $Var[X_t] = \sigma^2 < \infty$
3. $Cov(X_t, X_{t+h}) = 0$

정상시계열

백색잡음 검정

자기상관관계 검정

1. Portmanteau test
2. Ljung-Box test
3. McLeod-Li test

정상시계열

백색잡음 검정

정규성 검정

QQ plot	Kormogorov-Sminorv test	Jarque-Bera test
시각적으로 정규성을 따르는지 확인	자료의 평균, 표준편차, 히스토그램을 표준정규분포와 비교하여 적합도 검정	왜도와 첨도를 정규분포와 비교하여 정규성 검정

정상시계열

백색잡음 검정

정상성 검정

귀무가설 = '정상 시계열이다 '

대립가설 = '정상 시계열이다 '

Kpss test	ADF test	PP test
단위근 검정방법 중 하나	단위근 검정방법 중 하나	이분산이 있을 경우에도 사용가능한 검정방법