

시계열자료분석팀

5팀

오정민
강현주
김민정
김정민
배지현

INDEX

1. 시계열 자료
2. 정상성
3. 정상화 과정
4. 정상성 검정

시계열자료

시계열자료란

시계열 분석의 목적

회귀분석과 시계열 분석의 차이



: 계절, 월, 일 등 **시간의 흐름**에 따라 관측된 자료

GDP, 주가, 총인구, 강수량, 태양의 흑점수 등 다양한 분야에서 활용

정상성이란?

필요성

강정상성

약정상성

: 시계열의 **확률적 성질(평균, 분산 등)**들이
시간의 흐름에 따라 변하지 않는다는 것을 의미

약정상성, 강정상성

강정상성

필요성

강정상성

약정상성

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

모든 n 에 대하여 시간대가 바뀌어도 **결합분포가 같다**

같은 시차면 동일한 결합확률분포를 갖는다.

과거의 확률분포 및 상호관계가 미래에도 **그대로 유지!**

결합분포란?

확률 변수가 여러 개일 때
이들을 함께 고려하는 확률 분포

약정상성

필요성

강정상성

약정상성

1. $Var(X_t) = \sigma^2$ \longrightarrow 분산이 시점 t 에 관계없이 일정

2. $E(X_t) = m$ \longrightarrow 평균이 시점 t 에 관계없이 일정

3. $\gamma(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$
 \longrightarrow 공분산이 시점 t 가 아닌 시차 h 에만 의존

정상화 방법

분산 안정화

추세/계절성 제거

분산 안정화

■ Box – Cox 변환

$$f(x; \lambda) = \frac{x^{\lambda} - 1}{\lambda} \quad (\lambda \neq 0)$$

$$f(x; 0) = \ln(x) \quad (\lambda = 0)$$

$$\text{■ 로그 변환} \quad f(x) = \ln(x)$$

$$\text{■ 제곱근 변환} \quad f(x) = \sqrt{x}$$

여러 변환이 있다!
여럿 적용해보고
가장 잘 안정화 시키는
변환 방법을 취하자

정상화 방법

분산 안정화

추세/계절성 제거

평활 (Smoothing)

추세

- 이동평균 평활법
- 지수평활법

계절성

‘같은 주기를 갖는
값들의 평균값’을
사용

회귀 + 평활

추세+계절성

classical
decomposition



데이터에 따라 다른 방법을 사용한다!

정상화 방법


차분

추세/계절성 제거


차분 (differencing)

- 특정 시점 값과 그 바로 앞 시점 값의 차이

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



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



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

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

백색잡음

백색잡음(White Noise)

-대표적인 정상 시계열의 예시

-확률변수들이 IID일 때

(IID : 서로 독립이면서 동일한 분포를 따름)

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

White Noise

백색잡음

백색잡음(White Noise)

백색잡음은 다음과 같은 가정을 만족해야한다!

1 $E(Y_t) = 0$

2 $Var(Y_t) = \sigma^2$

3 $Cov(Y_t, Y_{t+k}) = 0 \rightarrow \text{Uncorrelated!!}$

(약정상성 조건에 '공분산 = 0' 조건 추가)

검정

독립성 검정

정규성 검정

정상성 검정

1. 독립성 검정

Portmanteau test, Ljung-Box test, McLeod and Li test

 H_0 (귀무가설): 잔차 간 독립이고 동일한 분포를 따른다 (iid), H_1 (대립가설): 아니다

귀무가설이 받아들여져야 한다!

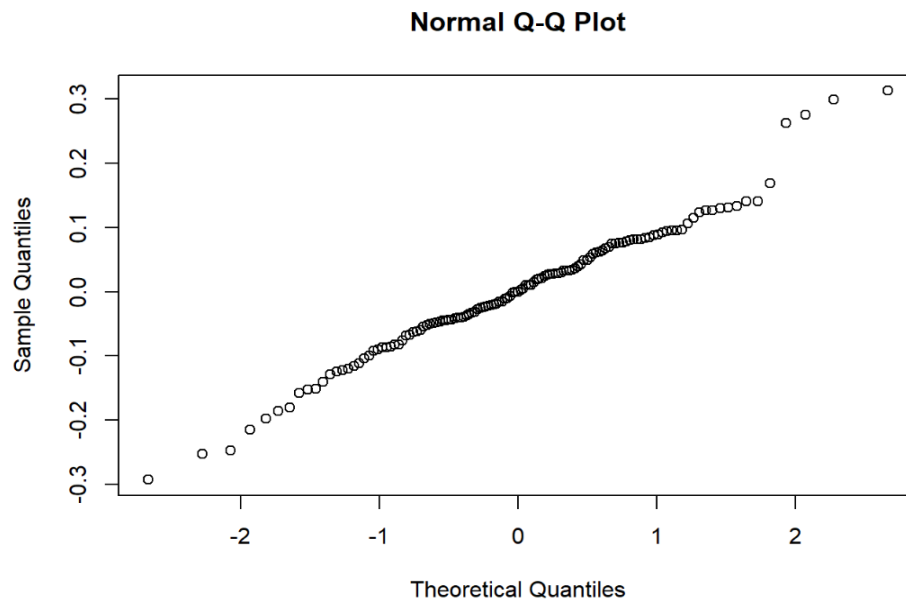
```
## Null hypothesis: Residuals are iid noise.
## Test                Distribution Statistic    p-value
## Ljung-Box Q          Q ~ chisq(20)         61.24         0 *
## McLeod-Li Q          Q ~ chisq(20)         37.7          0.0096 *
## Turning points T      (T-86)/4.8 ~ N(0,1)       97          0.0217 *
## Diff signs S          (S-65)/3.3 ~ N(0,1)        63          0.5465
## Rank P                (P-4257.5)/251.3 ~ N(0,1)  4124         0.5953
```

검정

독립성 검정 정규성 검정 정상성 검정

2. 정규성 검정

QQ plot 그려봐서 확인, kolmogrov-Smirnov test, Jarque-Bera test



정규분포를 따르는 경우
직선에 가깝게 그려짐!

검정

독립성 검정 정규성 검정 **정상성 검정**

3. 정상성 검정

kpss test

→ H_0 : 정상성을 가진다.귀무가설이 **반대**

adf test, pp test

→ H_0 : 비정상성을 가진다.

```
adf.test(diffall)
```

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

```
##
## Augmented Dickey-Fuller Test
##
## data: diffall
## Dickey-Fuller = -5.1619, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

```
kpss.test(diffall)
```

```
## Warning in kpss.test(diffall): p-value greater than printed p-value
```

```
##
## KPSS Test for Level Stationarity
##
## data: diffall
## KPSS Level = 0.081634, Truncation lag parameter = 4, p-value = 0.1
```