20기 정규세션

ToBig's 19기 강의자

최태순

# Time Series Analysis

## 시계열 분석

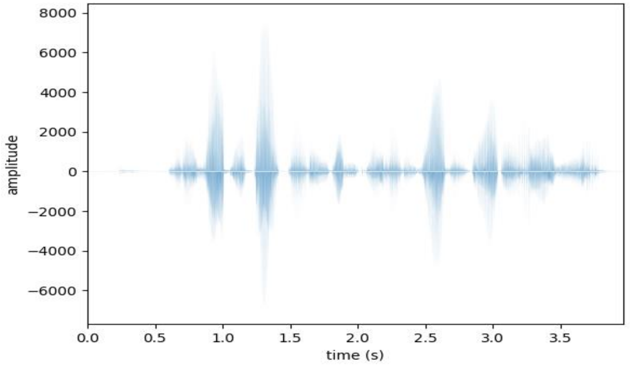
# Contents

Contents

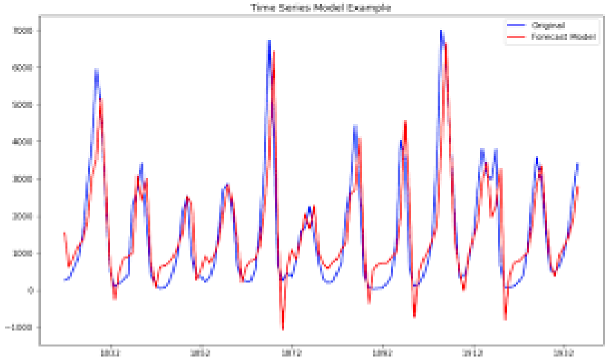
# 시계열 자료란?

- 주가 예측


- 음성 분석




- 수요 예측



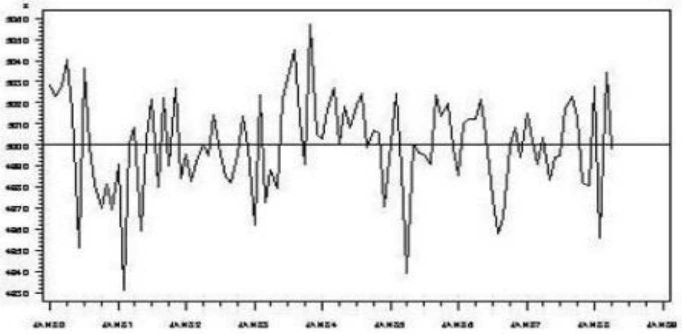

**시간의 흐름에 따라 측정한 자료**

- 수집된 시계열자료들을 분석하여 미래 예측
- 시계열자료가 생성된 시스템 또는 확률과정을 모형화하여 이를 이해하고 제어할 수 있도록 함

# 시계열 자료의 성분

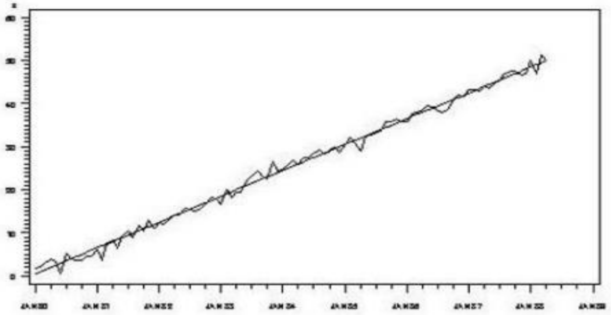
체계적 성분

## 불규칙 성분



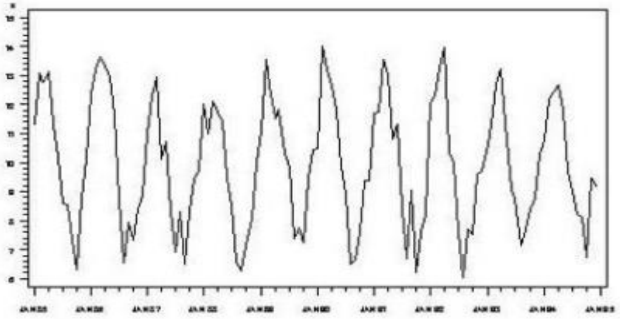
- 시간에 따른 규칙성이 존재하지 않음
- 랜덤한 원인에 의해 발생

## 추세 성분



- 시간에 따라 지속적으로 증가 혹은 하강하는 추세를 갖음
- 여러 형태를 가질 수 있음

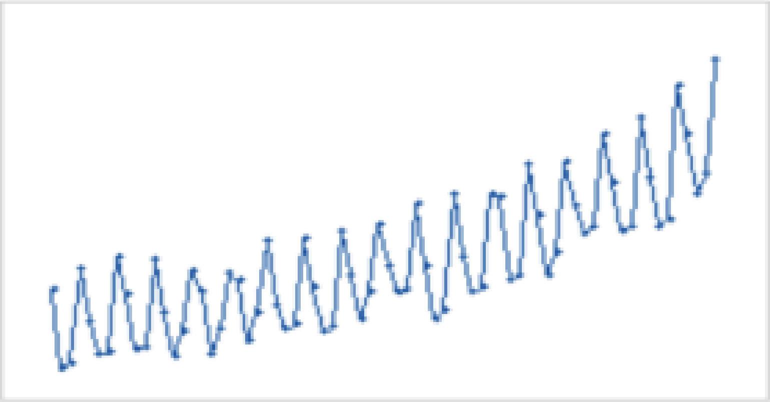
## 계절(순환) 성분



- 일정한 주기를 갖고 변화함
- 계절성분은 일반적으로 주기가 짧음
- 순환성분은 계절성분에 비해 주기가 긴 성분

# 시계열 기본모형(모델링)

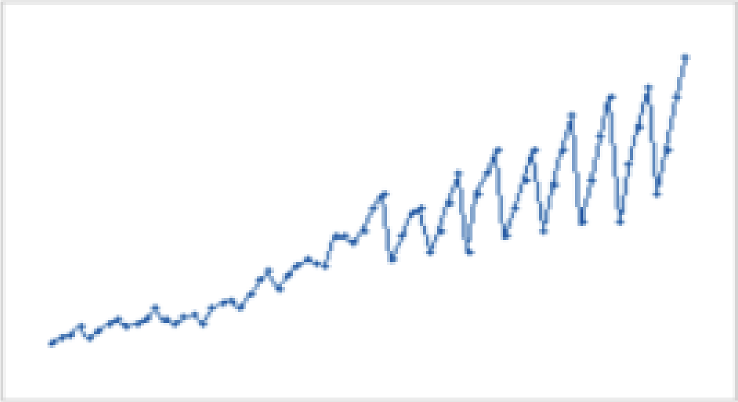
- 가법모형



$$Z_t = T_t + C_t + S_t + I_t$$

- 관측값이 네가지 변동 성분의 합으로 구성됐다는 가정 하에 분석
- 계절성분의 진폭이 시계열의 수준에 관계없이 일정할 때 주로 사용

- 승법모형



$$Z_t = T_t \times C_t \times S_t \times I_t$$

- 관측값이 네가지 변동 성분의 곱으로 구성됐다는 가정 하에 분석
- 계절성분의 진폭이 시계열의 수준에 따라 달라질 때 주로 사용

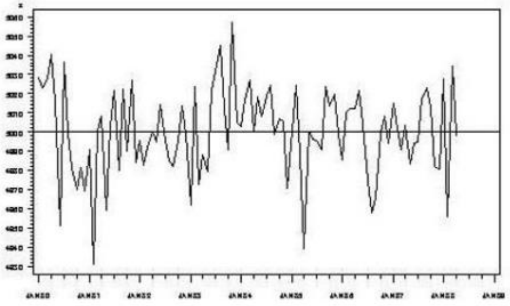
$Z_t$ : 관측값
$T_t$ : 추세 성분
$C_t$ : 순환 성분
$S_t$ : 계절 성분
$I_t$ : 불규칙 성분

Contents

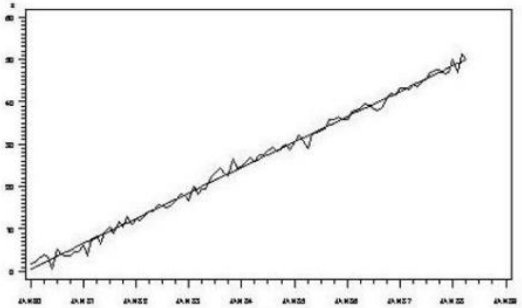
# 시계열 분석 기법(1) - 추세

관측값 $Z_t$를 시간의 함수로서 표현한 방법

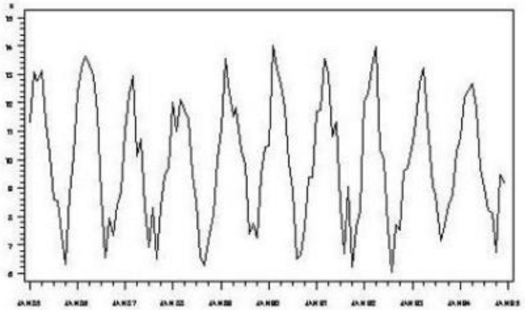
$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t$ **->** 선형다항추세모형**(linear polynomial trend model)**



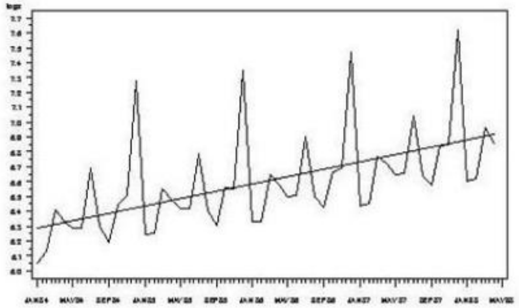
- 상수평균모형



- 선형추세모형



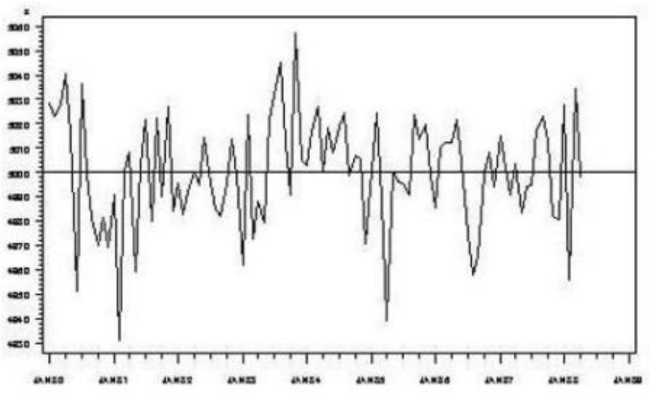
- 계절추세모형



- 선형계절추세모형

# 시계열 분석 기법(1) - 추세

## • 상수평균모형



불규칙성분만을 갖는경우

$Z_t = \beta_0 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$

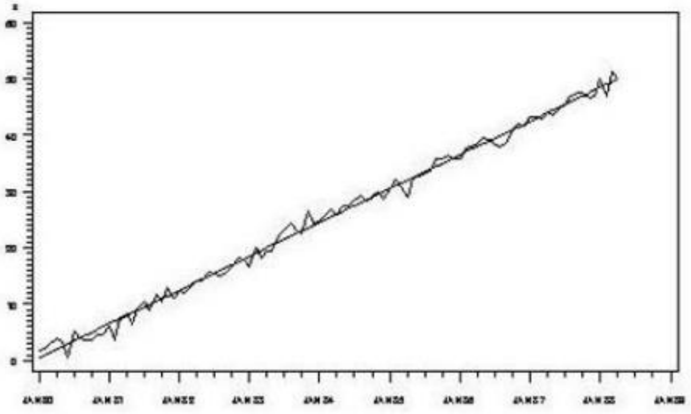
① 모수 $\beta_0$의 LSE : $\widehat{\beta_0} = \frac{1}{n}\sum_{t=1}^{n} Z_t = \bar{Z}$

② 시점 n을 원점으로 했을 때 $l-$시차 후 추정된 MMSE 예측값 : $\hat{Z}_n(l) = \widehat{\beta_0} = \bar{Z}, \quad l = 1,2,\cdots$

③ 예측오차 : $\widehat{e_n}(l) = Z_{n+1} - \hat{Z}(l) = \beta_0 + \varepsilon_{n+l} - \bar{Z}$

# 시계열 분석 기법(1) - 추세

## • 선형추세모형



추세성분만을 갖는경우

$$Z_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$
$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$
$$\vdots$$
$$Z = X\beta + \epsilon$$

## • 선형추세모형 분석의 단계적 방법

① 시계열자료에 대한 추세그래프를 통해 n차 추세모형 적합

② 잔차 시계열 그림을 통해 오차의 2가지 가정 확인

③ 만약 잔차 시계열 그림에서 이분산성이 보일 경우 분산 안정화 변환 실시(로그변환, Box-Cox transformation 등)

④ 오차의 독립성 가정의 검정을 위해 잔차 그래프의 형태 오차항의 자기상관관계 파악 가능 (Durbin-Watson검정 실시)

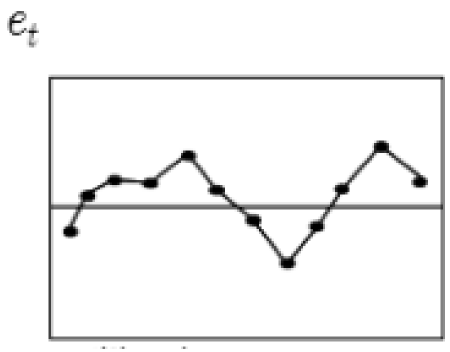
⑤ 자기상관관계가 있을 경우 독립성 가정을 만족하지 못하므로 오차의 공분산을 이용한 일반화최소제곱법(GLSE)를 이용 or 자기회귀오차모형(AR모형)을 적용해 이를 해결할 수 있음.

# 시계열 분석 기법(1) - 추세

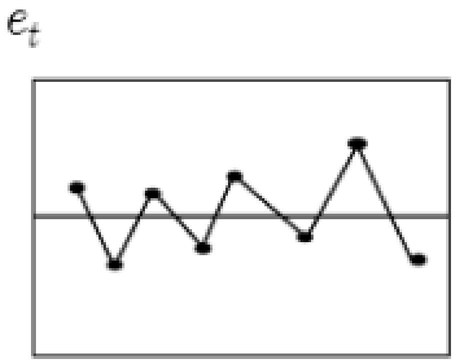
- **Durbin-Watson 검정(DW 검정) :** 오차항들 간의 1차 자기상관 존재 여부에 대한 유의성 검정

$$\widehat{\rho_k} = \frac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{n} e_t^2}, \quad k = 1,2,\cdots$$ (k차 표본자기상관계수)

$|\widehat{\rho_k}| > 2n^{-(\frac{1}{2})}$ 이면 k-시차의 자기상관관계가 존재함으로 판단



양의 상관관계

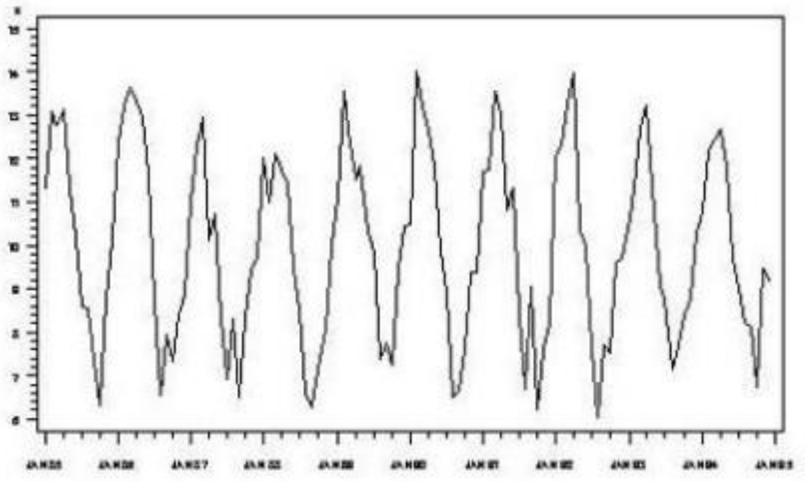


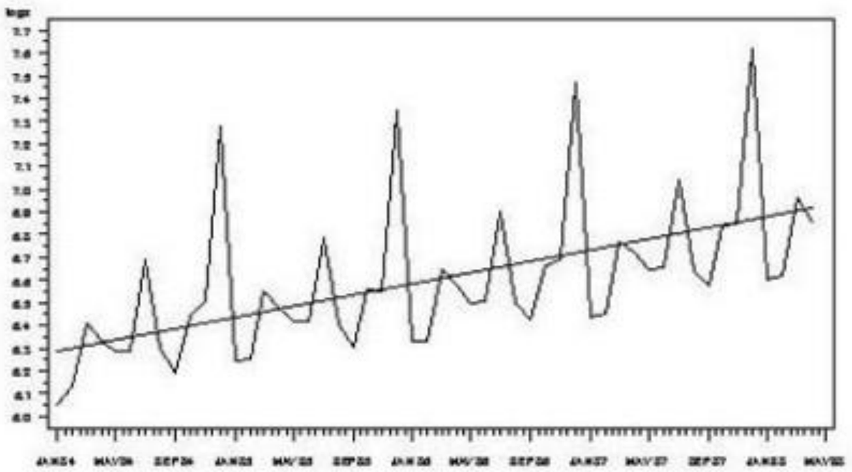
음의 상관관계

위와 같은 자기상관계수를 통해 유의성 검정 실시

- 가설 "$H_0$: $\emptyset = 0$ $vs$ $H_1$: $\emptyset \neq 0$"
- D(DW 검정통계량) $\simeq 2(1 - \widehat{\rho_1})$

# 시계열 분석 기법(1) - 추세





- **계절추세모형 – 삼각함수 또는 지시함수 이용**

삼각함수 : $Z_t = \beta_0 + \sum_{i=1}^{m} A_i \sin\left(\frac{2\pi i}{s} t + \emptyset_i\right) + \epsilon_t$

지시함수 : $Z_t = \beta_0 + \sum_{i=1}^{s} \beta_i IND_{ti} + \epsilon_t$

$A_i$ : 진동폭(amplitude)

$\phi_i$ : 편각 (phase shift)

$f_i = 2\pi/s$ : 제 $i$번째 주기항의 빈도(frequency)

$s/i$ : 제 $i$번째 주기항의 주기(period)

$\epsilon_t$ : 서로 독립이고 평균 0, 분산 $\sigma_\epsilon^2$을 갖는 오차항

$\beta_0$, $A_i$ 와 $\phi_i$ : 미지의 모수

- **선형계절추세모형**

삼각함수 : $Z_t = \beta_0 + \beta_1 t + \beta_2 \sin(\frac{2\pi t}{12}) + \beta_3 \cos(\frac{2\pi t}{12}) + \epsilon_t$

지시함수 : $Z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{s} \beta_{si} IND_{ti} + \epsilon_t$

# 시계열 분석 기법(2) - 평활

**평활법** : 시계열 자료의 특징인 다양한 변화(추세, 계절, 불규칙)로 인해 생기는 효과를 줄이고 자료를 평탄(smoothing)하게 변환시키는 방법

평활법

지수평활법

- 최근 자료에 더 큰 가중값, 과거로 갈수록 가중값을 지수적으로 줄여나가는 방법
- 시스템에 변화가 있을 때 변화에 쉽게 대처가능
- 계산이 쉽고 예측이 주목적
- 단순지수평활법, 이중지수평활법, Winter의 계절지수 평활법 등

이동평균법

- 평활에 의해 계절성분 또는 불규칙성분을 제거하여 전반적인 추세를 뚜렷하게 파악
- 관측값 전부에 동일한 가중값을 주는 대신에 일부 자료에만 동일한 가중값을 줌
- 분해법에서 계절조정을 하는데 주로 사용
- 단순이동평균법, 선형이동평균법 등

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 단순 지수 평활법

$Z_t = \beta_0 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$ ($\beta_0$는 시간에 따라 변화하는 모수값)

- 시계열 자료를 위와 같은 모형으로 가정
- 새로운 자료가 관측될 때마다 변화의 정보를 수용하여 $\beta_0$의 추정값을 갱신

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 단순 지수 평활법

- ✓ $Z_t = \beta_0 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$ ($\beta_0$는 시간에 따라 변화하는 모수값)

- ✓ $\hat{Z}_{n-1}(1) = \hat{\beta}_{0,n-1} = \frac{1}{n-1}\sum_{t=1}^{n-1} Z_t = \bar{Z}_{n-1}$
  - 시점 n-1로부터 1시차 후의 MMSE 예측값(미래 n시점의 예측)

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 단순 지수 평활법

✓ $Z_t = \beta_0 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$ ($\beta_0$는 시간에 따라 변화하는 모수값)

✓ $\hat{Z}_{n-1}(1) = \hat{\beta}_{0,n-1} = \frac{1}{n-1}\sum_{t=1}^{n-1} Z_t = \bar{Z}_{n-1}$

- 시점 n-1로부터 1시차 후의 MMSE 예측값(미래 n시점의 예측)

✓ $\hat{e}_{n-1}(1) = Z_n - \hat{Z}_{n-1}(1) = Z_n - \hat{\beta}_{0,n-1}$

- 예측오차

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 단순 지수 평활법

✓ $Z_t = \beta_0 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$  ($\beta_0$는 시간에 따라 변화하는 모수값)

✓ $\hat{Z}_{n-1}(1) = \hat{\beta}_{0,n-1} = \frac{1}{n-1}\sum_{t=1}^{n-1} Z_t = \bar{Z}_{n-1}$

- 시점 n-1로부터 1시차 후의 MMSE 예측값(미래 n시점의 예측)

✓ $\hat{e}_{n-1}(1) = Z_n - \hat{Z}_{n-1}(1) = Z_n - \hat{\beta}_{0,n-1}$

- 예측오차

✓ $\hat{\beta}_{0,n} = \hat{\beta}_{0,n-1} + \omega \hat{e}_{n-1}(1) = \hat{\beta}_{0,n-1} + \omega(Z_n - \hat{\beta}_{0,n-1})$

- 새 관측값 $Z_n$ 관측 후, 시점 n에서의 $\beta_0$의 추정값을 예측오차를 이용해 수정

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 단순 지수 평활법

$\hat{\beta}_{0,n} = S_n^{(1)}$ 이라 놓으면

$$S_n^{(1)} = S_{n-1}^{(1)} + \omega(Z_n - S_{n-1}^{(1)}) = \omega Z_n + (1-\omega)S_{n-1}^{(1)}$$

$= \ \vdots$ (수학적 귀납법 이용)

$$= \omega \sum_{j=0}^{n-1}(1-\omega)^j Z_{n-j} + (1-\omega)^n S_0^{(1)}$$

시점 n에서 1-시차 후 예측값 : $\boldsymbol{\hat{Z}_n(1) = \hat{\beta}_{0,n} = S_n^{(1)}}$

$\omega$ : 평활상수(smoothing constant)

$S_0^{(1)}$ : 초기 평활값

$S_n^{(1)}$ : 단순지수평활(simple exponential smoothing)통계량

$$\lim_{n\to\infty} \omega\sum_{j=0}^{n-1}(1-\omega)^j = \frac{\omega}{1-(1-\omega)} = 1$$

$$\lim_{n\to\infty}(1-\omega)^n S_0^{(1)} = 0$$

$S_n^{(1)} = \omega\sum_{j=0}^{\infty}(1-\omega)^j Z_{n-j}$ : 시점 $n$까지의 자료의 가중평균(weighted average)값

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 단순 지수 평활법

- **초기평활값 $S_0^{(1)}$ 선택방법**

(1) Brown(1962)과 Montgomery와 Johnson(1976) : $S_0^{(1)} = \frac{1}{n}\sum_{t=1}^{n} Z_t = \bar{Z}$
(2) Makridakis 와 Wheelwright(1978) : $S_0^{(1)} = Z_1$
(3) Abraham(1986) : $Z_0$의 후향예측값(back forecast) $S_1^*$

- **평활상수 $\omega$의 선택**

(1) Brown(1962) : 0.05 ~ 0.3
(2) Montgomery(1976) : $\omega$ = 1-$0.8^{1/trend}$ (이때, trend는 평활의 차수, 단순지수에서는 1)
(3) SSE($\omega$)를 최소로하는 $\omega$

# 시계열 분석 기법(2) - 평활

## 지수평활법 – 이중 지수 평활법

➢ **모형 가정**

$$Z_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$

➢ **지수 평활통계량**

$$S_n^{(1)} = \omega Z_n + (1-\omega) S_{n-1}^{(1)}$$
$$S_n^{(2)} = \omega S_n^{(1)} + (1-\omega) S_{n-1}^{(2)}$$

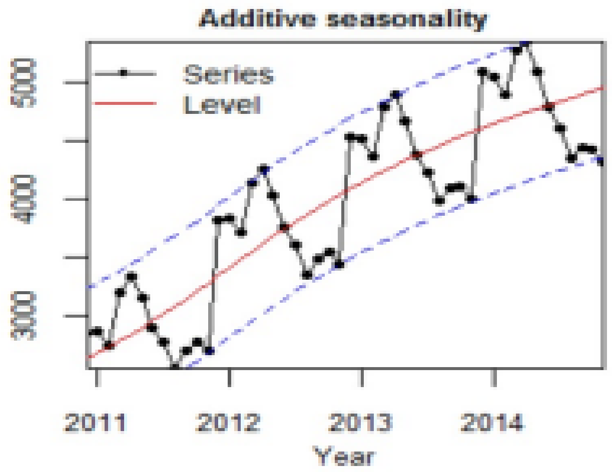
➢ **모수 추정량**

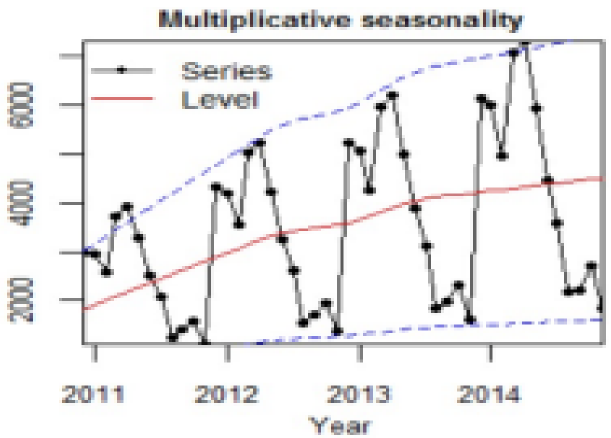
$$\hat{\beta}_{0,n} = 2S_n^{(1)} - S_n^{(2)} - \hat{\beta}_{1,n} n$$
$$\hat{\beta}_{1,n} = \frac{\omega}{1-\omega}\{S_n^{(1)} - S_n^{(2)}\}$$

# 시계열 분석 기법(2) - 평활

## 지수평활법 – Holt-Winters 계절 지수 평활법

- Holt-Winters 계절 지수 평활법 : 관측치가 평균과 추세와 계절성을 갖는 시계열 데이터로 가정
- 계절 변동의 두 가지 형태
  - 가법적 계절변동 : 시간에 따른 계절적 진폭의 크기 일정
  - 승법적 계절변동 : 계절적 진폭의 크기가 평균 수준에 비례하여 점차적으로 증가하거나 감소



# 시계열 분석 기법(2) - 평활

## 이동평균법 – 단순 이동 평균

✓ $Z_t = \beta_0 + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$ ($\beta_0$는 시간에 따라 변화하는 모수값)

✓ $M_n^{(1)} = \frac{Z_n + Z_{n-1} + \cdots + Z_{n-m+1}}{m} = \sum_{t=n-m+1}^{n} Z_t / m$

- n시점에서 최근 m개의 관측값만을 이용해 구한 추정량

✓ $M_{n+1}^{(1)} = \frac{Z_{n+1} + Z_n + \cdots + Z_{n-m+2}}{m} = M_n^{(1)} + (Z_{n+1} - Z_{n-m+1})/m$

- 새로운 값 $Z_{n+1}$이 추가될 경우 추정량(예측값) 갱신

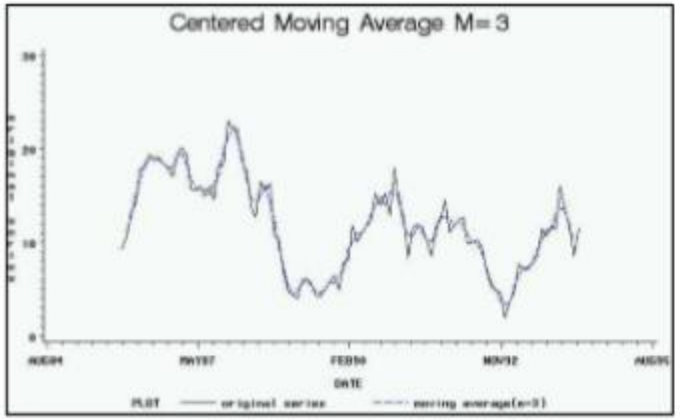
**직전에 구한 예측값($M_n^{(1)}$)과 가장 최근 관측값($Z_{n+1}$), 그리고 m-시차 전 관측값($Z_{n-m+1}$)으로 쉽게 갱신가능!!**

# 시계열 분석 기법(2) - 평활

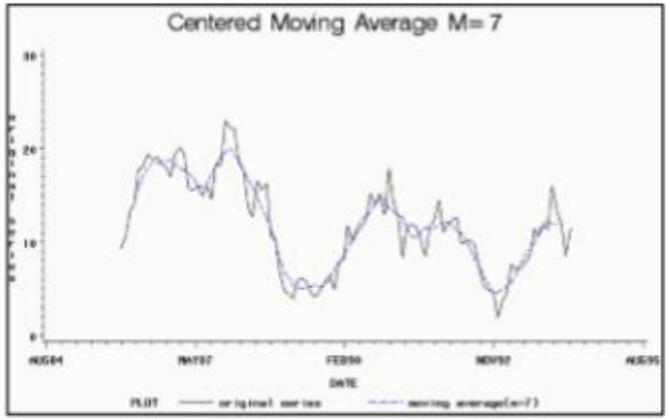
## 이동평균법 – 단순 이동 평균

- **이동평균에서의 m 역할**

m의 값이 크면 $M_n^{(1)}$의 값은 지엽적인 변화에 둔감하게 서서히 변함
m의 값이 작으면 $M_n^{(1)}$의 값은 지엽적인 변화를 빨리 반영함




( m=3 )




( m=7 )

m↑ = 평활 효과 ↑ = 지엽적 x
m↓ = 평활 효과 ↓ = 지엽적 o

# 시계열 분석 기법(2) - 평활

## 이동평균법 – 선형(이중) 이동 평균

### ➢ 모형 가정

$$Z_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$

### ➢ 이동평균 통계량

$$M_n^{(1)} = \frac{Z_n + Z_{n-1} + \cdots + Z_{n-m+1}}{m} = \sum_{t=n-m+1}^{n} Z_t / m$$

$$M_n^{(2)} = \frac{M_n^{(1)} + M_{n-1}^{(1)} + \cdots + M_{n-m+1}^{(1)}}{m} = \sum_{t=n-m+1}^{n} M_t^{(1)} / m$$

### ➢ 모수 추정량

$$\hat{\beta}_{0,n} = 2M_n^{(1)} - M_n^{(2)} - \hat{\beta}_{1,n} n$$

$$\hat{\beta}_{1,n} = \frac{2}{m-1}\{M_n^{(1)} - M_n^{(2)}\}$$

# 시계열 분석 기법(3) - 분해

**분해법 :** 체계적 성분과 불규칙 성분으로 시계열 자료를 분해하여 분석하는 방법

### ➢ 시계열변동 성분

- 불규칙 성분 ($I_t$)
- 추세 성분 ($T_t$)
- 계절 성분 ($S_t$)
- 순환 성분 ($C_t$)

### ➢ 분해법의 목적

- 미래 예측
- 계절조정

### ➢ 분해법의 기본모형

- 가법모형 : $Z_t = T_t + C_t + S_t + I_t$
- 승법모형 : $Z_t = T_t \times C_t \times S_t \times I_t$

***** 로그 가법모형 **:** $lnZ_t = lnT_t + lnC_t + lnS_t + lnI_t$

# 시계열 분석 기법(3) - 분해

## 분해법 – 추세분석을 이용한 분해

- 추세성분($T_t$)과 계절성분 ($S_t$)을 다음과 같이 정의

$T_t = \beta_0 + \sum_{i=1}^{k} \beta_i t^i$
$S_t = \sum_{i=1}^{s} \delta_i IND_{ti}$

- 추세성분($T_t$)과 계절성분 ($S_t$) 이 서로 독립 가정
- 가법모형에서는 각 성분의 차이로 계산
- 승법모형에서는 각 성분의 나누기로 계산

**추세분석 분해법 단계**

1) 원시계열 $\{Z_t\}$에 추세모형 $Z_t = \beta_0 + \sum_{i=1}^{k} \beta_i t^i + \varepsilon_t$ 적합하여 추세성분 추정값 $\widehat{T}_t$ 구함

2) $\{Z_t\}$에서 $\{\widehat{T}_t\}$를 뺀 잔차계열 $\{Z_t - \widehat{T}_t\}$에 계절추세모형 $Z_t - \widehat{T}_t = S_t + \varepsilon_t = \sum_{i=1}^{s} \delta_i IND_{ti} + \varepsilon_t$ 을 적합시켜 계절성분 추정값 $\widehat{S}_t$ 구함

3) $\{\widehat{I}_t\}$=$\{Z_t - \widehat{T}_t - \widehat{S}_t\}$의 불규칙성분 추정값을 통해 잔차분석 실시하여 각 성분에 대한 정보가 남아있는지 검토

# 시계열 분석 기법(3) - 분해

## 분해법 – 이동평균을 이용한 분해

이동평균을 통해 원시계열에서 계절성분과 불규칙성분을 동시에 평활시켜줄 수 있음!!

### 이동평균 분해법 단계

1) 원시계열 $\{Z_t\}$에 계절성분 주기 s에 적절한 항의 이동평균을 적용해 계절성분과 불규칙성분이 제거된 추세·순환성분 추정값 $\widehat{T_t + C_t}$를 구함

2) $\{Z_t\}$에서 $\{\widehat{T_t + C_t}\}$를 빼면 계절·불규칙 성분 $\widehat{S_t + I_t}$를 구할 수 있음

3) 계절·불규칙 성분의 추정계열$\{\widehat{S_t + I_t}\}$에 계절성분의 주기 s와 일치하지 않은 개수 항의 이동평균을 적용시키면 불규칙성분이 제거되어 계절성분의 추정계열 $\{\widehat{S_t}\}$만 남게 됨

4) 추세·순환성분 $\widehat{T_t + C_t}$를 종속변수로 하는 추세다항식을 적합시켜 추정한 값들을 추세성분의 추정값 $\widehat{T_t}$로 함

5) 추세·순환성분 $\widehat{T_t + C_t}$에서 추세성분의 추정값 $\widehat{T_t}$을 빼면 순환성분 추정값 $\widehat{C_t}$을 얻을 수 있음

6) $\{\widehat{I_t}\}$=$\{Z_t - \widehat{T_t} - \widehat{C_t} - \widehat{S_t}\}$의 불규칙성분 추정값을 통해 잔차분석 실시하여 각 성분에 대한 정보가 남아있는지 검토

# Contents

# • 확률과정이란?

= 확률법칙에 의해 생성되는 일련의 통계적 현상
= 시간 t를 축으로하는 확률변수 $Z_t$의 집합 즉, 시간(t)에 따라 확률값 $Z_t$이 변화하는 과정을 의미

# • 확률과정의 예

1. 백색잡음 과정 : 서로 독립이고 동일한 분포를 따르는 i.i.d 확률변수들의 계열로 구성된 확률과정

   $Z_t = \varepsilon_t, \quad t = 1,2,\cdots, \varepsilon_t$ : 백색잡음 과정 WN(0, $\sigma_\epsilon^2$)
2. 확률보행과정 : $Z_t = Z_{t-1} + \varepsilon_t,\ Z_0 = 0,\ \ t = 1,2,\ \cdots$
3. 이동평균과정 : $Z_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1},\ Z_0 = 0,\ \ t = 1,2,\ \cdots$
4. 선형과정 : $Z_t = \mu + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \varphi_2\varepsilon_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \varphi_j\, \varepsilon_{t-j},\ \varphi_0 = 1$
5. 자기회귀과정 : $Z_t - \mu = \emptyset(Z_{t-1} - \mu) +\ \varepsilon_t$

- **정상성**

= 다양한 확률과정 중 특정한 확률과정에서 실현된 시계열 모형의 조건
= 뚜렷한 추세가 존재하지 않고 시간(t)의 흐름에 따라 시계열의 진폭(변동)이 일정

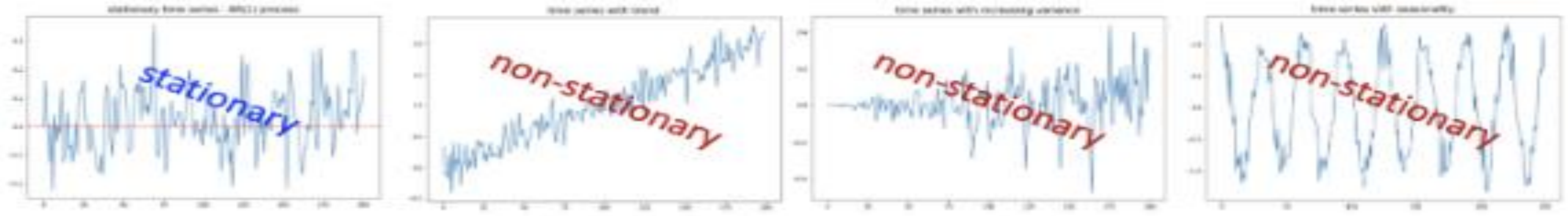
1. 강정상성(strict stationarity) : 임의의 자연수 $t_1, t_2, \cdots, t_n$와 k에 대해 $f(Z_{t1}, Z_{t2}, \cdots, Z_{tn}) = f(Z_{t1+k}, Z_{t2+k}, \cdots, Z_{tn+k})$
   -> 매우 복잡하고 유한개의 관측값만을 갖고는 이를 보이는 것은 현실적으로 불가능

2. 약정상성(weak stationarity) : 확률과정이 유한한 2차 적률($E(X^2) < \infty$)을 갖고 아래의 3가지를 만족
✓ $E(Z_t) = \mu$ (시간 t에 관계없이 동일)
✓ Var($Z_t$) = $\sigma^2$ (시간 t에 관계없이 동일)
✓ $E[(Z_t - \mu)(Z_{t+k} - \mu)] = Cov(Z_t, Z_{t+k}) = \gamma_k$ : 시차(lag) k에만 의존하고 시점 t와는 무관

시계열 과정에서의 **정상성 = 약정상성**

# • 정상성 판단

1. 목측법 : 시계열 자료의 그래프를 통해 추세, 계절성 등을 확인하여 정상성 여부를 판단할 수 있음




2. 자기 상관 함수(autocorrelation function)

- **자기 상관 함수(ACF, SACF)**

: k-시차 이후의 관측치 간의 상관계수 함수

$$ACF(k) = \rho_k = Corr(Z_t, Z_{t+k}) = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t) \cdot Var(Z_{t+k})}}$$

- **부분 자기 상관 함수(PACF, SPACF)**

: t 시차에 대해, 관측치 $Z_t$와 $Z_{t+k}$ 사이의 관측치 효과를 제거한 상관계수 함수 (시간의 효과 제거)

$$PACF(k) = \emptyset_{kk} = Corr(Z_t^*, Z_{t+k}^*)$$

: 최적 선형예측법, Cramer 공식, Durbin-Sevinson 알고리즘 이용

# • 확률적 시계열 모형

✓ Auto Regressive(AR) = 자기회귀과정
: t 시점에서의 관측값($Z_t$)이 과거 p개의 관측값들의 함수 형태로 나타난 확률과정 .......(가역성 항상 만족)

$$Z_t - \mu = \emptyset_1(Z_{t-1} - \mu) + \cdots + \emptyset_p(Z_{t-p} - \mu) + \varepsilon_t = \sum_{n=1}^{p} \emptyset_n(Z_{t-n} - \mu) + \varepsilon_t$$

✓ Moving Average(MA) = 이동평균과정
: t 시점에서의 관측값($Z_t$)이 과거 q개의 오차들의 함수 형태로 나타난 확률과정 ......(정상성 항상 만족)

$$Z_t - \mu = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

✓ Auto Regressive – Moving Average(ARMA) = 자기회귀이동평균과정
: t 시점에서의 관측값($Z_t$)이 과거 p개의 관측값들의 함수와 과거 q개의 오차들의 함수가 합쳐진 형태로 나타난 확률과정

$$Z_t - \mu = \emptyset_1(Z_{t-1} - \mu) + \cdots + \emptyset_p(Z_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

# • AR, MA의 쌍대성

✓ 유한차수의 정상 AR 과정은 무한차수의 MA과정으로 표현 가능
유한차수의 가역성을 가지는 MA과정은 무한차수의 AR과정으로 표현 가능

✓ 유한차수의 AR과정의 ACF와 유한차수의 MA과정의 PACF는 지수적으로 감소하는 형태
유한차수의 AR과정의 PACF와 유한차수의 MA과정의 ACF는 절단형태

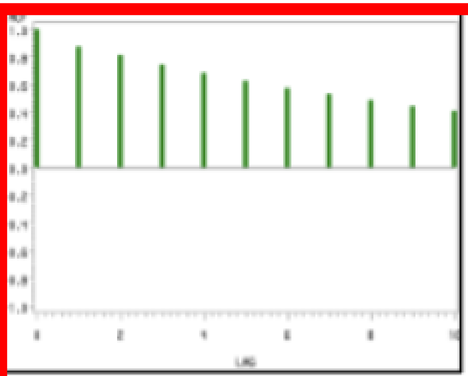
✓ 유한차수의 AR과정 정상성 조건 : AR작용소가 0이되는 근($\emptyset(B) = 0$)들이 단위원 밖에 존재
유한차수의 MA과정 가역성 조건 : MA작용소가 0이되는 근($\theta(B) = 0$)들의 단위원 밖에 존재

# • Correlogram

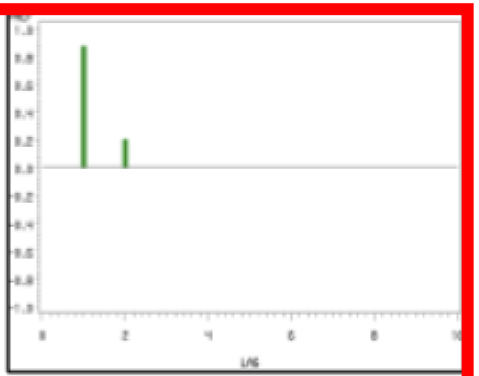
| 확률과정 | ACF | PACF |
|---|---|---|
| $AR(p)$ | 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차 $p$ 이후에는 0으로의 절단 형태 |
| $MA(q)$ | 시차 $q$ 이후에는 0으로의 절단 형태 | 지수적으로 감소하거나 소멸하는 싸인함수 형태 |
| $ARMA(p, q)$ | 시차$(q-p)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차$(p-q)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 |

# Correlogram

| 확률과정 | $ACF$ | $PACF$ |
|---|---|---|
| $AR(p)$ | 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차 $p$ 이후에는 0으로의 절단 형태 |
| $MA(q)$ | 시차 $q$ 이후에는 0으로의 절단 형태 | 지수적으로 감소하거나 소멸하는 싸인함수 형태 |
| $ARMA(p, q)$ | 시차$(q-p)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차$(p-q)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 |



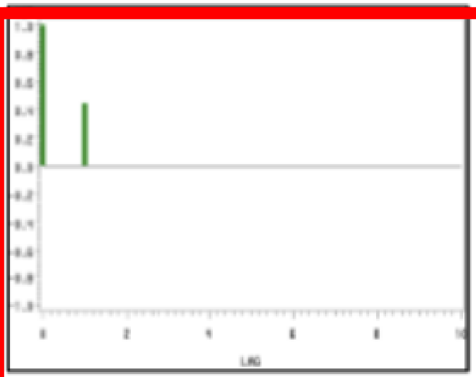
(e) $ACF$ $\phi_1 = 0.7$ $\phi_2 = 0.2$



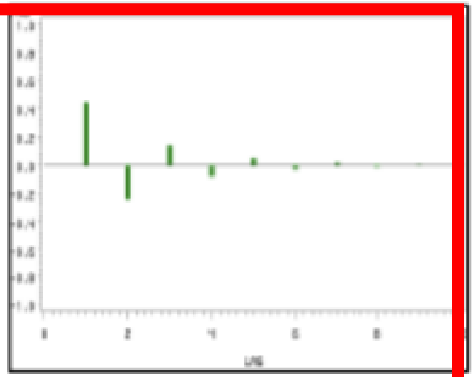
(f) $PACF$ $\phi_1 = 0.7$ $\phi_2 = 0.2$

# • Correlogram

| 확률과정 | $ACF$ | $PACF$ |
|---|---|---|
| $AR(p)$ | 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차 $p$ 이후에는 0으로의 절단 형태 |
| $MA(q)$ | 시차 $q$ 이후에는 0으로의 절단 형태 | 지수적으로 감소하거나 소멸하는 싸인함수 형태 |
| $ARMA(p, q)$ | 시차$(q-p)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차$(p-q)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 |



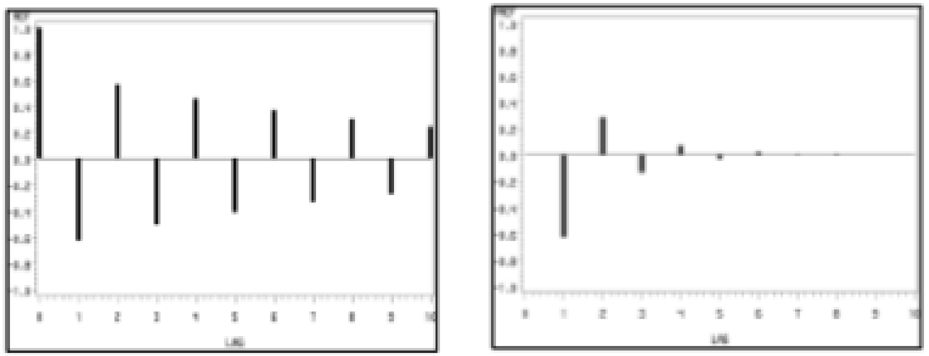
(c) $ACF$ $\theta=-0.6$



(d) $PACF$ $\theta=-0.6$

# Correlogram

| 확률과정 | *ACF* | *PACF* |
|---|---|---|
| $AR(p)$ | 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차 $p$ 이후에는 0으로의 절단 형태 |
| $MA(q)$ | 시차 $q$ 이후에는 0으로의 절단 형태 | 지수적으로 감소하거나 소멸하는 싸인함수 형태 |
| $ARMA(p, q)$ | 시차$(q-p)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 | 시차$(p-q)$ 이후에는 지수적으로 감소하거나 소멸하는 싸인함수 형태 |



(a) *ACF* $\phi=-0.9$ $\theta=-0.5$ (b) *PACF* $\phi=-0.9$ $\theta=-0.5$

# • 비정상성 시계열

✓ 시계열의 수준이 시간대에 따라 다름
✓ 시계열이 추세를 갖음
✓ 시계열이 계절성을 보임
✓ 시계열의 분산이 시간대에 따라 변함(이분산성)

1차 차분 $Z_t' = Z_t - Z_{t-1} = \nabla^1 Z_t$
2차 차분 $Z_t'' = Z_t' - Z_{t-1}' = \nabla^2 Z_t$

d차 차분 $Z_t'^{*d} = Z_t'^{*(d-1)} - Z_{t-1}'^{(d-1)} = \nabla^d Z_t$

# • 비정상성 시계열의 정상화

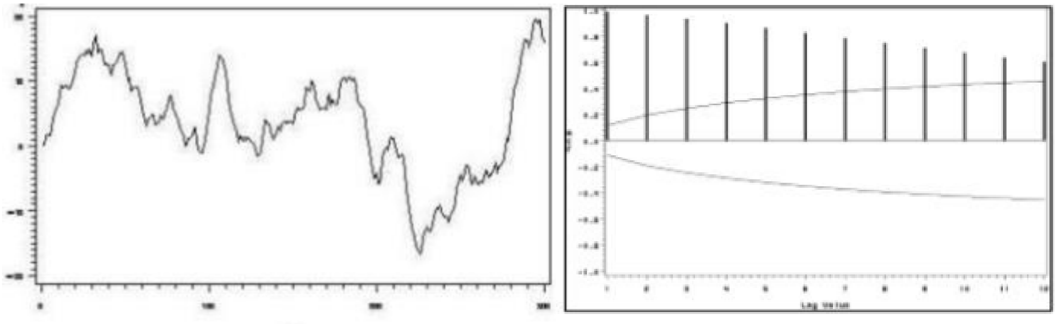
✓ 이분산성 : 분산안정화 변환

✓ 추세 :
- 결정적 추세 –선형추세모형 or 분해법(추세성분과 계절성분을 추출하고 불규칙성분에 대해 ARMA모형 적합)
- 확률적 추세 – **차분** (결정적 추세도 가능)

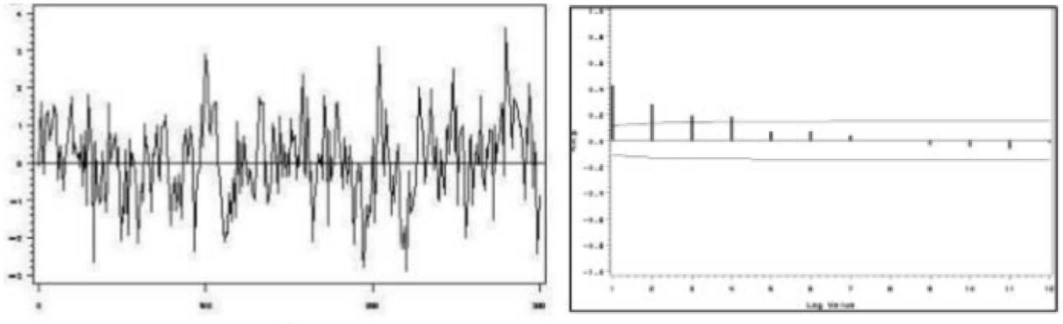
# • ARIMA 모형(자기회귀 누적 이동평균 과정)

- D차 차분 후의 $Z_t^{\prime * d}$수준이 ARMA(p,q) 과정을 따를 때 기존의 관측값 $\{Z_t\}$는 ARIMA(p,d,q) 과정을 따른다고 함
- 차분의 차수 d는 주로 시계열그림과 SACF를 참고하여 결정 – 시계열이 추세를 갖거나 SACF가 서서히 감소하는 경우에 차분이 필요하다고 볼 수 있음

ARIMA(1,1,1) 과정 시계열 그림



ARIMA(1,1,1) 과정의 시계열을 1차 차분한 후의 그림



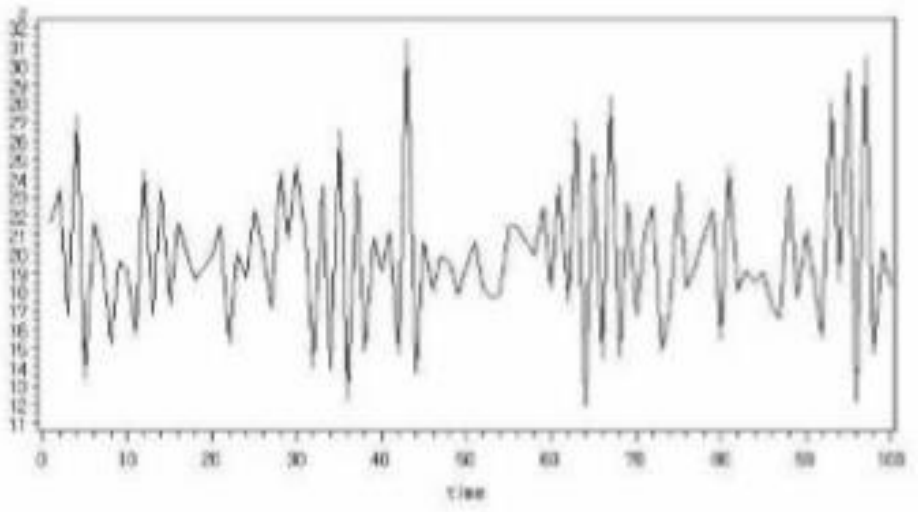
Contents

# Unit 04 | Box-Jenkins 방법론

| | |
|---|---|
| 식별 | • 자료의 시계열도, ACF, PACF 등 이용<br>• 차분의 필요 여부와 모형의 차수를 잠정적으로 결정 |
| 추정 | • 최소제곱법, 최대가능도법, 비선형최소제곱법 등 이용 |
| 진단 | • 잔차의 시계열도, ACF, PACF 및 포트맨토 통계량을 이용한 잔차분석<br>• AIC, SBC 등의 통계값 등을 이용해 가장 설명력 높은 모형 선택 |
| 예측 | • 적합한 최종모형으로 이후 값 예측 |

식별

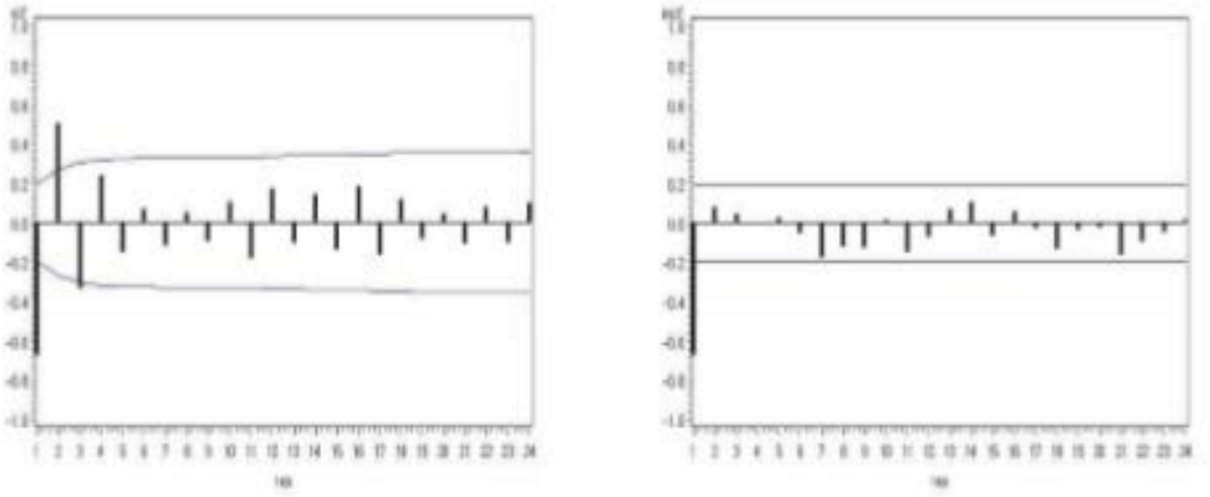
- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정



✓ 데이터를 이용해 시계열도를 그려 정상성을 확인!
평균이 일정하고 불규칙적인 모습을 보이므로 정상 시계열로 판단가능

식별

- 자료의 시계열도, ACF, PACF 등 이용
- 차분의 필요 여부와 모형의 차수를 잠정적으로 결정



**✓ ACF, PACF 통해 모형 식별!**
**ACF가 Lag 3 이후로 구간안에 위치하므로 MA(3) 모형을 잠정모형으로 채택!**
**PACF가 Lag 1 이후로 구간안에 위치하므로 AR(1) 모형을 잠정모형으로 채택!**

# Unit 04 | Box-Jenkins 방법론

추정

- 최소제곱법, 최대가능도법, 비선형최소제곱법 등 이용

조건부 최소제곱추정의 적합결과

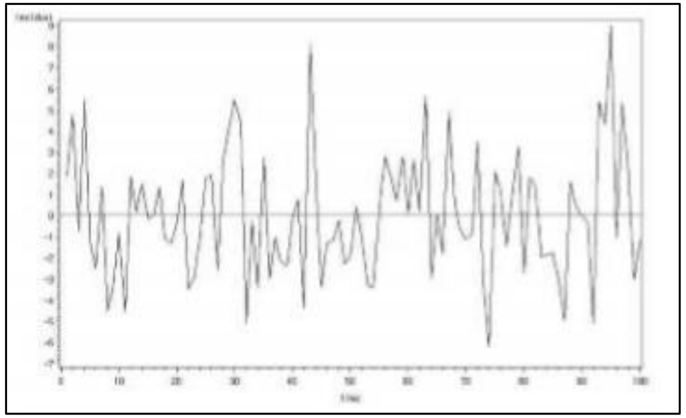
| 모수 | 추정값 | 표준오차 | $t-$값 | 유의확률 |
|---|---|---|---|---|
| $\mu$ | 19.834 | 0.179 | 110.88 | <0.0001 |
| $\phi_1$ | −0.677 | 0.074 | −9.09 | <0.0001 |

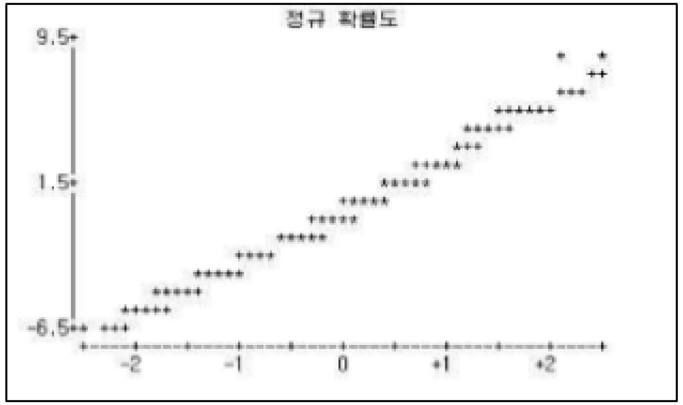
✓ **AR(1) 모형의 모수 추정!**

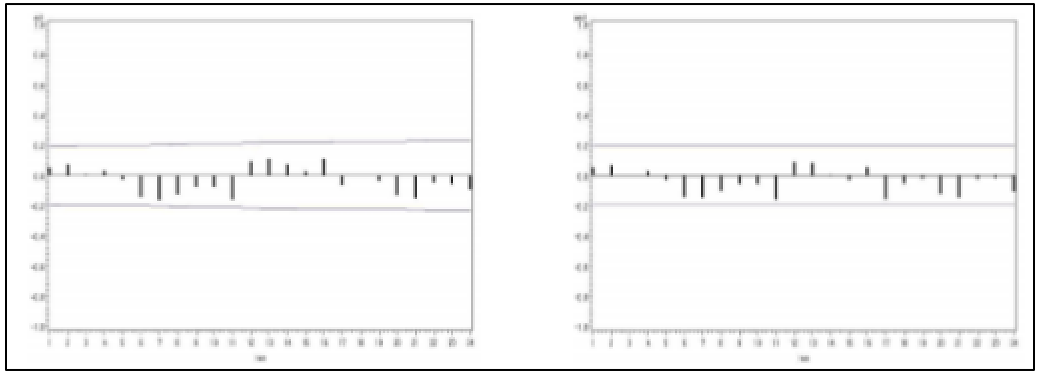
# Unit 04 | Box-Jenkins 방법론

**진단**

- 잔차의 시계열도, ACF, PACF 및 포트맨토 통계량을 이용한 잔차분석
- AIC, SBC 등의 통계값 등을 이용해 가장 설명력 높은 모형 선택





**✓ 잔차의 시계열도, 정규확률그림을 통해 정규성 확인!**
**✓ 잠정 모형의 ACF, PACF를 통해 유의한 차수 확인!**

# Unit 04 | Box-Jenkins 방법론

진단

- 잔차의 시계열도, ACF, PACF 및 포트맨토 통계량을 이용한 잔차분석
- AIC, SBC 등의 통계값 등을 이용해 가장 설명력 높은 모형 선택

**AR(1) 모형 적합 후 포트맨토 검정**

| Lag | 카이제곱 | 자유도 | 유의확률 | 자기상관계수 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 3.18 | 5 | 0.672 | 0.052 | 0.069 | 0.006 | 0.029 | -0.028 | -0.143 |
| 12 | 13.35 | 11 | 0.271 | -0.164 | -0.128 | -0.080 | -0.080 | -0.162 | 0.093 |
| 18 | 17.33 | 17 | 0.432 | 0.108 | 0.070 | 0.023 | 0.107 | -0.067 | 0.002 |
| 24 | 24.78 | 23 | 0.362 | -0.035 | -0.134 | -0.153 | -0.050 | -0.058 | -0.095 |

✓ **포트맨토 검정을 통해 모형의 적합성 여부 검정!**

# Unit 04 | Box-Jenkins 방법론

진단

- 잔차의 시계열도, ACF, PACF 및 포트맨토 통계량을 이용한 잔차분석
- AIC, SBC 등의 통계값 등을 이용해 가장 설명력 높은 모형 선택

| 모수 | 추정값 | 표준오차 | $t-$값 | 유의확률 |
|---|---|---|---|---|
| $\mu$ | 19.835 | 0.194 | 102.28 | <0.0001 |
| $\phi_1$ | −0.625 | 0.101 | −6.17 | <0.0001 |
| $\phi_2$ | 0.076 | 0.101 | 0.75 | 0.4527 |

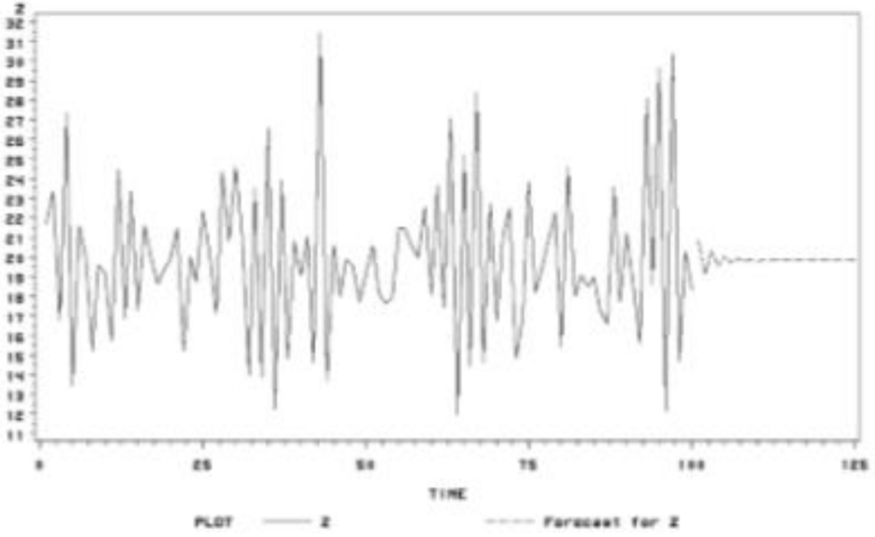
| 모수 | 추정값 | 표준오차 | $t-$값 | 유의확률 |
|---|---|---|---|---|
| $\mu$ | 19.835 | 0.191 | 103.83 | <0.0001 |
| $\theta_1$ | −0.101 | 0.148 | −0.68 | 0.5000 |
| $\phi_1$ | −0.731 | 0.102 | −7.18 | <0.0001 |

- ✓ **과대적합을 통한 모형진단 진행!**
- ✓ **AR(1)모형을 잠정모형으로 가정할 경우, MA(1)과 ARMA(1,1) 모형도 추가 진단 (모수를 하나씩 추가)**
- ✓ **AIC, BIC값을 비교하여 낮은 모형을 최종 예측모형으로 선정!**

# Unit 04 | Box-Jenkins 방법론

예측

- 적합한 최종모형으로 이후 값 예측




✓ 예측모형을 통해 미래 예측!

# [과제] 시계열 분석 실습

데이터 : 삼성 주가 예측

출처 : Yahooo finance

Box-Jenkins 방법론 과정을 따라 자유롭게 시계열분석을 진행하고 결과를 예측해주세요!

Python을 사용해 과제를 제출해주세요.

# 참고자료

- ToBig's 19기 정규세션 Time Series 강의 (김정우님)
- https://datalabbit.tistory.com/70
- https://seungbeomdo.tistory.com/26
- SAS/ETS와 R을 이용한 시계열분석 5판 (조신섭 · 손영숙 · 성병찬 지음)

# Q & A

들어주셔서 감사합니다.