시계열 분석 기법과 응용

# Week 1. 시계열 평활기법

## 1-1. 이동평균법

전치혁 교수
(포항공과대학교 산업경영공학과)

# 시계열 분석 개요

## 시계열 분석 (Time Series Analysis)

- 하나의 변수에 대한 시간에 따른 관측치를 시계열 또는 시계열 데이터라 함
- 시계열 분석 목적
  - 시계열의 특성 (추세, 계절성 등)을 요약하고 시간에 따른 패턴 (자기 상관성 등) 분석
  - 시간에 따른 패턴을 바탕으로 모형화하고 미래값을 예측
- 회귀모형과는 달리 다른 변수를 도입하지 않고 자신의 변수의 과거 패턴이 미래에도 계속된다는 가정하에 변수의 과거값을 바탕으로 미래값 예측
- 시계열 패턴은 수평, 추세, 계절성이 복합된 것으로 간주

## 본 과정에서 다룰 시계열 모형

- 평활화 모형: 이동평균, 지수평활, 윈터스 모형, 분해법
- 정상적 ARMA 모형: AR 모형, MA 모형, ARMA 모형
- 비정상적 모형: ARIMA 모형, 계절성 ARIMA 모형
- 오차 이분산 모형: ARCH 모형, GARCH 모형
- 다변량 시계열: 벡터회귀모형 (VAR)
- 상태공간모형

# 이동평균법

## 이동평균 (Moving Average)

- 매 시점에서 직전 N개 데이터의 평균을 산출하여 평활치로 사용

## 단순이동평균 (Simple Moving Average)

- 시계열 데이터 $\{X_1, X_2, \ldots\}$가 수평적 패턴인 경우 사용

## 이중 이동평균 (Double Moving Average)

- 시계열 데이터 $\{X_1, X_2, \ldots\}$가 추세 패턴을 따르는 경우 사용

# 단순 이동평균법

- 시계열 데이터 $\{X_1, X_2, \dots\}$가 수평적 패턴인 경우 사용
- 시점 t에서의 단순 이동평균: $M_t = \frac{1}{N}(X_{t-N+1} + \cdots + X_t)$
  - 시점 t+1에서의 이동평균: $M_{t+1} = \frac{1}{N}(X_{t-N+2} + \cdots + X_{t+1})$

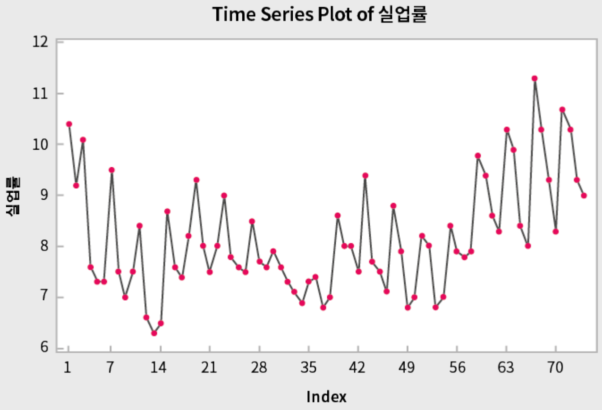
    $= M_t + \frac{X_{t+1} - X_{t-N+1}}{N}$
  - 시점 T에서 시점 T+1의 값 예측(한 단계이후 예측)

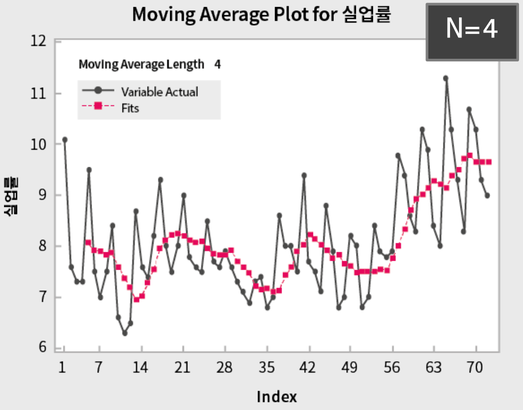
    $f_{T,1} = M_T$
  - N이 클수록 평활효과가 큼

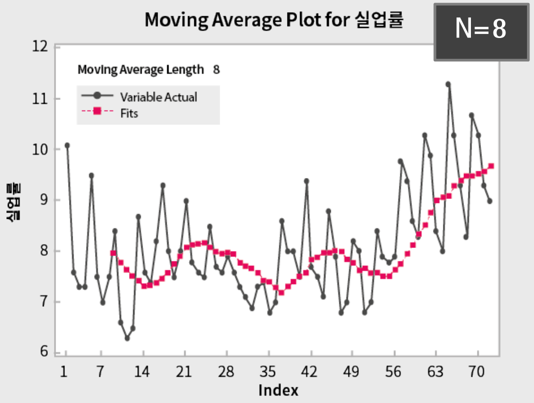
| 시간 | 시계열 | 이동평균 |
|---|---|---|
| $t-N+1$ | $X_{t-N+1}$ | |
| $t-N+2$ | $X_{t-N+2}$ | |
| … | | |
| t-1 | $X_{t-1}$ | |
| t | $X_t$ | $M_t$ |
| t+1 | $X_{t+1}$ | $M_{t+1}$ |

# 단순 이동평균법

- 예 (청년 실업률) 다음은 우리나라 분기별 (2000-2017) 청년(15-29세) 실업률(%)을 나타낸 것이다.
- N=4 와 N=8일 때 단순이동평균을 구하여 보자




#1

# 이중 이동평균법

- 시계열이 다음과 같이 선형추세를 갖는다 하자.

  $X_t = c + bt + a_t$

  – 단순이동평균 $M_t$은 추세를 늦게 따라감

  $$E[M_t] = c + bt - \frac{N-1}{2}b \quad \Rightarrow E[M_t] + \frac{N-1}{2}b = c + bt$$

  – 이를 보정하기위해 이중 이동평균을 활용

- 이중 이동평균 (double moving average)

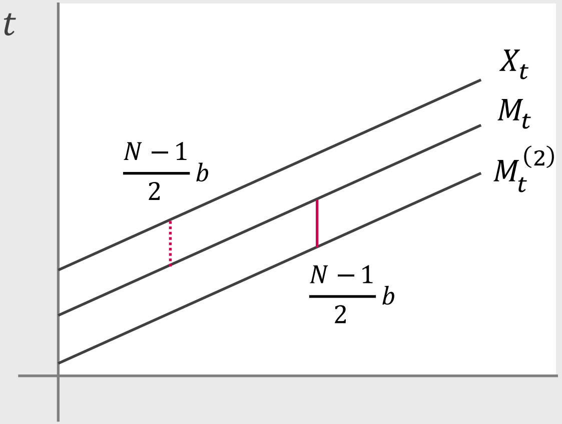
  $$M_t = \frac{1}{N}(X_{t-N+1} + \cdots + X_t)$$

  $$M_t^{(2)} = \frac{1}{N}(M_{t-N+1} + \cdots + M_t)$$

  $$E\left[M_t^{(2)}\right] = c + bt - (N-1)b$$

  $$E[M_t] - E\left[M_t^{(2)}\right] = \frac{N-1}{2}b$$

$$\hat{b} = \frac{2}{N-1}(M_T - M_t^{(2)})$$



# 이중 이동평균법

예측

- 시점 T에서 다음 시점의 예측치 (한단계 이후 예측)

$$f_{T,1} = E[X_{T+1}|X_T, X_{T-1}, \dots] = c + b(T+1)$$

$$\hat{f}_{T,1} = \hat{c} + \hat{b}(T+1) = 2M_T - M_T^{(2)} + \hat{b}$$

$$\hat{b} = \frac{2}{N-1}(M_T - M_T^{(2)})$$

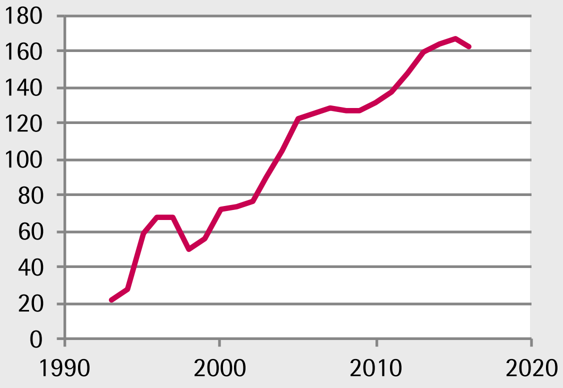
- k-단계 이후 예측치

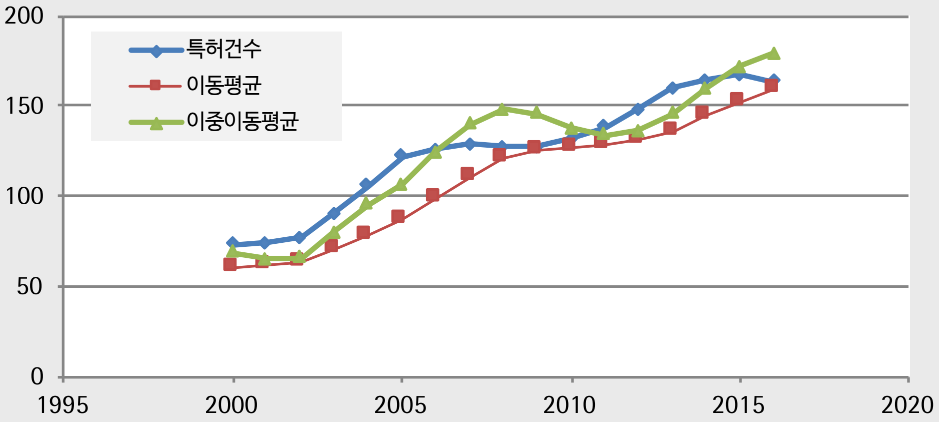
$$f_{T,k} = E[X_{T+k}|X_T, X_{T-1}, \dots] = c + b(T+k), \ \ k = 1,2,\dots$$

$$\hat{f}_{T,k} = \hat{c} + \hat{b}(T+k) = 2M_T - M_T^{(2)} + k\hat{b}$$

# 이중 이동평균법

- 예 (특허건수) 아래는 우리나라 연도별 (1993-2016) 특허건수 (천건)를 나타낸 것이다.
- 이동평균 및 이중이동평균법을 적용하여 시간에 따라 한단계 이후를 예측해 보자. (N=4 사용)



# 예측성능 척도

## 예측 오차

- 주로 특정시점에서 다음 시점을 예측하고 다음 시점의 실제값과 비교하여 예측 오차를 산출, 즉 시점 t에서 한단계 예측오차는 다음으로 산출
  - $e_{t,1} = X_{t+1} - f_{t,1}$

- 총 n개 시점에서 예측 오차를 산출하는 경우 다음과 같은 척도가 사용됨
  - 평균제곱오차: $MSE = \frac{1}{n}\sum_{t=1}^{n} e_{t,1}^2$
  - 제곱근 평균제곱오차: $RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_{t,1}^2}$
  - 평균절대오차: $MAD = \frac{1}{n}\sum_{t=1}^{n} \left|e_{t,1}\right|$
  - 평균절대 퍼센트오차: $MAPE = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{e_{t,1}}{X_{t+1}}\right|$

# 예측성능 척도

- 예 (특허건수) : 본 데이터에 대하여 단순이동평균 과 이중이동평균을 사용하여 1999년부터 한단계 이후 예측에 대한 예측오차를 구한 후 여러 예측성능척도를 산출하여 비교하여 보자.

| | 단순이동평균 사용 | 이중이동평균 사용 |
|---|---|---|
| MSE | 327.08 | 132.47 |
| RMSE | 18.08 | 11.51 |
| MAD | 15.66 | 10.11 |
| MAPE | 13.49 | 8.54 |

| 년도 | 건수 | 예측치 | 오차 | 절대오차 | 상대오차 |
|---|---|---|---|---|---|
| 2000 | 72.8 | 68.93 | 3.87 | 3.87 | 5.32 |
| 2001 | 73.7 | 64.68 | 9.02 | 9.02 | 12.24 |
| 2002 | 76.6 | 65.86 | 10.74 | 10.74 | 14.02 |
| | | | | | |
| | | | | | |
| 2015 | 167.3 | 171.64 | -4.34 | 4.34 | 9.65 |
| 2016 | 163.4 | 179.17 | -15.77 | 15.77 | 8.55 |

**이중이동평균 사용**

#2

# Reference


#1. KOSIS 국가통계포털 http://kosis.kr/ 2019.12

#2. KOSIS 국가통계포털 http://kosis.kr/ 2019.12