

# 이동평균법

---

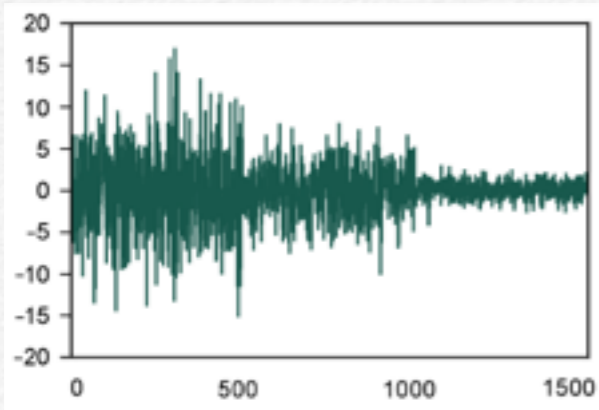
## 시간도표

### 정의

- Y-축은 시계열 관측값, X-축은 시간순서로 하여 관측값을 연결한 이차원 그래프
- 시간적 순서를 가진 시계열 데이터 의 값의 시간적 변동(변화)를 보기 위한 그래프

### 활용

- 관측값의 시간적 변화 패턴 : 시계열 성분 경향 trend, 계절성 seasonality 에 대한 시각적 진단
- 주기 폭 (분산)에 대한 시각적 진단 : 이분산 모형  $\Leftrightarrow$  정상성 stationarity

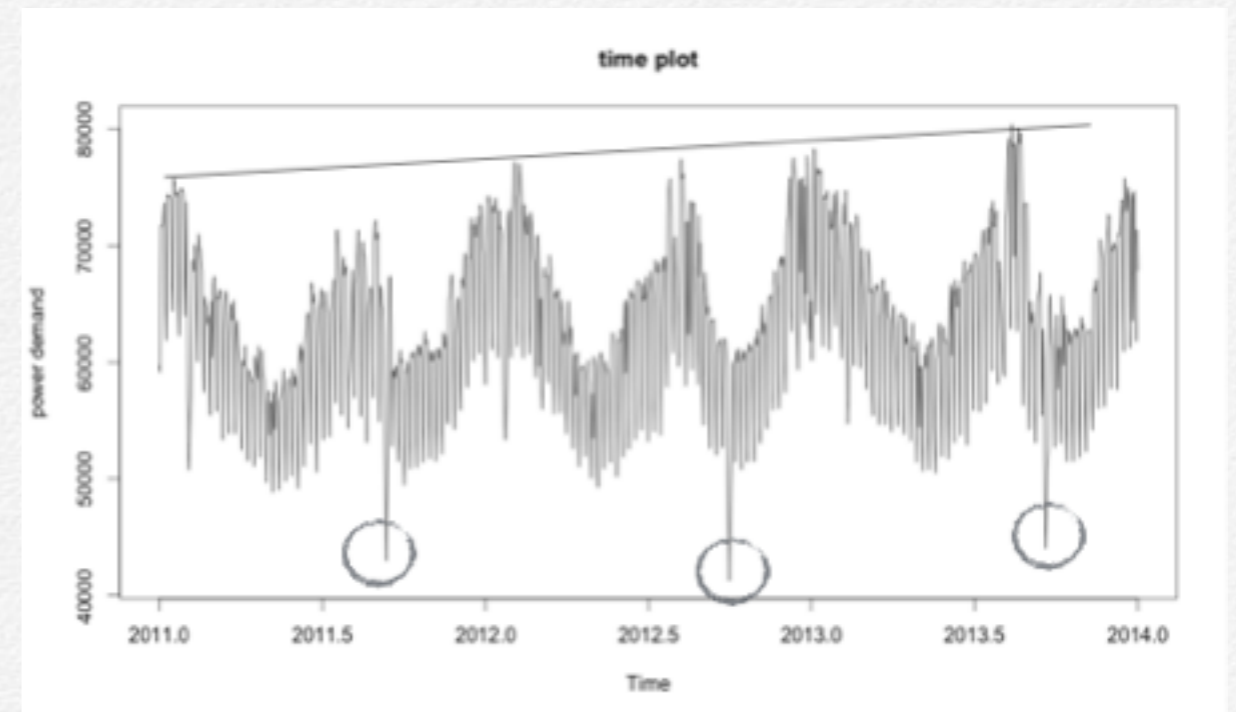


### 예제 데이터 1

[http://wolfpack.hnu.ac.kr/iBOOKs/about\\_R/power\\_demand.csv](http://wolfpack.hnu.ac.kr/iBOOKs/about_R/power_demand.csv)

2011년 1월 1일부터 13년 12월 31일, 3년간 전력수요

“PEAK\_HOUR”-최대수요시각 “REAL\_LOAD”-전력수요 “AVG\_TMPR”-평균 온도, “MIN\_TMPR”-최저기온 “MAX\_TMPR”-최고기온 “MAX\_WND\_SPD”-최대풍속, “AVG\_WND\_SPD”-평균풍속 “RAIN\_AMT”-강수량 “AVG\_HMDT”-평균습도 “CLD\_AMT”-구름량 “SUN\_HOUR”-일조량 “HOLIDAY\_YN”-휴일여부



- 일별 자료이므로 당연히 7일(계절성) 주기 있음-전력수요이므로 월별주기(계절성) 있음
- 추석연휴 - 동그라, 전력수요가 이상적으로 낮음
- 직선적 증가 경향을 (직선) 보이고 있음 - 전력수요 연도별 증가 경향 보임

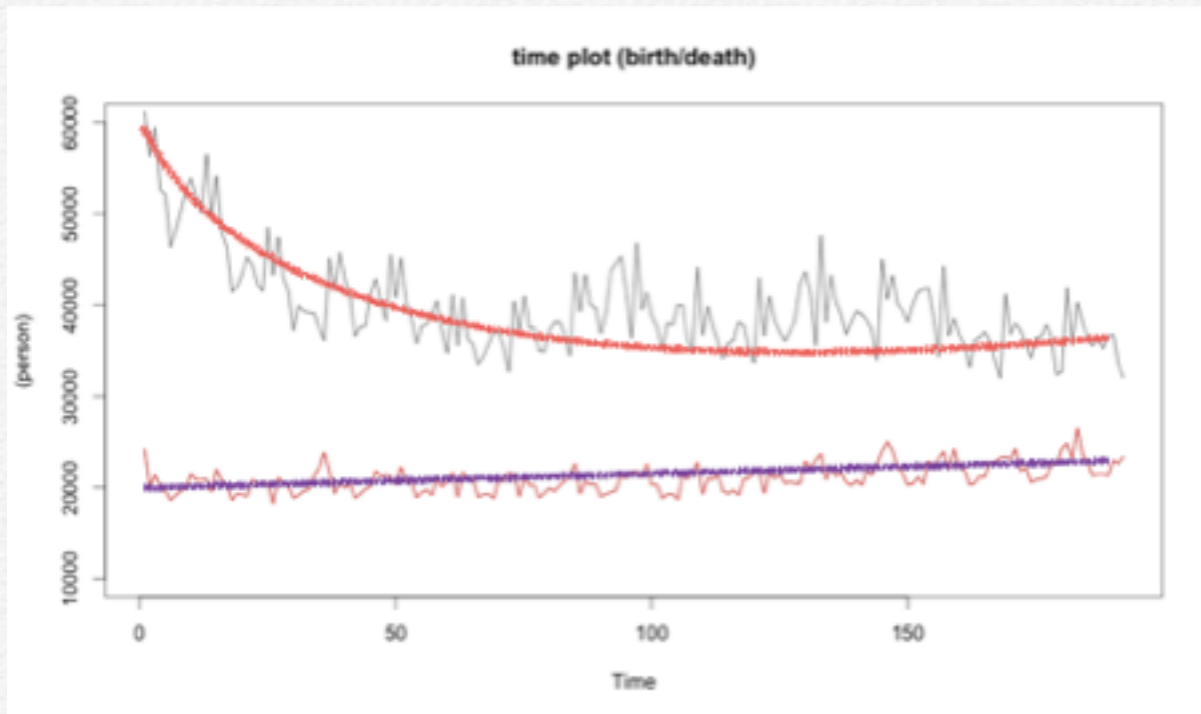
## 예제 데이터 2

[http://wolfpack.hnu.ac.kr/iBOOKs/about\\_R/population\\_vital.csv](http://wolfpack.hnu.ac.kr/iBOOKs/about_R/population_vital.csv)

국내 인구동태 월별 (2000년 1월 ~ 2015년 12월)

출생(명) birth / 사망(명) death / 결혼(건) marriage

이혼(건) divorce / 경제활동인구(명)



- 지수적 감소하는 경향을 보임 (출생자 수)
- 사망자 수는 직선적 증가
- 월별 자료이므로 12주기 계절성 있음

## Moving Average 이동평균법

시계열 데이터는 주기나 불규칙성을 가지고 있으므로 과거의 몇 개 관측치를 평균하여 전반적인 추세를 파악할 수 있는 방법을 이용하여 예측치를 구할 수 있는데 이를 이동평균법(Moving average)이라 한다.

이동 평균법은 과거 몇 개의 과거치의 평균으로 미래 값을 예측하는 방법이다. 지수평활법과는 달리 이동평균법에서는 과거치에 적용되는 가중치는 동일

### 모형

$$Y_t = \mu + e_t, e_t \sim N(0, \sigma^2)$$

관측치 -  $\{Y_t; t = 1, 2, \dots, T\}$ , 평균모수 -  $\mu$

### MA 계산

$$MA_t = \frac{\sum_{j=1}^m Y_{t-j}}{m}$$

### 중심점 이동평균 Centered MA

계절성이 존재하는 경우 사용,  $m = 2k + 1$ ,  $m$ 이 짝수이면 +1 사용하지 않음

$$\text{분기별 데이터} : MA_t = \frac{1}{m} \sum_{j=k}^k Y_{t+j}$$

### 예측

$$\hat{Y}_{t+1} = MA_t : \text{차기 1만 예측가능}$$

## 주기 M 결정

- 주기의 배수를 활용
- 주가의 경우 : M=5, 20(한달), 60(분기), 120(반년)
- 월별데이터 : M=12, 24, 36, ...

## 특징

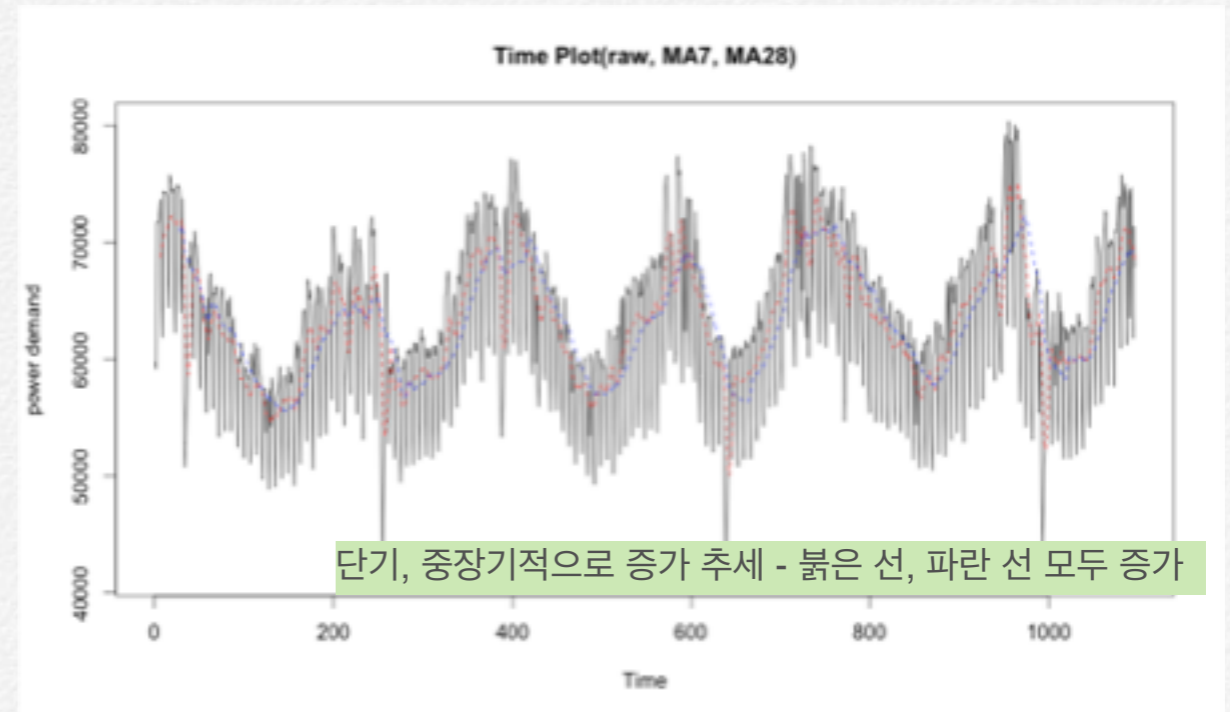
- 계절성, 불규칙성을 제거하여 전반적인 추세 파악 가능 - 직관적 정보 제공
- 주기가 길면 장기 패턴, 짧으면 단기 패턴을 진단할 수 있음 - 주기가 길어질수록 주기(사이클)는 사라지고 직선 형태가 된다.
- 자신의 과거치 (pattern, trace)이용하여 미래 값을 예측한다.
- 자신의 m 개 관측치 평균으로 시계열 자료 {Yt}의 패턴 인식
- 가중치는 1/m으로 동일하다.

## 이동평균법 문제점

- 과거치에 대한 가중치가 동일하다. - 시간이 먼 관측값과 가까운 값의 영향을 동일하게 간주하므로 왜곡 가능성이 높음
- 시계열 데이터에 경향과 계절성이 없다면 문제가 없으나, 존재한다면
- 차기 1기만 예측이 가능함 - 이전 추세분석

## In R

- 전력수요 : 일별 - 주기 m=7, 28
- 출생율 : 월별 - 주기 m=12, 60

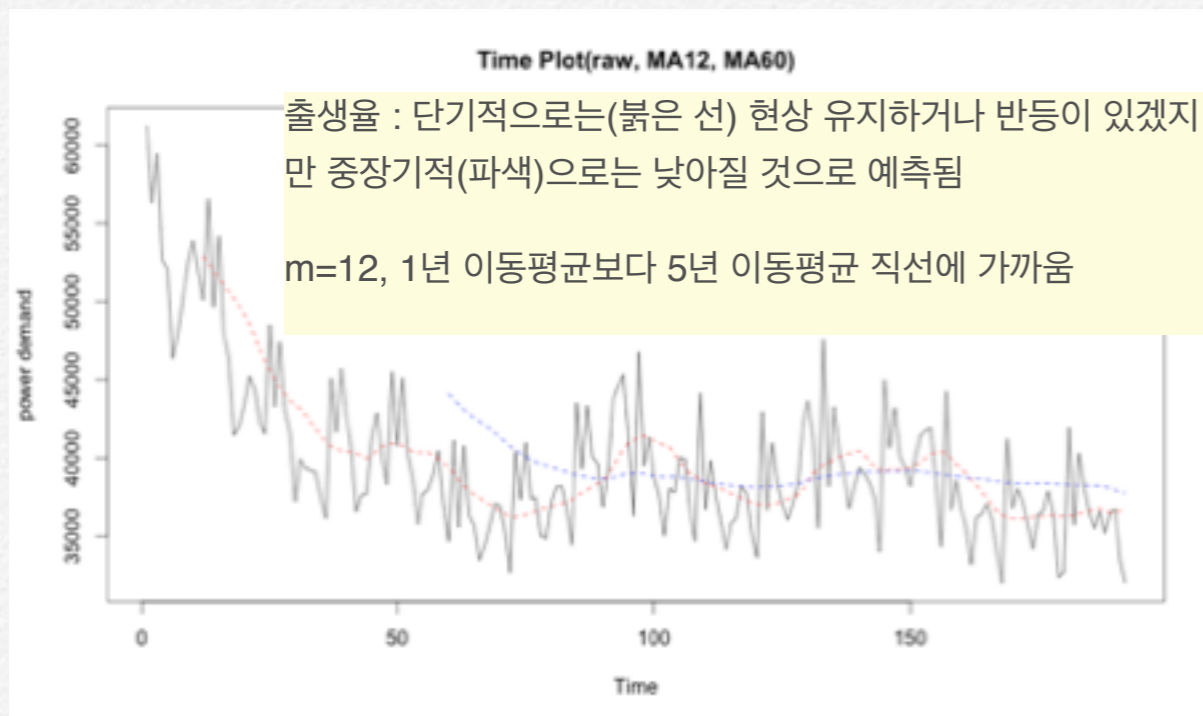


```
library(TTR) #Smoothing 함수 불러오기
fit.yt=SMA(yt.pd,n=7) #주기(m)=7 이동평균법
fit2.yt=SMA(yt.pd,n=28) #주기(m)=28 이동평균법
fit3.yt=filter(yt.pd,rep(1/7,7),sides=1) #가중치1/7, 7개관측치, 1=이동평균,
2=중심이동평균
ds.pd2=data.frame(cbind(yt,fit.yt,fit2.yt))
plot.ts(ds.pd2$yt,main="Time Plot(raw, MA7, MA28)",ylab="power
demand",col="black") #원데이터
lines(ds.pd2$fit.yt,main="Time Plot(raw, MA7, MA28)",ylab="power
demand",col="red",lty="dashed") # m=7 MA
lines(ds.pd2$fit2.yt,main="Time Plot(raw, MA7, MA28)",ylab="power
demand",col="blue",lty="dashed") # m=28 MA
```

```

library(TTR) #Smoothing 함수 불러오기
fit.yt.bt=SMA(yt.bt,n=12) #주기(m)=12 이동평균법
fit2.yt.bt=SMA(yt.bt,n=60) #주기(m)=60 이동평균법
ds.pv2=data.frame(cbind(yt.bt,fit.yt.bt,fit2.yt.bt))
plot.ts(ds.pv2$yt.bt,main="Time Plot(raw, MA12, MA60)",ylab="power
demand",col="black") #원데이터
lines(ds.pv2$fit.yt.bt,main="Time Plot(raw, MA12, MA60)",ylab="power
demand",col="red",lty="dashed") # m=12 MA
lines(ds.pv2$fit2.yt.bt,main="Time Plot(raw, MA12,
MA60)",ylab="power demand",col="blue",lty="dashed") # m=60 MA

```



## 함께 그리기

```

ds.pv=read.csv("population_vital.csv") #csv 외부데이터 R 데이터 읽기
ds.pv0=ds.pv[,-1]
yt.pv=ts(ds.pv0,frequency=12,start=c(2000,1)) #12일 주기,2000년 1월
plot.ts(yt.pv,main="time plot",ylab="(person/item)") #제목, y-축제목,

```

