



# 시계열 분석:

## 시계열 데이터의 특징, 시계열 분해법, 자기상관

# Key words

#시계열데이터 #이동평균법  
#지수평활법 #시계열분해법 #정상성  
#자기상관 #자기회귀모델(AR)

# 시계열 분석

## 시계열 자료의 개념

- 시간의 흐름에 따라 순차적으로 관찰된 자료들의 집합.

## 시계열 자료의 예시

- 월별 소비자 물가지수, 일별 주가지수.
- 일일 강수량, 일일 기온, 연간 지진의 발생 수.
- 월별 상품판매량, 월별 상품재고량, 일별 상품매출액.
- 월별 교통사고 건수, 월별 범죄 발생 수.



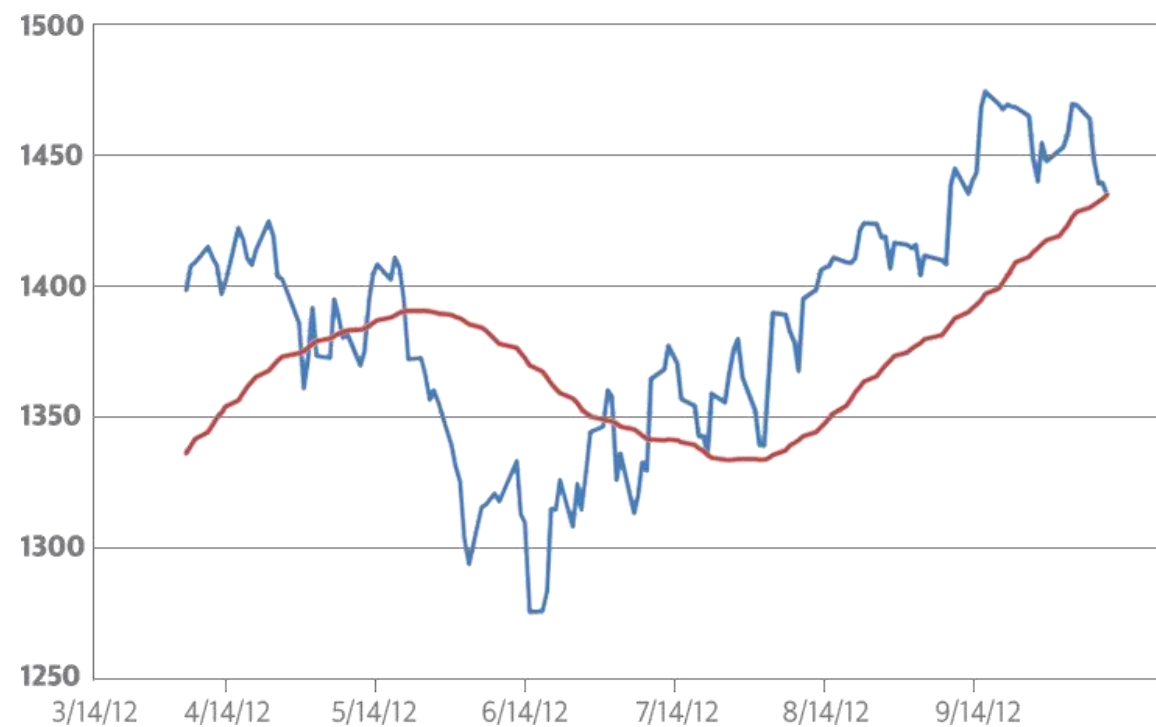


# 탐색적 시계열 분석

## I 이동 평균 평활법(Moving Average Smoothing)

- 가장 최근의  $m$ -기간 동안의 자료들의 단순 평균을 다음 기간의 예측값으로 추정.
- $Z_n$  을 시점  $n$ 에서의 실제값,  $\hat{Z}_{n+1}$  을 시점  $n$ 에서 추정한 시점  $n + 1$ 의 예측값이라고 하면,

$$\hat{Z}_{n+1} = \frac{1}{m} (Z_n + Z_{n-1} + \dots + Z_{n-m+1})$$



# 탐색적 시계열 분석

## 예제

- 다음은 어느 회사의 지난 24년간의 판매액을 나타낸 자료이다.  
 $m = 5$  인 단순이동평균을 이용하여 1992년 이후의 값들을 예측하시오.

연도	$t$	판매액( $z_t$ )	예측값( $\hat{z}_{n+1}$ )
1987	1	312	
1988	2	381	
1989	3	317	
1990	4	337	
1991	5	399	
1992	6	387	349.2
1993	7	375	364.2
1994	8	349	363
...	...	...	...
2009	23	454	388.2
2010	24	365	409.4

# 탐색적 시계열 분석

## I 지수평활법(Exponential Smoothing)

- 이동평균평활법과 같이 과거의 관측값을 활용하여 다음 기간의 값을 예측하지만, 최근의 자료에 더 많은 가중치를 부여하는 방법임.

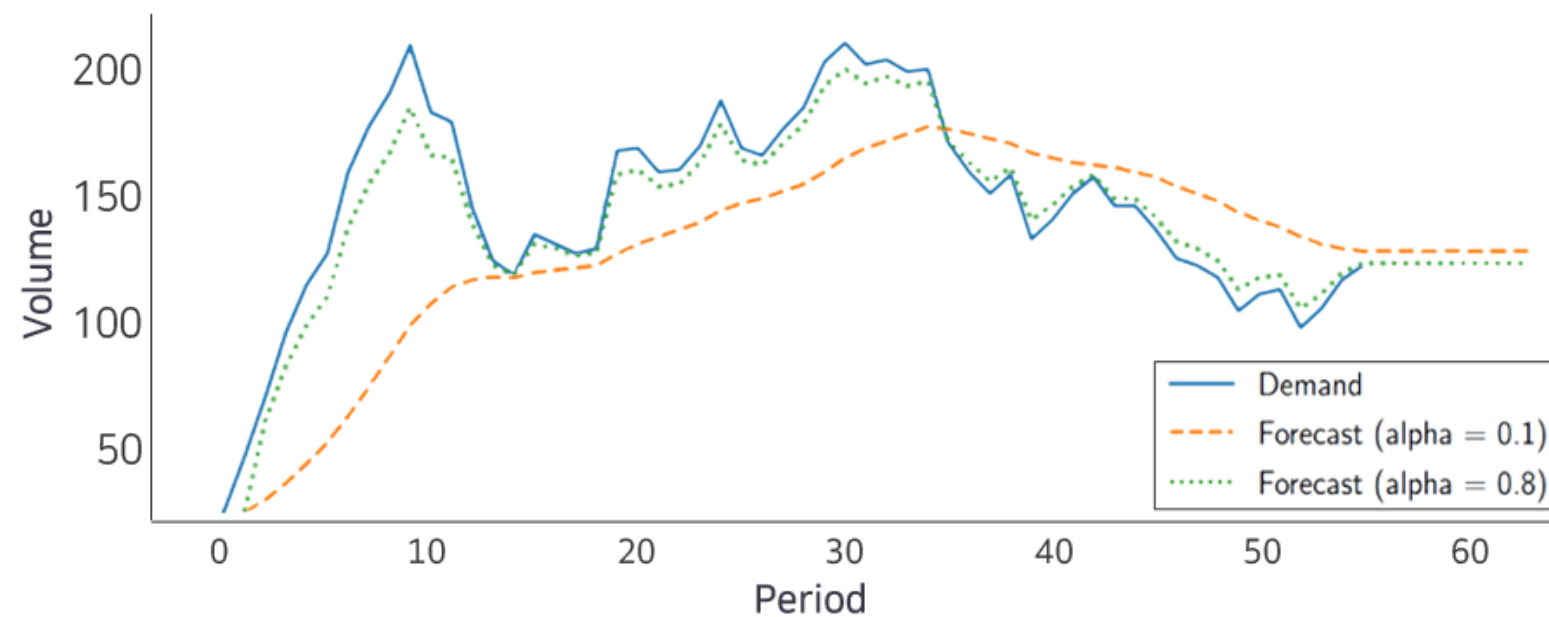
# 탐색적 시계열 분석

## 지수평활법(Exponential Smoothing)

- $Z_n$  을 시점  $n$ 에서의 실제값,  $S_n$  을 시점  $n$ 에서 지수평활 예측값( $\hat{Z}_{n+1}$ )이라고 하고, 평활상수  $\alpha$ 는 0~1 사이의 값이라고 할 때,

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

$$\Leftrightarrow S_n = \alpha Z_n + \alpha(1 - \alpha)Z_{n-1} + \alpha(1 - \alpha)^2 Z_{n-2} + \dots$$



# 탐색적 시계열 분석

## 예제

- $\alpha = 0.2$  인 지수평활값을 구하여라.

연도	$t$	판매액( $z_t$ )	예측값( $s_t$ )
2010	1	39	39
2011	2	37	38.6
2012	3	61	43.1
2013	4	58	46.1
2014	5	18	40.5
2015	6	56	43.6
2016	7	82	51.2
2017	8	27	46.4
2018	9	41	45.3
2019	10	69	50.1
2020	11	49	49.8



# 탐색적 시계열 분석

## I 분해법 (추세요인, 계절요인)

- 변동 요인의 분해 과정

- 원계열  $Z_t$ 에 추세선(예. 단순선형회귀)을 적합하여 추세성  $\hat{T}_t$ 을 추정.

$$Z_t \rightarrow \hat{T}_t = \hat{\alpha} + \hat{\beta}t$$

- 원계열  $Z_t$ 과 추세성  $\hat{T}_t$ 의 비율  $\hat{S}_t$ 을 정의

$$\hat{S}_t = \frac{Z_t}{\hat{T}_t}$$

- $k$  번째 계절에 대한 계절지수  $\hat{S}_k$ 는 해당 계절의  $\hat{S}_t$ 들의 평균으로 구함.

# 탐색적 시계열 분석

## I 분해법 (추세요인, 계절요인)

- 예측

- $T$  시점의 예측값  $\hat{Z}_T$ 는

- $\hat{Z}_T = (\hat{\alpha} + \hat{\beta} T) \hat{S}_{I_K}$  ( $\hat{S}_{I_K}$ :  $T$ 시점이  $K$ 번째 계절이라고 할 때의 계절지수)

# 탐색적 시계열 분석

## 예제

- 다음은 어느 호텔의 분기별 Occupancy Rate를 기록한 것이다. 추세효과와 계절효과를 추정한 뒤, 이를 토대로 2018년도의 각 분기별 예측값을 구하여라.

연도	분기	$t$	$Z_t$	$T_t$	$S_t$
2013	1	1	0.561	0.645	0.870
	2	2	0.702	0.650	1.080
	3	3	0.800	0.655	1.221
	4	4	0.568	0.660	0.860
2014	1	5	0.575	0.666	0.864
	2	6	0.738	0.671	1.100
	3	7	0.868	0.676	1.284
	4	8	0.605	0.681	0.888
2015	1	9	0.594	0.687	0.865
	2	10	0.738	0.692	1.067
	3	11	0.729	0.697	1.046
	4	12	0.600	0.702	0.854
2016	1	13	0.622	0.708	0.879
	2	14	0.708	0.713	0.993
	3	15	0.806	0.718	1.122
	4	16	0.632	0.723	0.874
2017	1	17	0.665	0.729	0.913
	2	18	0.835	0.734	1.138
	3	19	0.873	0.739	1.181
	4	20	0.670	0.744	0.900

# 탐색적 시계열 분석

- 추세효과 추정

$$\hat{T}_t = 0.6394 + 0.00525 t$$

- 계절지수 추정

연도	1	2	3	4
2013	0.870	1.080	1.221	0.860
2014	0.864	1.100	1.284	0.888
2015	0.865	1.067	1.046	0.854
2016	0.879	0.993	1.122	0.874
2017	0.913	1.138	1.181	0.900
$\hat{S}I_k$	0.878	1.076	1.171	0.875

# 탐색적 시계열 분석

분기	$t$	$\hat{T}_t (= 0.6394 + 0.00525 t)$ 추세 예측값	$\hat{S}I_k$ 계절지수 예측값	$\hat{Z}_t$ Occupancy Rate 예측값
1	21	0.749	0.878	0.658
2	22	0.755	1.076	0.812
3	23	0.760	1.171	0.890
4	24	0.765	0.875	0.669

# 확률적 시계열 분석

## I 정상 시계열 모형

- 정상성(stationarity)의 조건
  - 평균이 시점에 의존하지 않음.
  - 분산이 시점에 의존하지 않음.
  - 공분산은 단지 시차에만 의존하고, 시점 자체에는 의존하지 않음.



# 확률적 시계열 분석

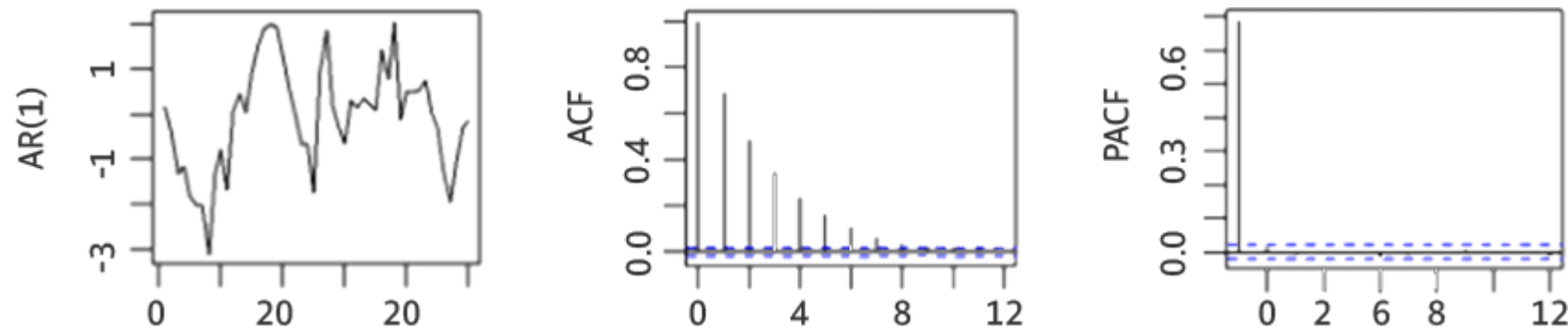
## I 정상 시계열 모형

- 정상성 조건으로 하나라도 만족하지 못하는 시계열 자료를 비정상 시계열이라고 부르며, 비정상 시계열은 정상성을 만족하는 정상시계열 자료로 만든 후 자기회귀(AR), 이동평균(MA), 자기회귀이동평균(ARMA) 모형을 적합하는 시계열 분석을 수행함.

# 확률적 시계열 분석

## 정상 시계열 모형

- 자기회귀(AR), 이동평균(MA) 모형의 식별을 위해 자기상관 및 부분자기상관함수의 형태를 이용함.
  - 자기상관계수 (autocorrelation, AC)  
일정한 간격으로 떨어진 두 시점에서의 시계열 자료 간의 선형적 연관성.
  - 부분자기상관계수 (partial autocorrelation, PAC)  
일정한 간격으로 떨어진 두 시점 사이의 상관관계 분석 시, 중간 시점의 영향을 제외하고 측정하는 선형적 연관성.



# 확률적 시계열 분석

## I 자기회귀과정(Autoregressive Model)

- AR(1) 모델

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon$$

- AR(1) 모델 적용사례

연도	CPI 변화율 ( $Y_t$ )
1979	11.3
1980	13.5
1981	10.4
1982	6.2
...	...
2004	2.7
2005	3.4
2006	3.2

# 확률적 시계열 분석

## I 자기회귀과정(Autoregressive Model)

- AR(1) 모델 적합결과

	계수추정치	표준오차	t 통계량	유의확률
절편	0.007	0.0044	1.58	0.12681
Y	0.761	0.0867	8.78	4.12E-09

$$\hat{Y}_t = 0.007 + 0.761Y_{t-1}$$

- AR(1) 모델 예측 : 2007년의 CPI 변화율?

$$\hat{Y}_{2007} = 0.007 + 0.761(3.2) = 2.44$$

# 시계열 분석

## ■ 시계열 자료

- 시간의 흐름에 따라 순차적으로 관찰된 자료

## ■ 탐색적 시계열 분석

- 평활법

- 이동 평균 평활법(Moving Average Smoothing)

- 최근  $m$ 기간의 자료 이용.
    - 단순평균값.

- 지수평활법(Exponential Smoothing)

- 과거 전체 자료 이용.
    - 최근 자료에 높은 가중치를 두며 과거로 갈수록 가중치는 지수적으로 감소함.

- 분해법

- 추세요인 : 시계열 자료가 갖는 장기적인 변화 추세. 단순선형회귀모델 등으로 추정.
  - 계절요인 : 분기, 계절, 월 등 일정한 주기에 따라 순환하며 변동하는 요인.

# 시계열 분석

## I 확률적 시계열 분석

- 정상성 (stationarity)
  - 평균이 일정함.
  - 분산이 시점에 의존하지 않음.
  - 공분산은 단지 시차에만 의존하고, 시점 자체에는 의존하지 않음.
- 확률적 시계열 모델
  - 자기회귀(AR), 이동평균(MA), 자기회귀이동평균(ARMA) 등.
  - 모델의 식별을 위해 자기상관계수, 부분자기상관계수를 이용함.
    - 자기상관계수 : 일정 간격 떨어진 두 시점에서의 자료 간의 선형적 연관성.
  - AR(1) 모델
    - 현재의 값  $Y_t$ 을 직전 시점의 값  $Y_{t-1}$ 으로 설명.
    - $Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon$