

시계열 분석

목 차

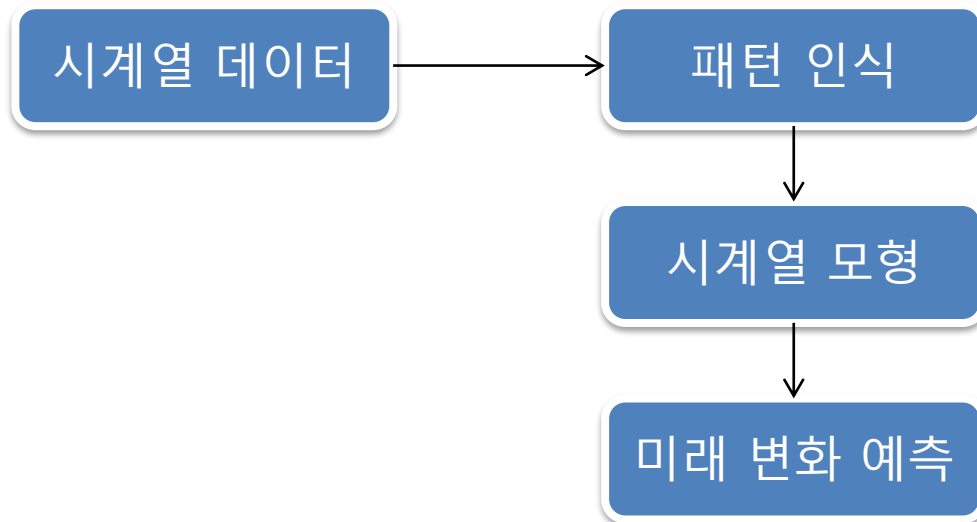
Timeseries Analysis 수업내용

1. 시계열 분석
2. 시계열 자료 분석
3. 시계열 자료 시각화
4. 시계열 분석 기법(ARIMA 모형 기반)
5. ARIMA 모형 시계열 예측

시계열 분석

● 시계열 분석(Timeseries Analysis)

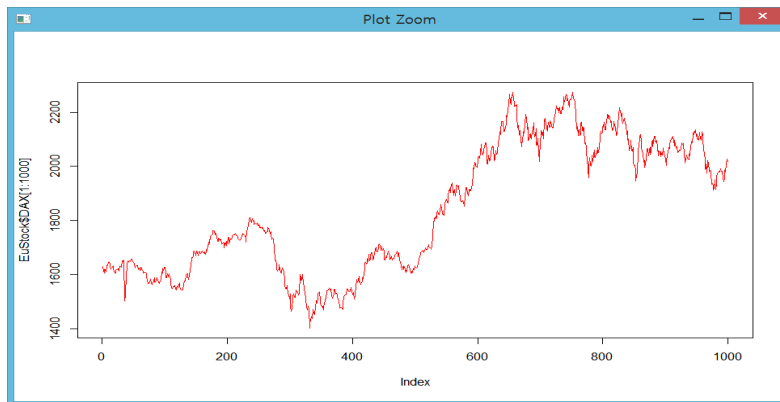
- 어떤 현상에 대해서 시간의 변화에 따라 일정한 간격으로 현상의 변화를 기록한 시계열 데이터를 대상으로 미래의 변화에 대한 추세를 분석하는 방법
- 시간 경과에 따른 관측 값의 변화를 패턴으로 인식하여 시계열 모형을 추정하고, 이 모형을 통해서 미래의 변화에 대한 추세를 예측하는 분석방법



시계열 분석

● 시계열 분석 특징

1. y변수 존재 : 시간 t 를 설명변수(x)로 시계열(Y_t)을 반응변수(y)로 사용.
2. 미래 추정 : 과거와 현재의 현상을 파악하고 이를 통해서 미래 추정.
3. 계절성 자료 : 시간 축을 기준으로 계절성이 있는 자료를 데이터 셋으로 이용.
4. 모수 검정 : 선형성, 정규성, 등분산성 가정이 만족해야 함.
5. 추론 기능 : 유의수준 판단 기준이 존재하는 추론통계 방식.
6. 활용분야 : 경기예측, 판매예측, 주식시장분석, 예산 및 투자 분석, 교통수요 등



x : 시간축
y : 통계량

시계열 분석

● 시계열 분석 적용 범위

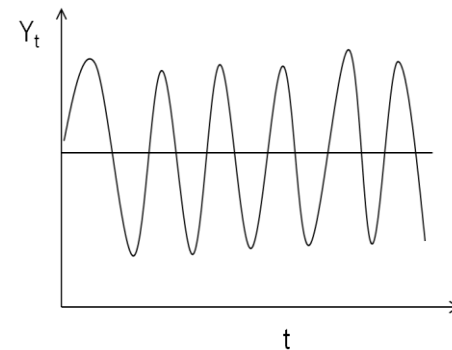
1. 기존 사실에 대한 결과 규명 : 주별, 월별, 분기별, 년도별 분석을 통해서 고객의 구매 패턴을 분석
2. 시계열자료 특성 규명 : 시계열에 영향을 주는 일반적인 요소(추세, 계절, 순환, 불규칙)를 분해해서 분석한다.(시계열 요소 분해법)
3. 가까운 미래에 대한 시나리오 규명 : 탄소배출 억제를 성공 했을 때와 실패 했을 때 지구 온난화는 얼마나 심각해질 것인가를 분석한다.
4. 변수와 변수의 관계 규명 : 경기선행지수와 종합주가지수의 관계를 분석한다.(국가 경제와 주가지수 관계)
5. 변수 제어 결과 규명 : 입력 변수의 제어(조작)를 통해서 미래의 예측 결과를 통제할 수 있다.(판매 촉진에 영향을 주는 변수 값을 조작할 경우 판매에 어떠한 영향을 미치는가?)

시계열 자료 분석

● 시계열 자료 분류

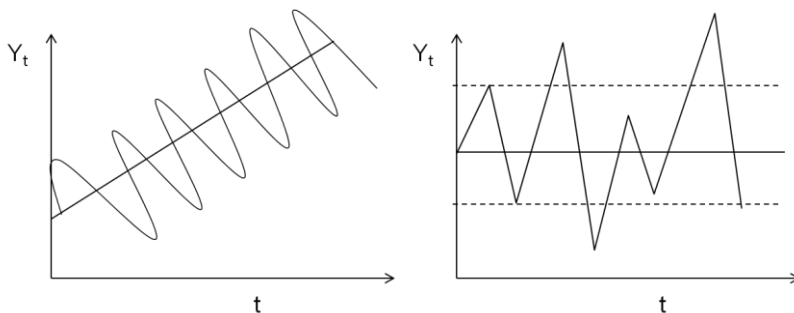
1. 정상성(stationary) 시계열

- 1) 어떤 시계열 자료의 변화 패턴이 평균값을 중심으로 일정한 변동폭을 갖는 시계열
- 2) 시간의 추이와 관계없이 평균과 분산이 일정



2. 비정상성(non-stationary) 시계열 : 대부분의 시계열자료

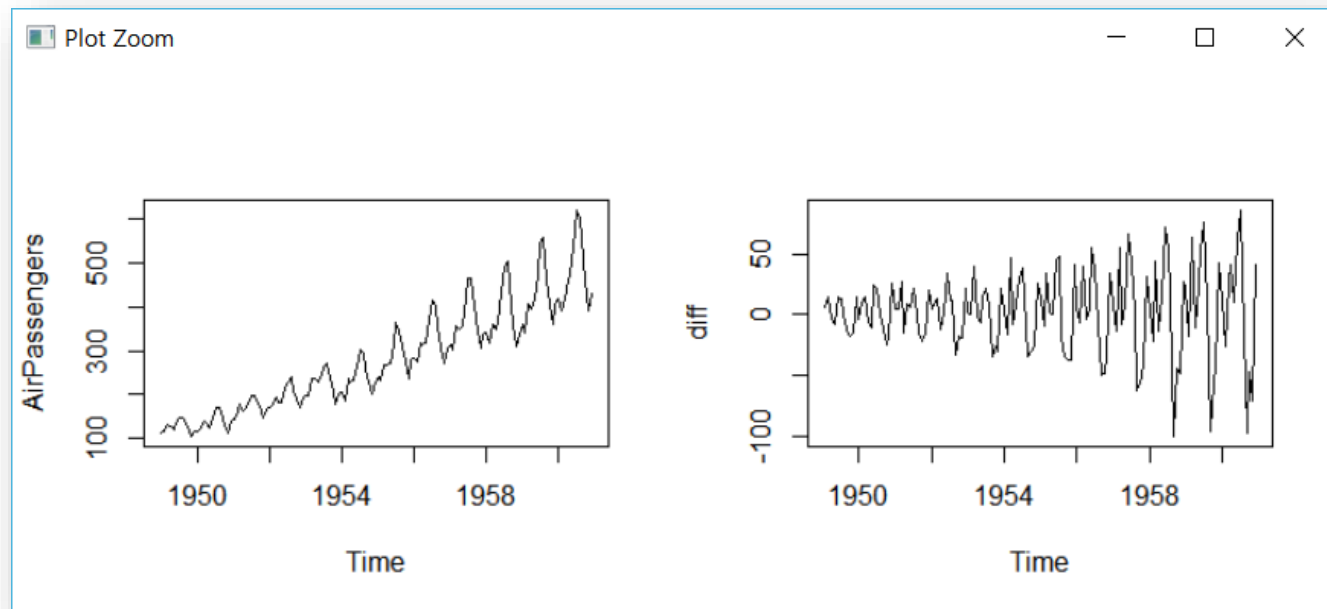
- 1) 시간의 추이에 따라서 점진적으로 증가하는 추세
- 2) 분산이 일정하지 않은 경우



시계열 자료 분석

- 시계열 자료 확인

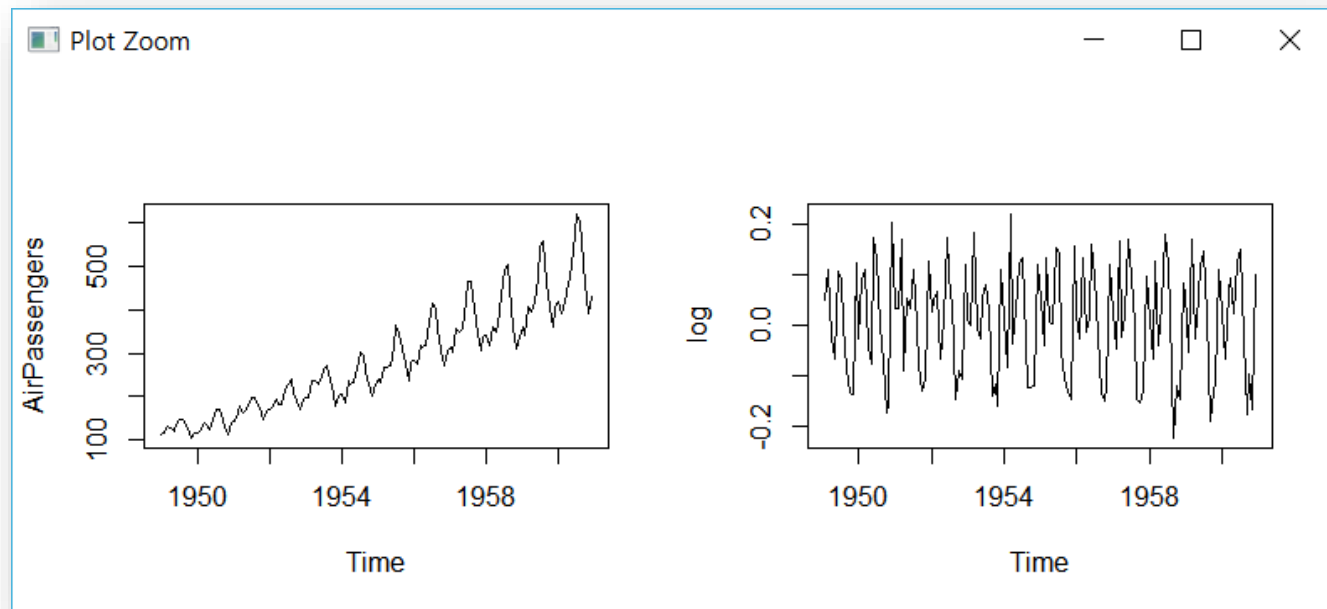
1. 비정상 시계열 -> 정상성 시계열 : 평균 정상화 : 차분



시계열 자료 분석

- 시계열 자료 확인

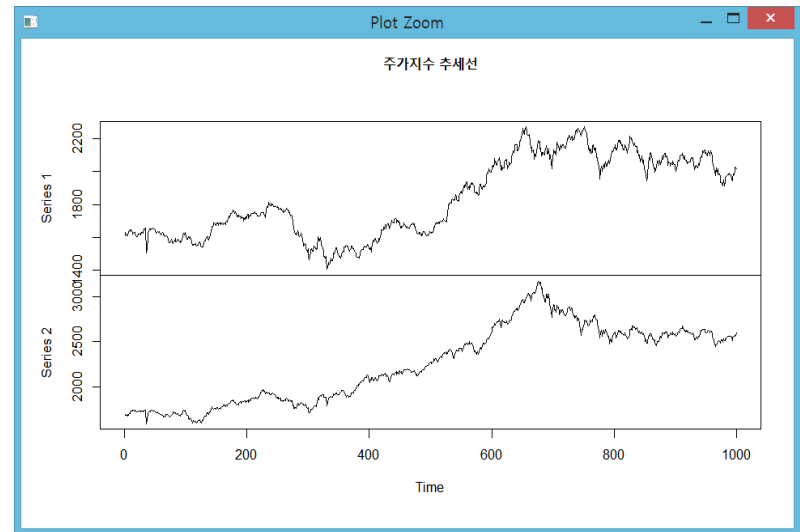
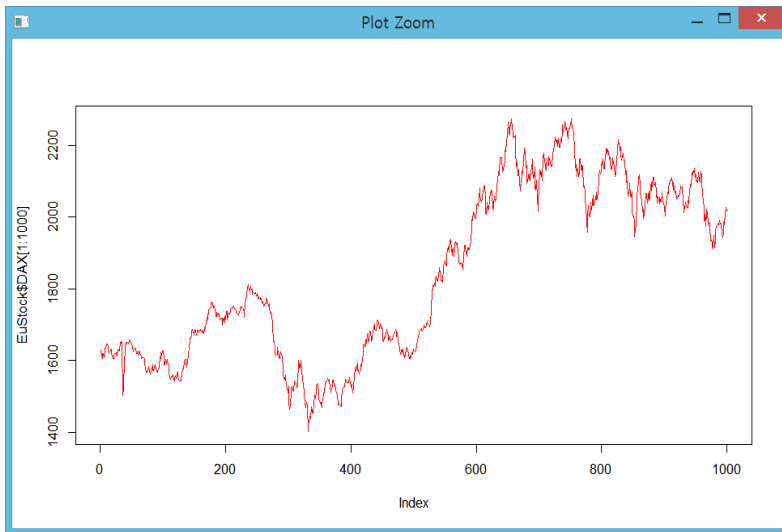
2. 비정상 시계열 -> 정상성 시계열 : 분산 정상화 : 로그 -> 차분



시계열 자료 시각화

- 시계열 자료 시각화

- ✓ 단일 시계열 vs 다중 시계열 시각화



시계열 요소 분해 시각화

● 시계열 데이터 특성

1. 추세 변동(Trend variation: T) :

✓ 인구 변동, 지각변동, 기술변화 등 상승과 하락의 영향(장기 변동요인)

2. 순환 변동(Cyclical variation: C)

✓ 2년~10년의 주기에서 일정한 기간 없는 반복적 요소(중. 장기 변동요인)

3. 계절 변동(Seasonal variation: S)

✓ 일정한 기간(월, 요일, 분기), 1년 단위 반복적 요소(단기 변동요인)

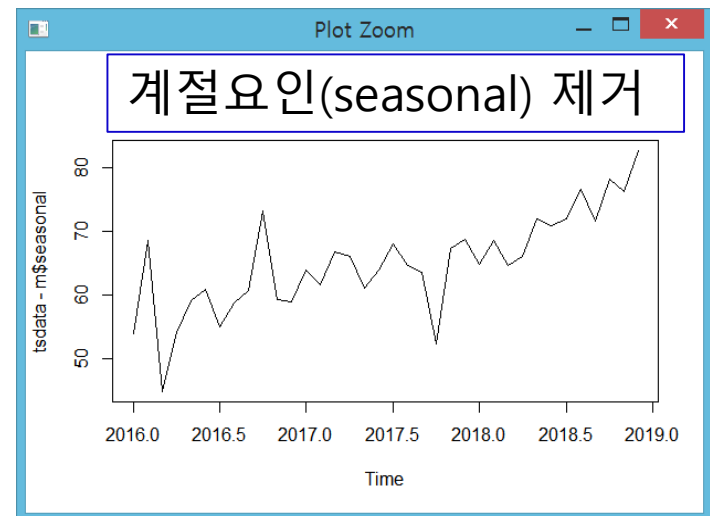
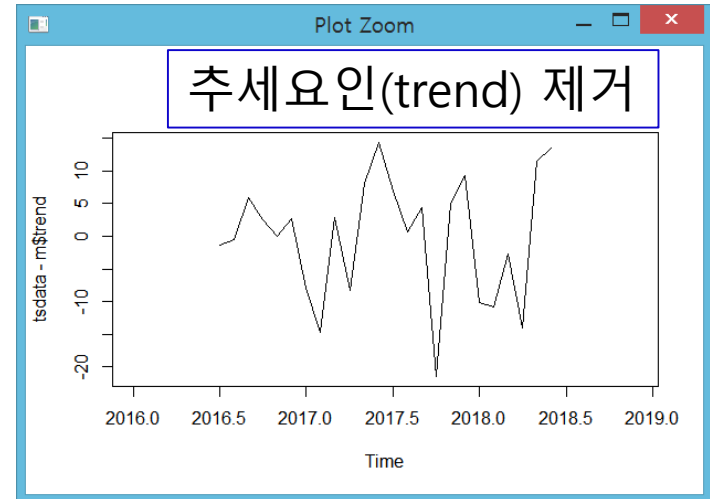
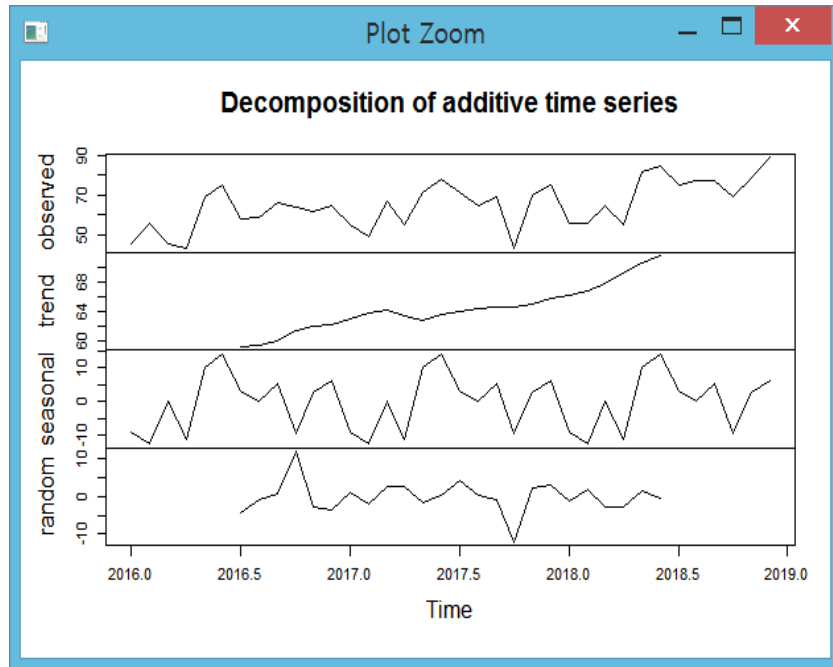
4. 불규칙변동(Irregular variation: I)

✓ 어떤 규칙 없이 예측 불가능한 변동요인(설명할 수 없는 요인)

✓ 실제 시계열 자료에서 추세, 순환, 계절 요인을 뺀 결과(회귀분석 오차)

시계열 요소 분해 시각화

- 시계열 데이터 특성 분석



ARIMA 모형 시계열

● 시계열 모형 생성

- 시계열 모형 생성의 대표적인 방법(현재 가장 많이 이용)

1. 정상성을 가진 시계열 모형

✓ 자기회귀모형(AR), 이동평균모형(MA),

자기회귀이동평균모형(ARMA)

2. 비정상성을 가진 시계열 모형(차수 적용)

✓ 자기회귀누적이동평균모형(ARIMA)

✓ 형식) $ARIMA(p, d, q)$: 3개의 인수

✓ p : AR모형 차수, d : 차분 차수, q : MA모형 차수

시계열 분석

- 시계열 분석 절차

- ARIMA 모델을 이용

[단계1] 시계열자료 특성분석(정상성/비정상성)

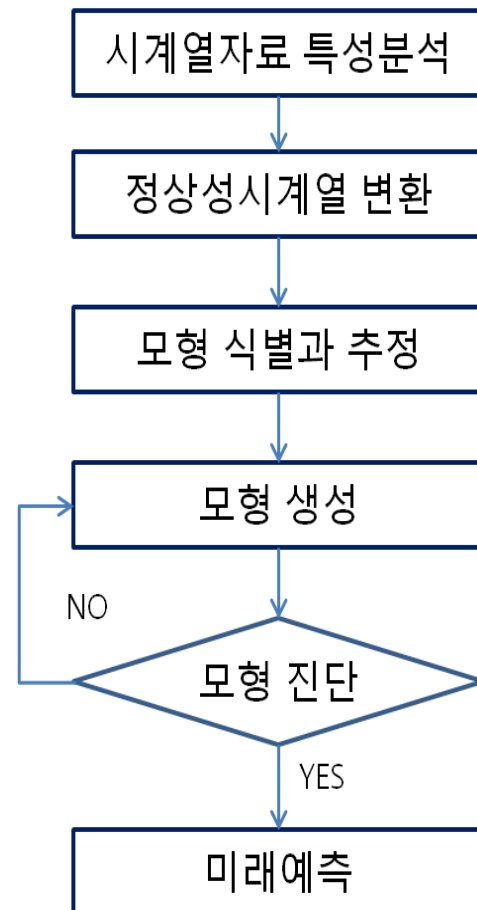
[단계2] 정상성시계열 변환

[단계3] 모형 식별과 추정

[단계4] 모형 생성

[단계5] 모형 진단(모형 타당성 검정)

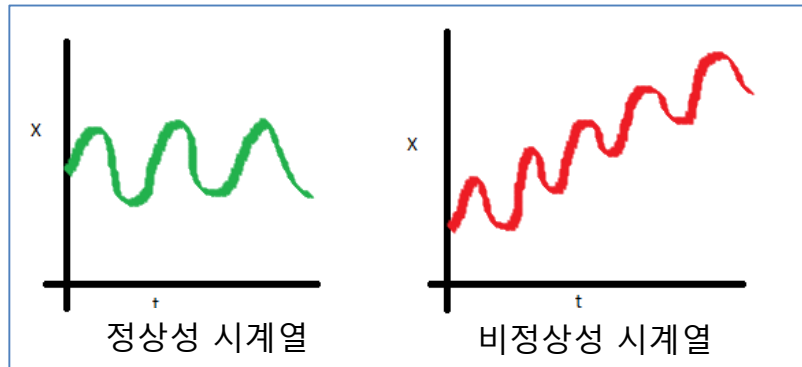
[단계6] 미래 예측(업무 적용)



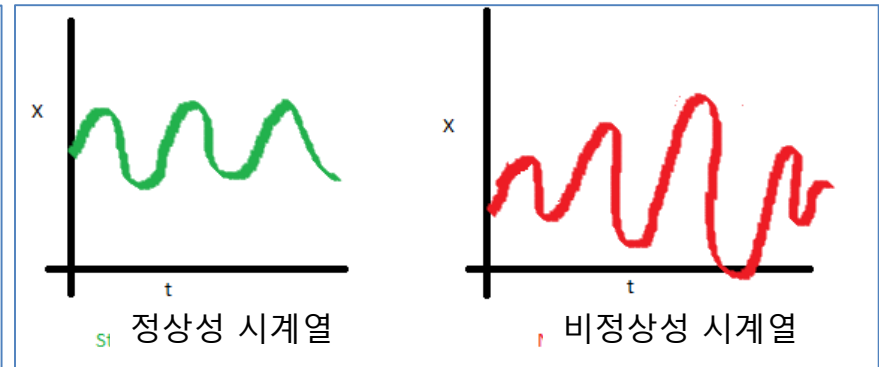
시계열 분석

1단계. 시계열 자료 특성 분석 : 비정상성과 정상성 시계열 확인

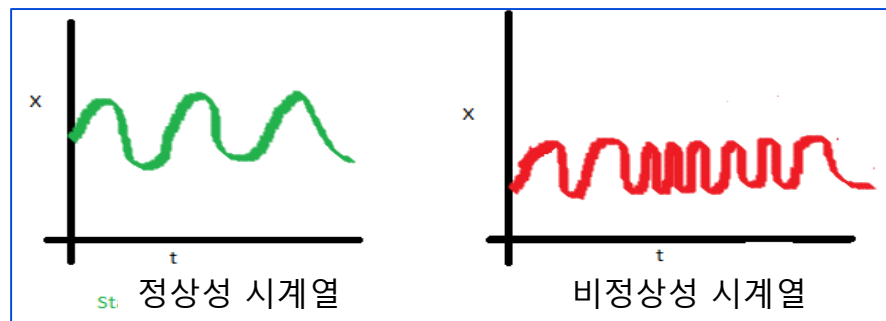
1. 시간의 추이와 관계 없이 평균이 불변



2. 시간의 추이와 관계 없이 분산이 불변



3. 두 시점 간의 자기상관(공분산)이 기준시점과 무관

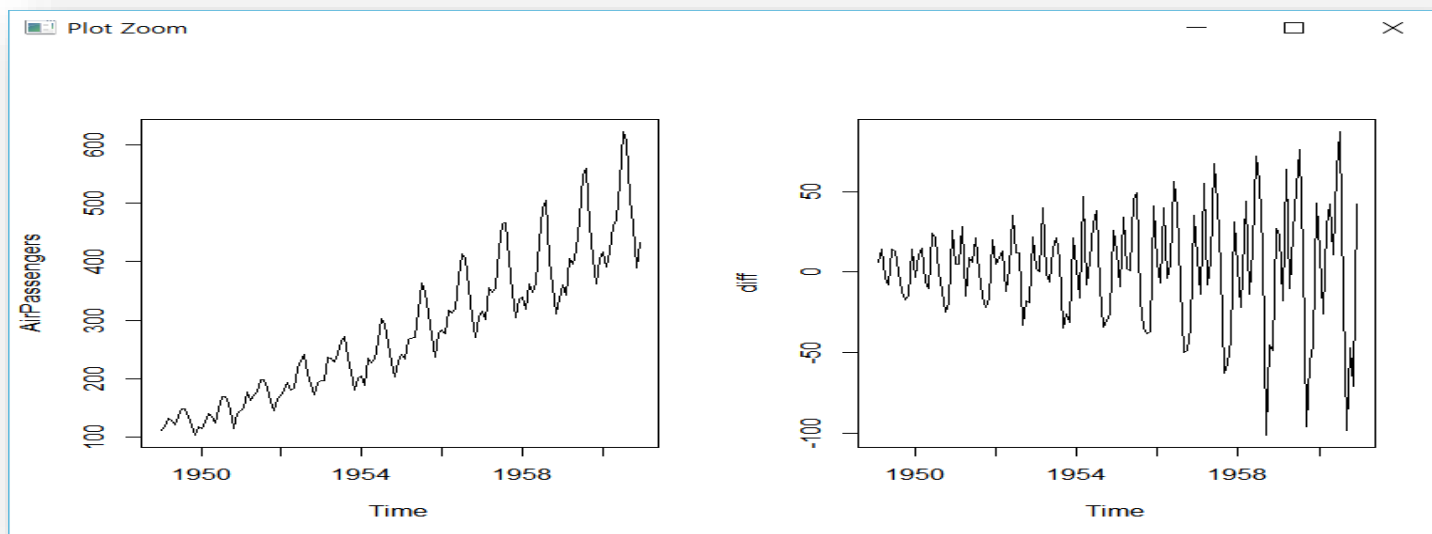


시계열 분석

2단계. 비정상성 시계열 -> 정상성 시계열 변환

✓ 차분(diff)과 로그함수(log) 적용

(1) 차분 : 점차적으로 하강/상승하는 추세 요인 제거(평균의 정상화)



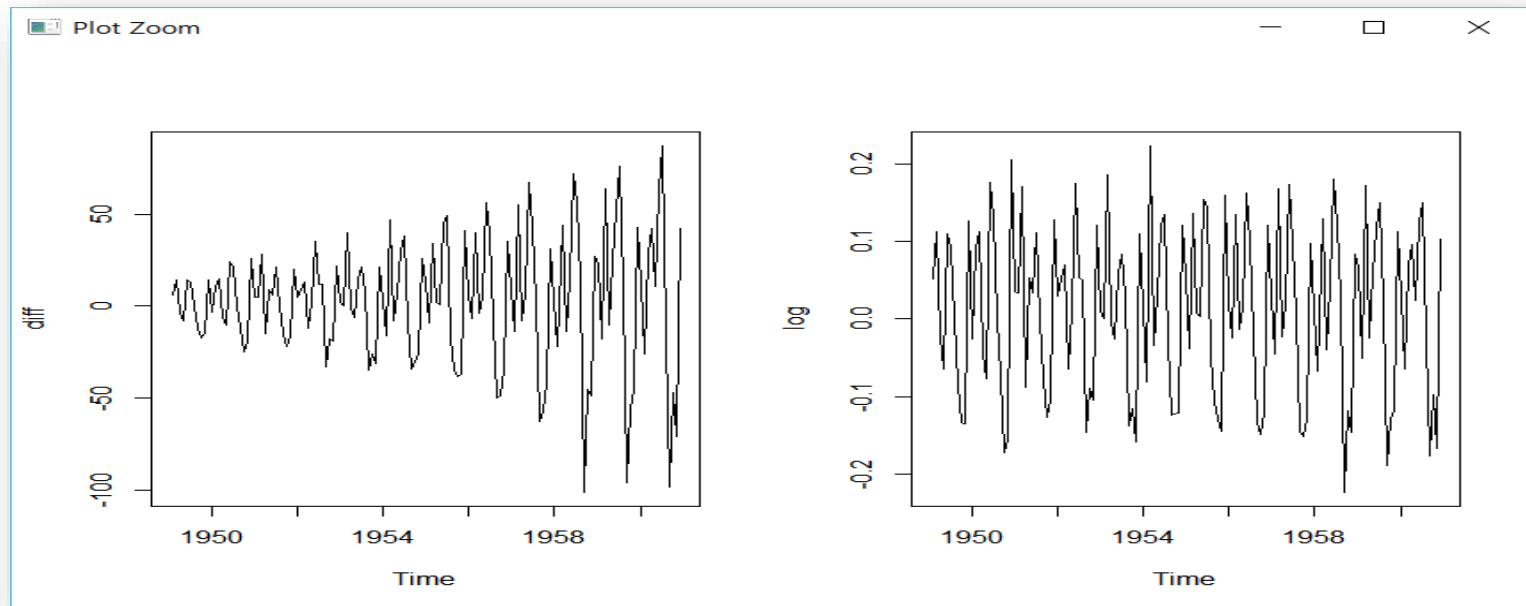
시계열 분석

2단계. 비정상성 시계열 -> 정상성 시계열 변환

✓ 차분(diff)과 로그함수(log) 적용

(2) 로그 함수 적용 : 시간의 변화에 따라서 변동 크기가 변화되는

분산의 정상화(로그변환 = 대수변환)



시계열 분석 기법

3단계. 모형 식별과 추정 : `auto.arima()` 함수 이용

- ✓ `auto.arima` 함수 : ARIMA 모형의 초적화된 파라미터 제공
- ✓ ARIMA : 비정상성을 가진 시계열 자료를 모형 생성

형식) $\text{ARIMA}(p, d, q)$: 3개 파라미터

✓ p : AR차수, d : 차분차수, q : MA 차수

✓ `auto.arima()` 함수 : 모형과 차수 제공

자기회귀모형(AR)

이동평균모형(MA)

자기회귀이동평균모형(ARMA)

[$\text{ARIMA}(p,d,q)$ 모형 \rightarrow 정상성 시계열 모형 식별]

$d=0$ 이면, $\text{ARMA}(p, q)$ 모형이며, 정상성을 만족한다.

$q=0$ 이면 $\text{IAR}(p, d)$ 모형이며, d 번 차분하면 $\text{AR}(p)$ 모형을 따른다.

$p=0$ 이면 $\text{IMA}(d, q)$ 모형이며, d 번 차분하면 $\text{MA}(q)$ 모형을 따른다.

시계열 분석 기법

```
library(forecast)
```

```
arima <- auto.arima(tsdata) # 모형 식별과 파라미터 예측
```

```
arima
```

```
Series: tsdata
```

```
ARIMA(1,1,0)
```

```
Coefficients:
```

```
ar1
```

```
-0.6891
```

```
s.e. 0.2451
```

```
sigma^2 estimated as 31644: log likelihood=-72.4
```

```
AIC=148.8 AICc=150.3 BIC=149.59
```

1번 차분한 결과가 정상성시계열의 AR(1) 모형으로 식별된
AIC(Akaike's Information Criterion)/BIC(Bayesian Information Criterion) :
이론적 예측력을 나타내는 지표 (값이 적은 모형 채택)

시계열 분석 기법

4단계. 모형 생성

이전 단계에서 식별된 모형과 파라미터를 이용하여 시계열 모형 생성

```
model <- arima(tsdata, order=c(1, 1, 0))
```

```
model
```

Call:

```
arima(x = tsdata, order = c(1, 1, 0))
```

Coefficients:

ar1

-0.6891

s.e. 0.2451

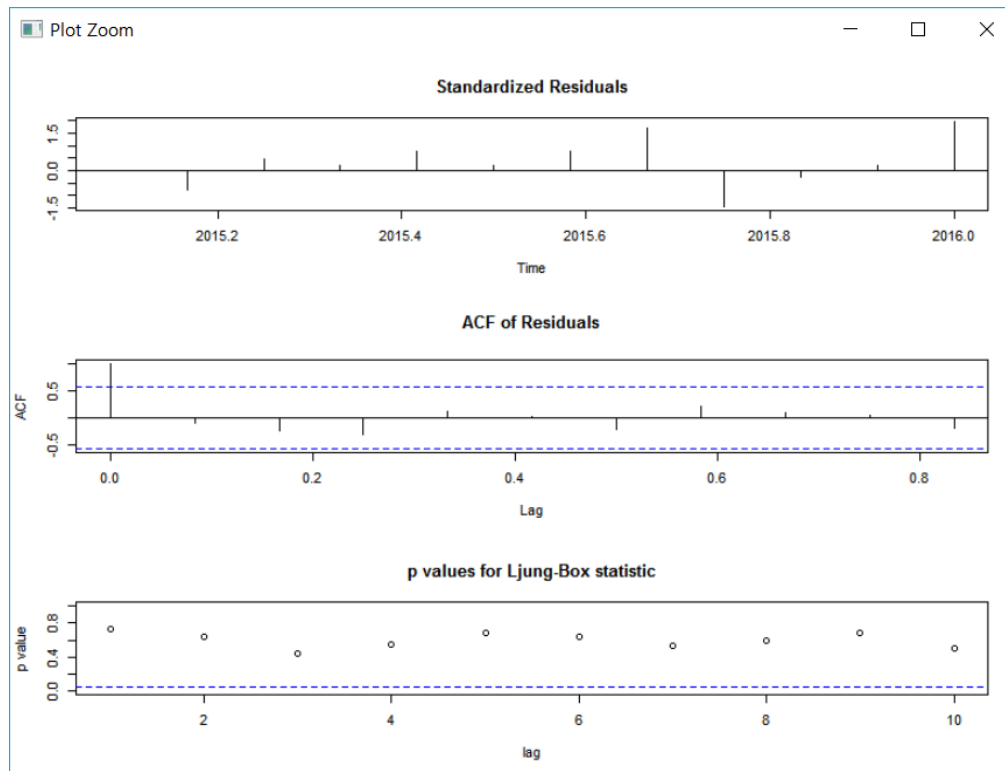
sigma^2 estimated as 28767: log likelihood = -72.4, aic = 148.8

시계열 분석 기법

5단계. 모형 진단(모형 타당성 검토)

잔차가 백색 잡음(white noise) 검토(모형의 잔차가 불규칙적이고, 독립적)

(1) 자기상관함수에 의한 모형 진단



좋은 시계열 모형은 잔차의 ACF에서 자기상관이 발견되지 않고, p value값이 0 이상으로 분포(ARIMA모형은 매우 양호한 시계열 모형)

시계열 분석 기법

(2) Box-Ljung에 의한 잔차항 모형 진단

```
Box.test(model$residuals, lag=1, type = "Ljung")
```

Box-Ljung test

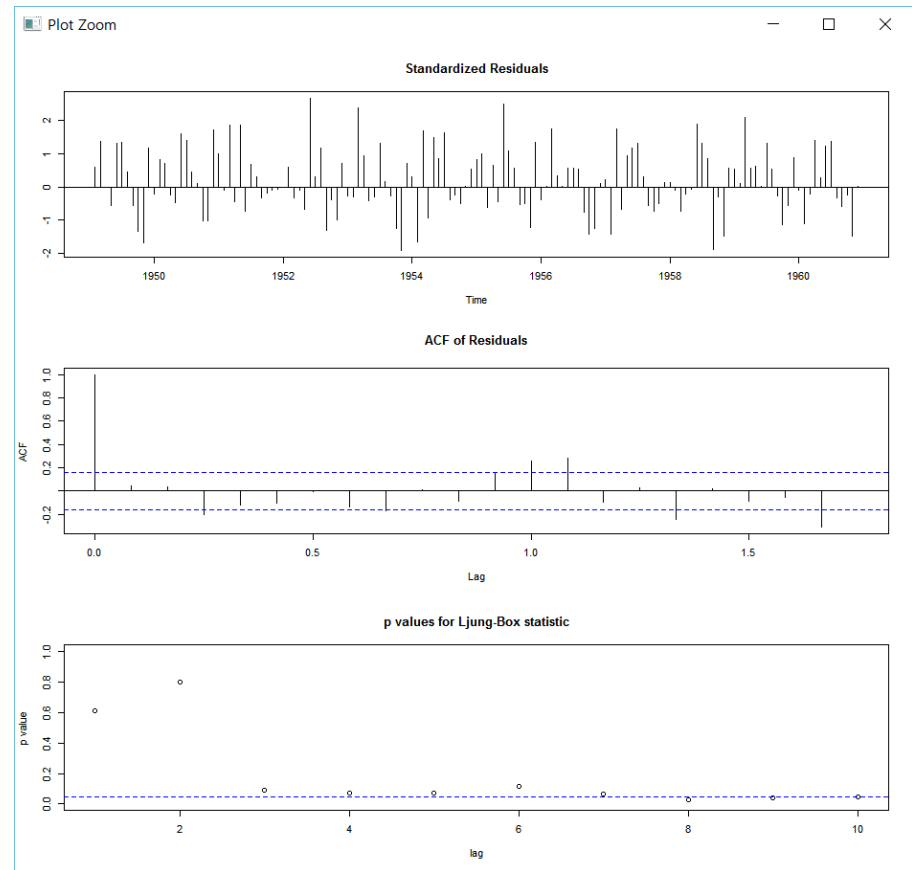
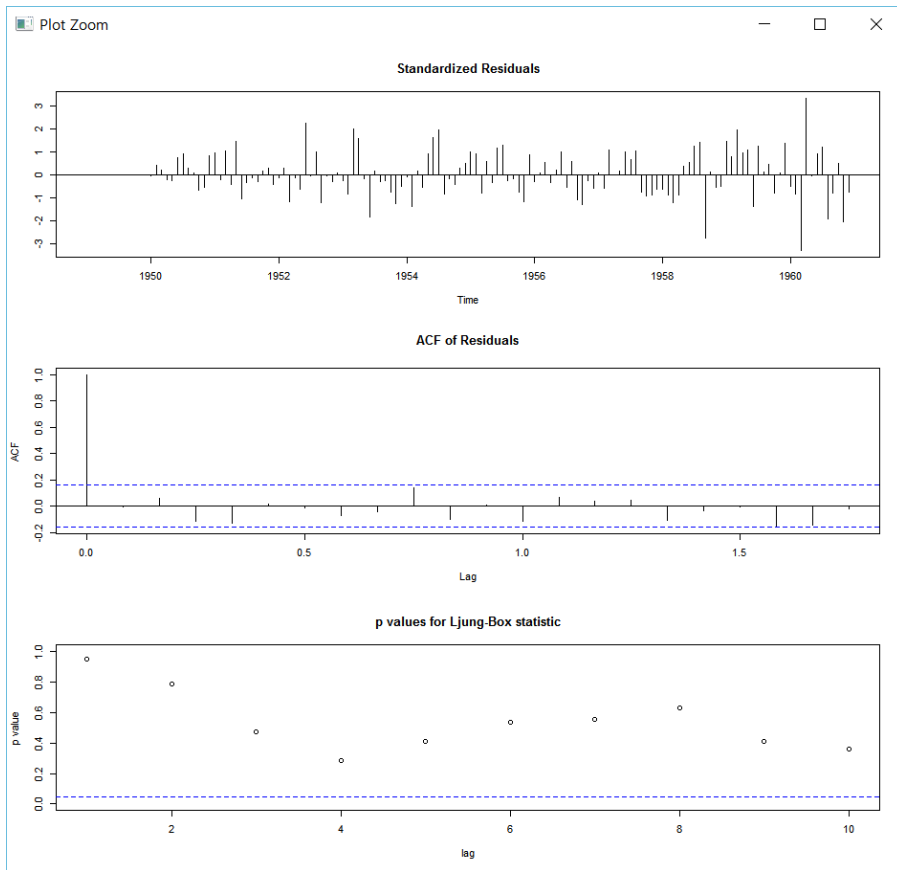
data: model\$residuals

X-squared = 0.12353, df = 1, p-value = 0.7252

Box-Ljung 검정방법은 모형의 잔차를 이용하여 카이제곱검정 방법으로 시계열 모형이 통계적으로 적절한지를 검정하는 방법으로 p-value가 0.05 이상이면 모형이 통계적으로 적절하다고 볼 수 있다.

시계열 분석 기법

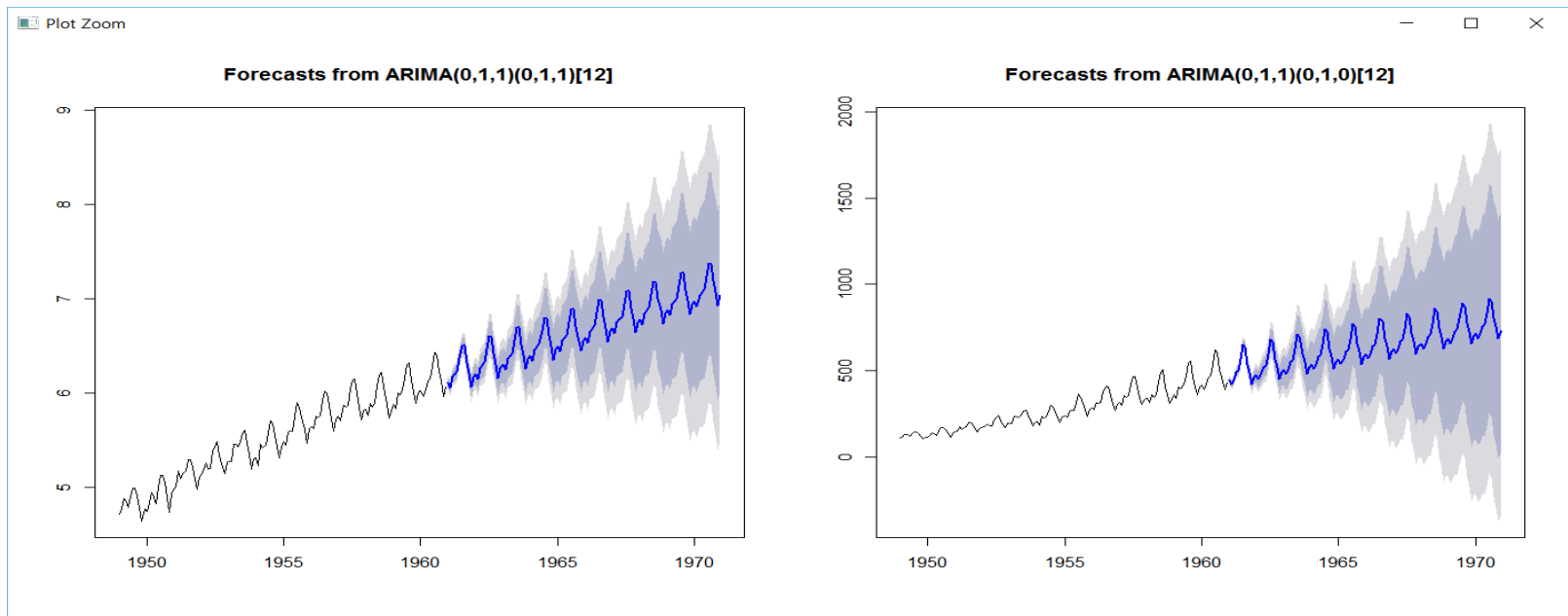
정상성 시계열 vs 비정상성 시계열



시계열 분석 기법

6단계. 미래 예측

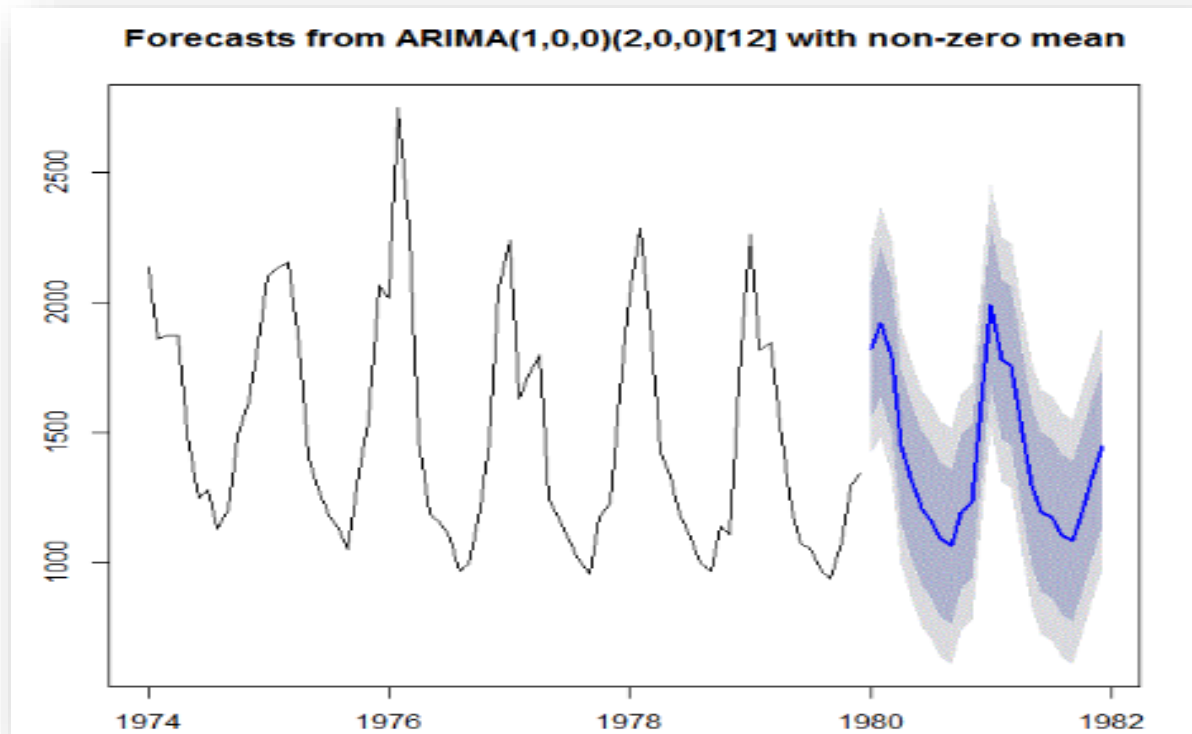
- 모형 진단을 통해서 적절한 모형으로 판단되면 이 모형으로 가까운 미래를 예측하는데 이용
- forecast 패키지에서 제공하는 forecast() 함수는 시계열의 예측치를 제공하는 함수로 기본 기간은 2년(24개월)



시계열 자료 미래 예측 예제

mdeaths 샘플 데이터 이용 시계열 데이터 미래 예측

mdeaths # 영국인 사망 관련 시계열 데이터
fit <- auto.arima(mdeaths)



R 시계열분석 Time Series ARIMA

1. 시계열 자료 - 시간의 흐름에 따라 관찰된 데이터
2. 정상성 대부분의 시계열 자료는 다루기 어려운 비정상성 시계열 자료이기 때문에 분석하기 쉬운 정상성 시계열 자료로 변환
 - (1) 평균이 일정 : 모든 시점에 대해 일정한 평균을 가진다.
 - 평균이 일정하지 않은 시계열은 차분(difference)을 통해 정상화.
 - 차분은 현시점 자료에서 이전 시점 자료를 빼는 것.
 - (2) 분산도 시점에 의존하지 않음 - 분산이 일정하지 않은 시계열은 변환(transformation)을 통해 정상화.
 - (3) 공분산도 시차에만 의존할 뿐, 특정 시점에는 의존하지 않음.

R 시계열분석 Time Series ARIMA

3. 시계열 모형

- (1) 자기회귀 모형 (Autoregressive model, AR) P 시점 이전의 자료가 현재 자료에 영향을 줌
오차항 = 백색잡음과정(white noise process) 자기상관함수(Autocorrelation Function, ACF) : k
기간 떨어진 값들의 상관계수 부분자기상관함수(partial ACF) : 서로 다른 두 시점의 중간에
있는 값들의 영향을 제외시킨 상관계수 ACF 빠르게 감소, PACF는 어느 시점에서 절단점을
갖는다 PACF가 2시점에서 절단점 가지면 AR(1) 모형
- (2) 이동평균 모형 (Moving average model, MA) 유한한 갯수의 백색잡음 결합이므로 항상 정상성
만족 ACF가 절단점을 갖고, PACF는 빠르게 감소
- (3) 자기회귀누적이동평균 모형 (Autoregressive integrated moving average model, ARIMA) 비정상
시계열 모형 차분이나 변환을 통해 AR, MA, 또는 이 둘을 합한 ARMA 모형으로 정상화
ARIMA(p, d, q) - d : 차분 차수 / p : AR 모형 차수 / q : MA 모형 차수
- (4) 분해 시계열에 영향을 주는 일반적인 요인을 시계열에서 분리해 분석하는 방법 - 계절
요인(seasonal factor), 순환 요인(cyclical), 추세 요인(trend), 불규칙 요인(random)