**Time-Series Analysis**

1. 안녕하세요. 데이터분석 캡스톤 디자인 사례 연구 발표를 하게 된 이인석 입니다.
2. 먼저 관심분야 사례에 대해 발표한 후 관심 프로젝트 주제 발표하는 순서로 진행하도록 하겠습니다.
3. 저는 시계열 데이터 분석에 대해 조사를 해봤습니다. 시계열 데이터 분석은 일정 기간 동안 변화하는 데이터 세트의 분석입니다. 그리고 이 시계열 데이터 세트는 다양한 시점 에서 동일한 변수 의 관찰을 통해 기록됩니다.
4. 시계열 데이터에서 가장 먼저 그려야 할 것은 시간 그래프(time plot)입니다. 즉, 관측값을 관측 시간에 따라 인접한 관측값을 직선으로 연결하여 그리는 것입니다. 다음 그래프는 호주 당뇨병 약의 월별 매출을 나타냅니다. 분명하게 증가하는 추세를 이 그림에서 확인할 수 있습니다. < 매년 연초마다 의약품 매출이 급감하는데, 이것은 환자들이 연말에 저렴하게 의약품을 비축할 수 있도록 하는 정부의 보조금 정책 때문입니다. 이 시계열을 이용하는 어떤 예측이라도 위에서 언급한 계절성 패턴과 더불어 변화의 추세가 느리다는 사실을 그 예측치 안에 담아내야 합니다.>

**시계열 분석 절차**

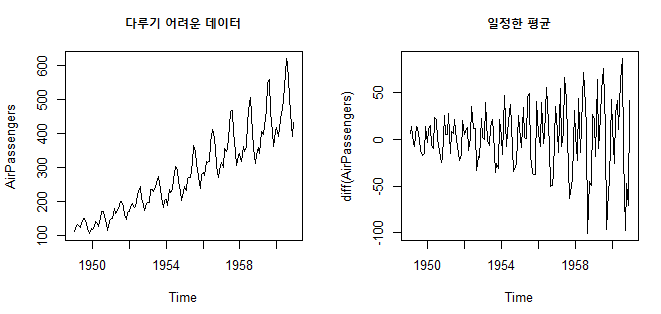
1. 시간그래프 그리기
2. 추세와 계절성을 제거하기
3. 잔차를 예측하기
4. 잔차에 대한 모델 적합하기
5. 예측된 잔차에 추세와 계절성을 더하여 미래를 예측하기
6. 시계열 데이터 특성 파악

시계열 분석에서 확인해야 할 가장 중요한 사항 중 하나는 데이터 특성 파악이다. 예를 들어 경향(trend), 주가(cycle), 계절성(seasonaity), 불규칙성(irregular), 정상성(stationary) 등이 있다. 이 중 가장 큰 특징 중 하나인 정상성은 정상성은 시계열의 수준과 분산에 체계적인 변화가 없고 엄밀하게 주기적 변동이 없다는 것으로 미래는 확률적으로 과거와 동일하다는 것을 의미한다.

정상성을 만족하는 시계열 은 다음 세가지 조건을 만족함

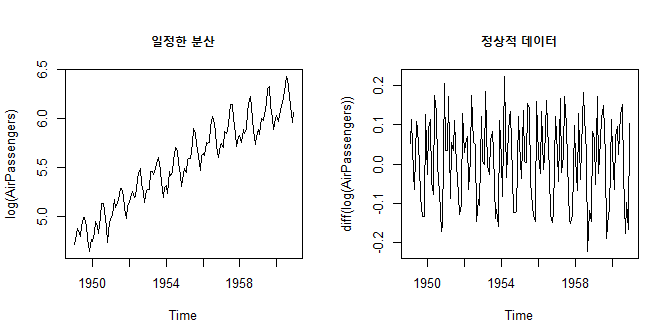
- 1. 평균값은 시간 t에 관계없이 일정한다.

모든 시점에서 일정한 평균을 가짐

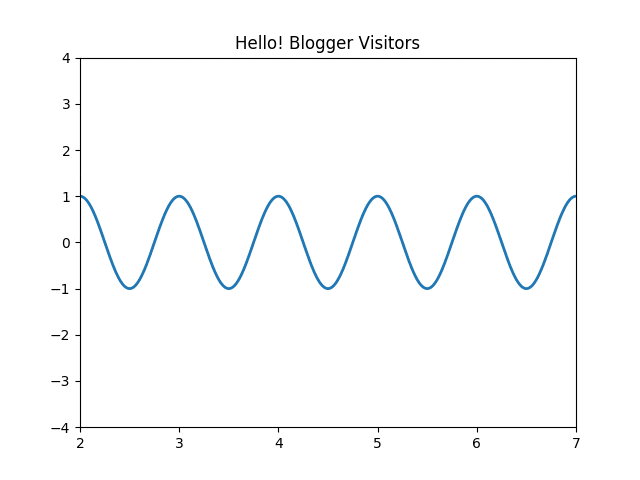


- 2. 분산값은 시간 t에 관계없이 일정하다.

시점에 의존하지 않고 일정해야 함



- 3. 공분산은 시간 t에 의존하지 않고 오직 시차에만 의존한다. 다음 그래프처럼 일정 시간마다 변해야 한다. 시간 차이를 기준으로 공분산이 막 불규칙하다면 그건 정상성을 만족하지 않아 정상 시계열이 아니다. 특정시점에서 t, s에 의존하지 않고 일정



위 세 가지 조건 중 하나라도 만족하지 않으면 정상 시계열이 아니다.

안타깝게도 대부분 시계열 자료는 비정상성 자료이다.

비정상 시계열을 정상 시계열로 전환하는 방법

- 1. 시계열 평균이 비일정하면 원시계열에 차분(현 시점에서 전 시점의 자료 값을 뺌)하면 정상 시계열이 된다.

- 2. 계절성을 갖는 비정상시계열은 정상시계열로 바꿀 때 계절차분을 사용한다.

- 3. 분산이 일정하지 않은 경우 원계열에 자연로그(변환)을 취하면 정상시계열이 된다.

대표 정상시계열인 백색잡음으로 만드는 백색잡음 과정은 시계열 et의 평균이 0이고 분산이 일정한 값 2 시그마이고 자기공분산이 0인 경우를 의미한다.

+ 자기상관 : 시점 t와 (t-1) 간 상관관계, 일정 기간동안 증가하거나 감소하는 경우 자기상관이 존재, 증감이 시점마다 반복되면 음의 자기상관인데 양의 자기상관이 일반적이다.

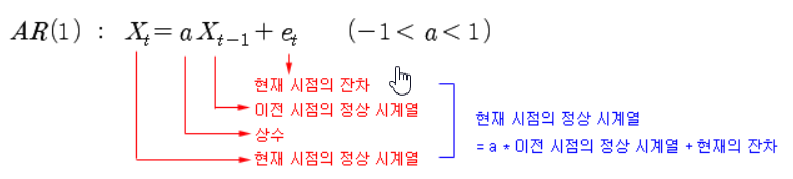
시계열 모형

1. **자기회귀모형(AR, AutoRegressive) :**

자신의 과거 값을 사용하는 모형. 현재 시계열 자료의 몇 번째 전 자료까지 영향을 주는가를 구하는 것이 목적.

=> 현시점의 시계열 자료에 과거 1시점 이전 자료만 영향을 준다면 1차 자기회귀모형, AR(1)

=> 다음이 수식



​

자기회귀모형인지 판단하기 위한 척도로 자신과의 상관성과 부분자기상관함수를 사용한다.

보통 자기회귀모형은 자기상관함수가 시차 증가에 따라 점차 감소하고 부분자기상관함수는 시차 후 급격히 감소해 절단된 형태를 취한다.

​

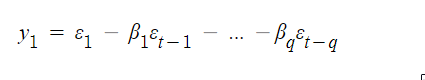
1. **이동평균모형(MA, Moving Average) :**

최근 데이터의 평균이나 중앙치를 예측치를 사용하는 방법으로 각 과거치에는 동일한 가중치가 주어진다.

=> 현시점의 자료를 유한 개의 백색잡음의 선형결합으로 표현되어 항상 정상성을 만족한다.

=> 자기회귀모형과 반대로 자기상관함수 p + 1 시차 이후 절단된 형태를 취한다.

=> 다음이 수식



​

1. **자기회귀누적 이동 모형(ARMIA) :**

대부분의 많은 시계열 자료가 따르는 모형으로 기본 비정상시계열이며 차분이나 변환으로 AR, MR, ARMA 모형으로 정상화할 수 있다.

=> 차수 p, d, q 값에 따라 모형의 이름이 바뀜

=> 차수 p는 AR과 관련, q는 MA와 관련, d는 ARIMA에서 ARMA로 정상화할 때 몇 번 차분했는지를 물음.

=>> AR : 과거 시점의 관측자료의 선형결합으로 표현

=>> MR : 과거 시점의 백색잡음 선형결합으로 표현

=>> ARMA : 위 두 모형을 합친 모형 (자기회귀 이동 모형)

1. 분해 시계열

시계열에 영향을 주는 일반적인 요인을 시계열에서 분리해 분석하는 방법

1. 추세요인

* 형태가 오르거나 내리는 추세, 선형, 이차식, 지수형태

1. 계절요인

* 요일, 월, 사분기 별로 변화하여 고정된 주기에 따라 자료가 변화

1. 순환요인

* 명백한 경제적, 자연적 이유없이 알려지지 않은 주기로 자료가 변화

1. 불규칙요인

* 위 세 가지의 요인으로 설명할 수 없는 회귀분석에서 오차에 해당하는 요인

출처: <https://m.blog.naver.com/dsz08082/222042702104>

* 자연언어처리-감성분석- 기업과 관련된 뉴스나 SNS 홍보물 등에 달린 댓글의 긍/부정을 판단하는 곳에 사용
* 머신러닝을 활용한 금융 사기 탐지
* Review 데이터를 사용한 책, 영화 추천 시스템 구축