시계열자료분석

1. 시계열자료 분석 기초

* 시계열 데이터?
* 관측치가 시간적 순서를 가진 데이터

Ex) 주식가격 데이터, 실업률, 기후 데이터 등

* 시계열 변동요인
* 추세요인: 장기간&점진적으로 상승하거나 하강
* 순환요인: 수년간의 간격으로 주기적인 패턴
* 계절요인: 단기적이고 주기적인 패턴을 갖고 반복적인 경향
* 불규칙요인: 명확히 설명할 수 없는 요인에 의한 우연변동

ex) 지진, 해일 등

* 시계열 자료분석 순서

1. 시계열 자료 구조파악
2. 시계열 자료의 구조와 특성을 토대로 미래의 값 예측

* 정상성 데이터인가? (미래를 예측하기 위해서는 정상성 데이터를 이용)
* 정상성의 기준?

i. 모든 시점에 대해 일정한 평균을 가짐

ii. 평균이 일정하지 않고 분산도 시점에 의존하지 않는다

iii. 공분산도 단지 시차에만 의존할 뿐 실제 특정 시점 t,s에는 의존하지 않는다.

1. 생성된 시스템 제어

* 목적: 시계열이 갖고 있는 법칙성을 발견하여 이를 모형화하고, 추정된 모형을 통하여 미래 예측
* 시계열 자료의 예측방법

1. 양적예측방법(경험적 법칙을 추정하여 예측)

* 가정: 과거의 패턴은 미래에서도 지속될 것이다.

|  |  |
| --- | --- |
| 시계열분석방법 | 종류 |
| 전통적 시계열분석방법 | 평활법, 분해법 |
| 확률적 시계열분석방법 | ARIMA |

1. 질적예측방법(주관적인 견해를 사용하여 예측)

* 과거의 정보가 없거나 불충분한 경우에 사용
* 시계열 데이터 예측 평가방법

1. 데이터를 관측가능기간과 예측기간으로 구분

* 관측가능기간의 데이터를 사전평가, 예측기간의 데이터를 사후평가로 구분

(미래의 데이터는 구할 수 없으므로 사후평가데이터라고 가정)

1. 사전평가에서 모형을 추정하고 사후평가에서 추정된 모형으로 예측

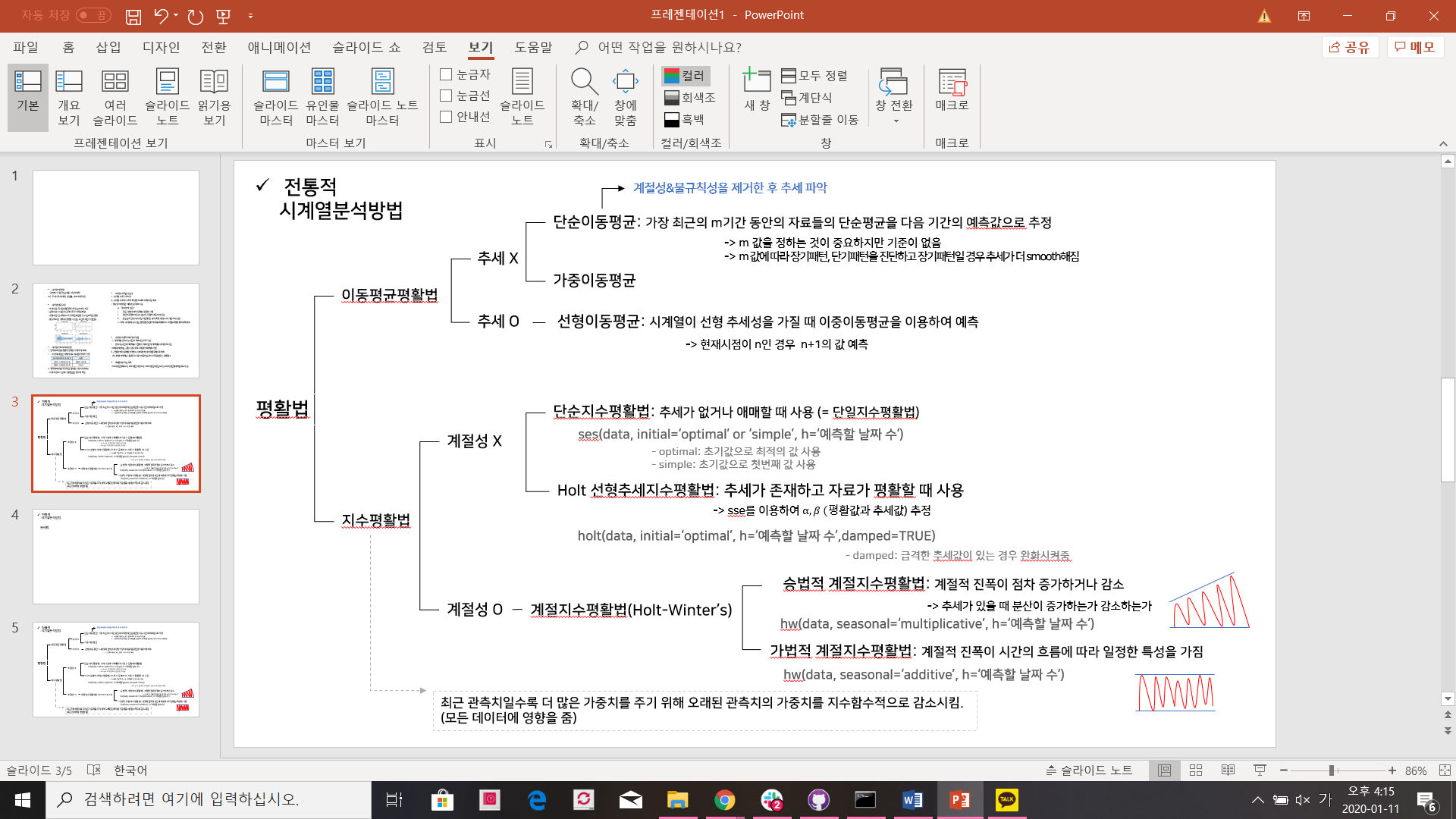
cf) 시계열데이터는 시간을 갖고 있기 때문에 순서가 바뀌면 안되기 때문에 샘플링 X.

* 시계열 모형을 평가하는 척도
* MAE(평균절대오차), MSE(평균제곱오차), RMSE(평균제곱근오차) 등 여러 척도가 있지만 이들은 의존성을 갖고 있기 때문에 MAPE(평균절대백분비오차), MASE로 측정하는 것을 추천

1. 시계열 예측방법

* 평활법을 이용한 시계열자료분석
* 특정 패턴이 시계열에 내재되어 있을 때, 원계열을 평활화하여 불규칙변동을 제거함으로써 추정이 가능하다는 것이 전제

|  |  |
| --- | --- |
| 평활법 종류 | |
| 이동평균평활법 | 각 데이터에 동일한 값만큼 가중치 부어 |
| 지수평활법 | 최근 관측치일수로 더 많은 가중치를 부여, 오래된 관측치의  가중치를 지수함수적으로 감소시킴 |



* 분해법을 이용한 시계열자료분석
* 시계열의 패턴을 개별 성분(추세-순환성분, 계절성분, 불규칙성분)으로 분해하여 시계열의 특성 분석

cf) 추세성분과 순환성분 왜 같이 쓰는가?

: 순환성분이 장기적이기 때문에 데이터가 많아지면 모호해짐

* 분해된 각 성분들을 개별적으로 예측한 후에 다시 결합시켜 예측
* 문제점: 데이터 손실
* ARIMA를 이용한 시계열자료분석(Box-Jenkins 방법론 사용)

1. 모형식별

* 정상성을 가진 데이터인지 통계검정을 이용하여 확인

\*평균이 일정하지 않으면 차분을 통해 정상화

\*분산이 일정하지 않으면 로그를 통해 정상화

-> 정상 데이터로 변환 후 재검정

* ACF(자기상관함수)와 PACF(편자기상관함수)를 이용하여 모형후보군 도출

\*AR모형: 현 시점의 자료가 과거 p시점 전의 자료와 관계가 있는 모형

(ACF가 빠르게 감소하고 PACF는 어느 시점에서 절단점을 갖고 있음)

\*MA모형: 유한한 백색잡음의 결합으로 언제나 정상성을 만족하는 모형

(PACF가 빠르게 감소하고 ACF는 어느 시점에서 절단점을 갖고 있음)

\*ARIMA모형: 불안정한 시계열 자료를 설명하기 위하여 차분을 통해 비정상성을 제거하여 ARMA모형으로 정상화

1. 모수추정

* AIC와 BIC를 통해 모수 추정

1. 모형 적합성

* 잔차의 ACF와 PACF 시각화 및 통계검정을 통해 모형 적합성 확인

1. 미래 예측값 추정