

#5 시계열 분석



11기 Java 김민혁

목차

5.1 시계열 데이터란

5.2 시계열 분해

5.3 정상성

5.4 Future Work




시계열 데이터란

- 개념
 - 일정한 시간 동안 수집된 일련의 순차적으로 정해진 데이터 세트의 집합
- 종류
 - 주가
 - ECG (심장 정보 데이터)
 - 언어



시계열 분해

- 시계열 데이터 구성 요소
 - 추세 (Trend)  시계열의 장기간에 걸친 점진적이고 지속적인 변화 상태
 - 계절성 (Seasonality)
 - 주기성 (Cycle)
 - 잔차 (Residual)

시계열 분해

- 시계열 데이터 구성 요소
 - 추세 (Trend)
 - 계절성 (Seasonality) ← 계절적 요인의 영향을 받아 1년, 혹은 일정 기간 안에 반복적으로 나타나는 패턴
 - 주기성 (Cycle)
 - 잔차 (Residual)

시계열 분해

- 시계열 데이터 구성 요소
 - 추세 (Trend)
 - 계절성 (Seasonality)
 - 주기성 (Cycle) ← 정해지지 않은 빈도, 기간으로 일어나는 상승 혹은 하락
 - 잔차 (Residual)

시계열 분해

- 시계열 데이터 구성 요소
 - 추세 (Trend)
 - 계절성 (Seasonality)
 - 주기성 (Cycle)
 - 잔차 (Residual) ← 시계열 분해 후 남는 나머지. 잔차의 정보량을 최소화해야 함.

시계열 분해

- 시계열 데이터 구성 요소
 - 추세 (Trend)
 - 계절성 (Seasonality)
 - 주기성 (Cycle)
 - 잔차 (Residual)

정보의 분리가 필요하다.



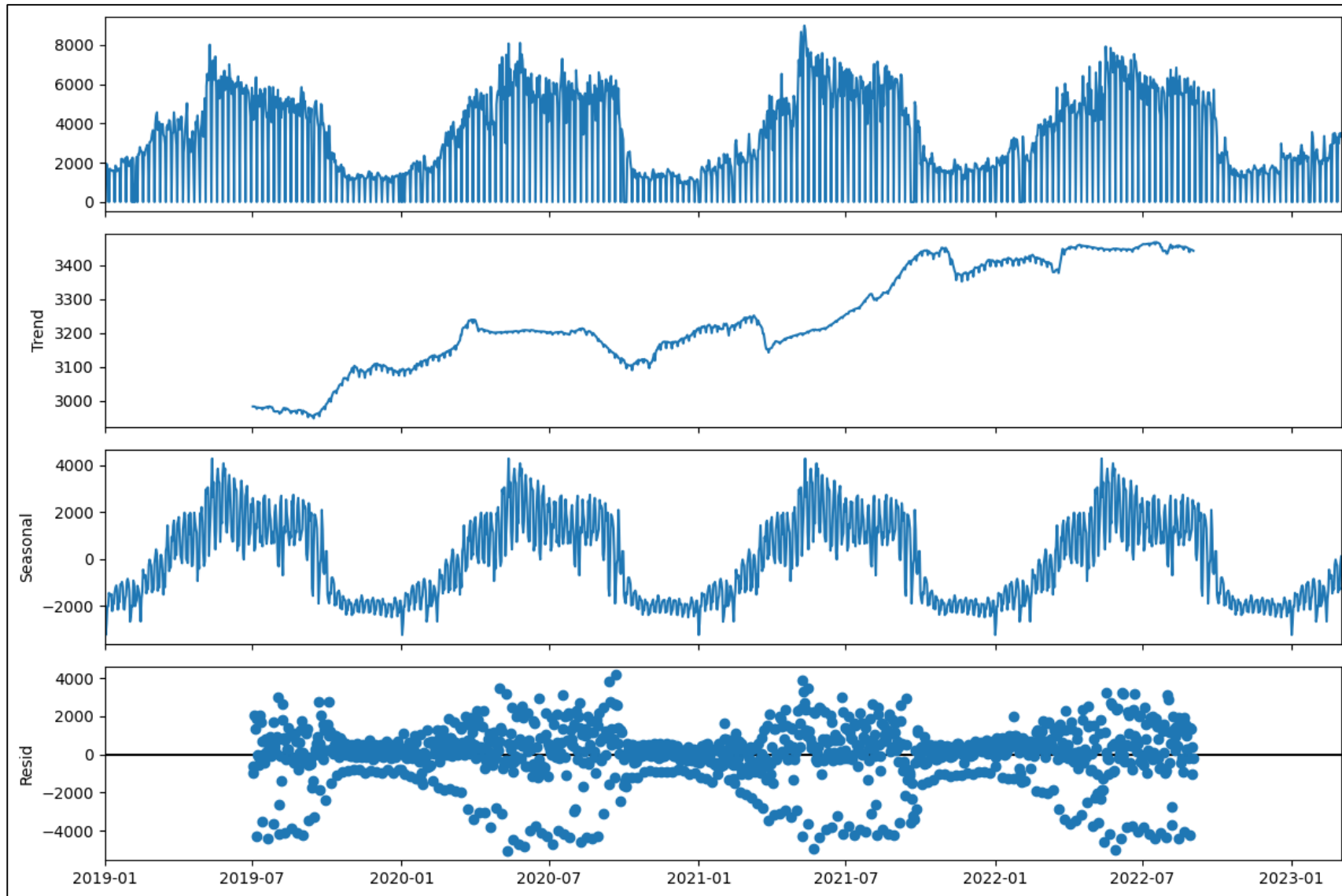
시계열 분해

```
import pandas as pd
from statsmodels.tsa.seasonal import seasonal_decompose
import matplotlib.pyplot as plt

train = pd.read_csv("/content/drive/MyDrive/data_0417/train.csv")
train = train[train['item']=='TG']
train.drop(['ID', 'item', 'corporation', 'location', 'supply(kg)'], axis = 1, inplace=True)
train = train.groupby('timestamp').mean()
train.index = pd.to_datetime(train.index)
result = seasonal_decompose(train, model='additive', period = 365)

plt.rcParams['figure.figsize'] = [12, 8]
result.plot()
plt.show()
```

시계열 분해



정상성

관측값

예측값

현재 시점
T

정상성

- 정의
 - 확률변수 간의 확률 분포가 시간에 상관없이 일정한 확률 과정.
 - 분포가 시간과 독립적, 따라서 확률변수의 기댓값이나 분산 등의 값도 시간과 독립
- 종류
 - 강 정상성
 - 약 정상성

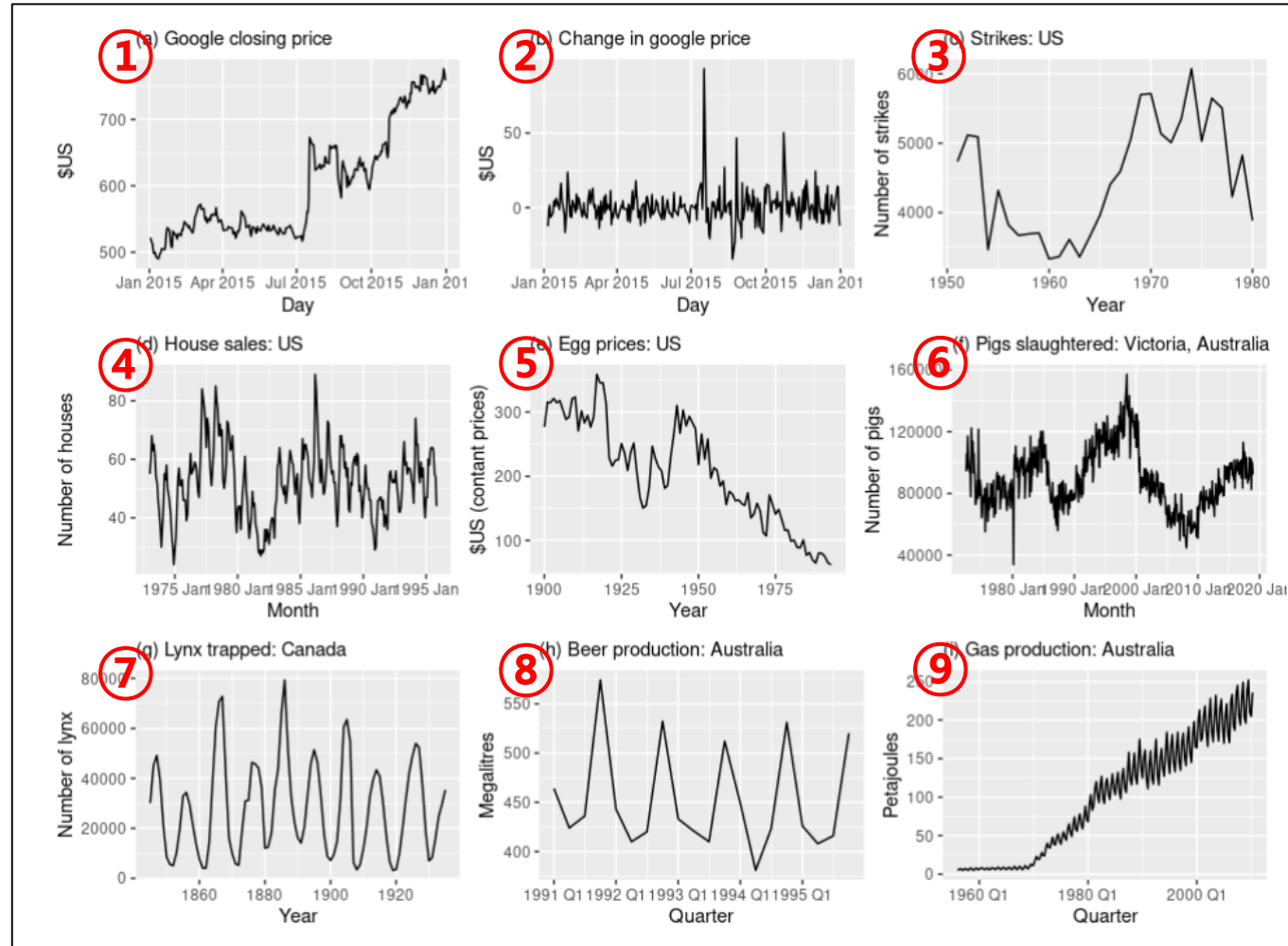
정상성

- 정의
 - 확률변수 간의 확률 분포가 시간에 상관없이 일정한 확률 과정.
 - 분포가 시간과 독립적, 따라서 확률변수의 기댓값이나 분산 등의 값도 시간과 독립
- 종류
 - 강 정상성 ← 기저를 이루는 확률 분포가 언제나 같음. (현실에선 어려움)
 - 약 정상성

정상성

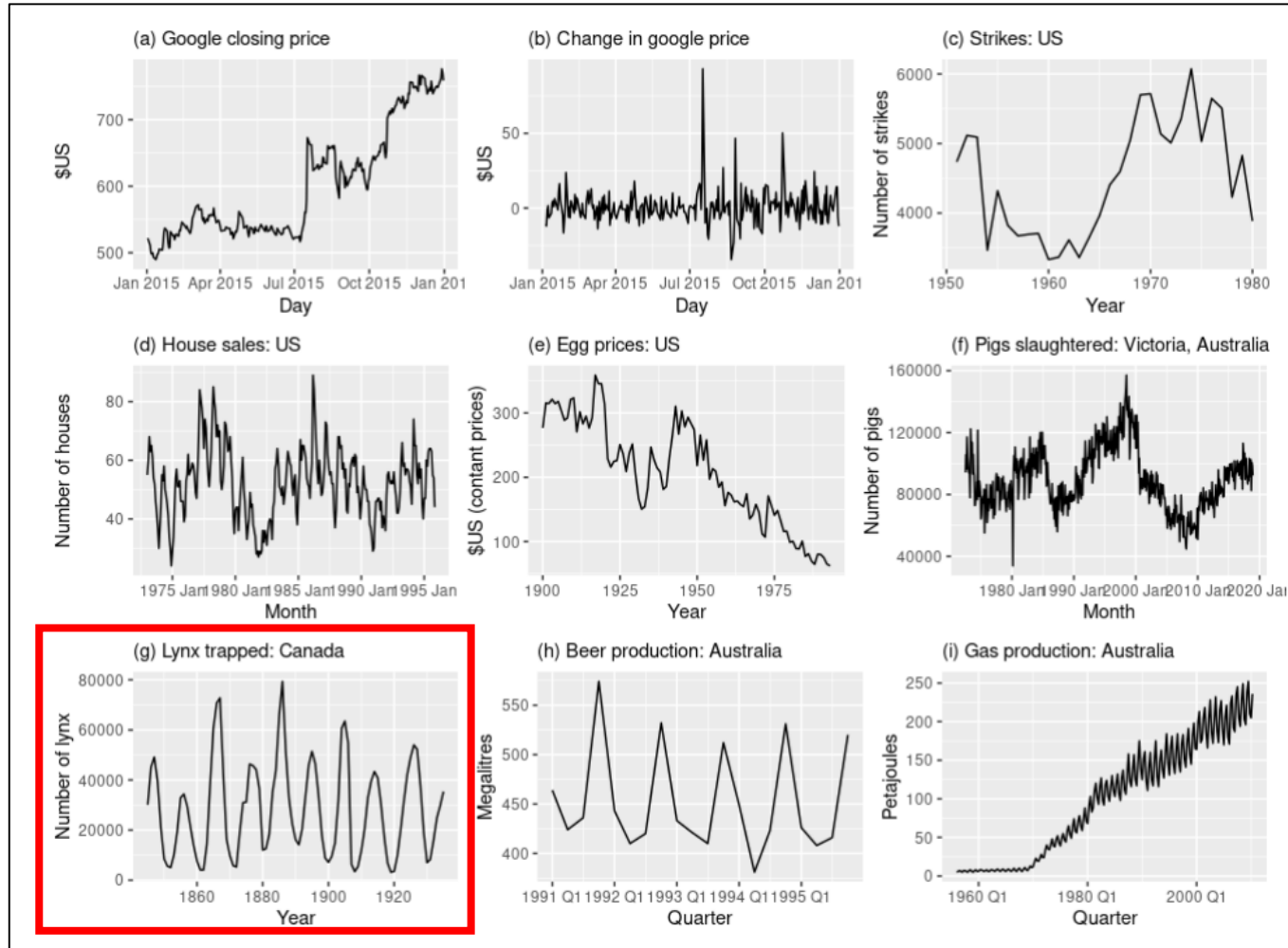
- 정의
 - 확률변수 간의 확률 분포가 시간에 상관없이 일정한 확률 과정.
 - 분포가 시간과 독립적, 따라서 확률변수의 기댓값이나 분산 등의 값도 시간과 독립
- 종류
 - 강 정상성
 - 약 정상성 ← 어느 시점(t)에 관측해도 확률 과정의 성질이 변하지 않음. (기댓값, 분산)

정상성



정상성

추세 X
계절성 X
분산 안정적



정상성

- 어떻게 정상적으로 만들 것인가?
 - 분산
 - Log Transformation
 - 추세, 계절성
 - 평활법 (Smoothing)
 - 차분 (Differencing)

정상성

- 어떻게 정상적으로 만들 것인가?
 - 분산
 - Log Transformation
 - 추세, 계절성
 - 평활법 (Smoothing) ← 회귀 분석 후, Residual로 추세와 계절성 제거
 - 차분 (Differencing)

정상성

- 차분
 - 계절성 차분 (Seasonal Differencing)

$$Y_{t'} = Y_t - Y(t - m)$$



주기

Future Work

- ARIMA
- ARCH



감사합니다.

