

09

정상성 확인 및 검정

[정상성 검정 및 확인]

정상성이란 추세와 계절성을 갖지 않고 진동하는 데이터입니다.

ARIMA 모델에 적용시키기 위해서는 정상성을 갖춘 데이터가 필요합니다.

00:01

정상성 검정 및 확인

정상성이란?

- 정상성은 추세와 계절성을 갖지 않는 진동하는 데이터이다.
- 주기성 행동을 가지고 있지만 주기가 특정 패턴을 갖지 않고 시계열을 관측하는 어떠한 시점에서도 똑같은 모양일 것이다..
- 그러므로 추세나 계절성이 없는 데이터 => ARIMA 모델을 적용하기 위한 조건
- 정상성을 띄지 않는다면 정상성(stationary)을 갖춘 형태로 변경해줘야 한다.

정상성 검정

Augmented Dickey-Fuller Test

- 정상성을 알아보기 위한 단위근 검정 방법
- 귀무가설 : 데이터가 정상성(stationary)을 갖지 않는다. (변동이 없다.)
- 대립가설 : 데이터가 정상성(stationary)을 갖는다

```
adfuller(x, maxlag, regression, autolag)
```

- x: 시계열 자료
- regression: 검정에 사용할 모형
{'nc': 추세 없음, 'n':상수항만 사용, 'ct':상수항&1차 추세, 'cct':상수항&2차 추세}
- autolag: 검정 AIC 모형의 지수를 자동 결정하는 알고리즘
{'AIC', 'BIC', 't-stat', None}

[전처리]

03:34

정상성 검정 및 확인

1. p-value
2. 사용된 시차의 수
3. 분석에 사용한 자료 수
4. 1%, 5%, 10% 수준에 해당되는 검정통계량 값

```
In [ ]: import pandas as pd
data = pd.read_csv('../data/arima_data.csv', names = ['day', 'price'])
data.head(3)
```

```
In [3]: data['day'] = pd.to_datetime(data['day'].format('%Y-%m-%d'))
data.set_index('day', inplace=True)
data.head(3)
```

```
Out[3]:
           price
day
2013-01-01  3794
2013-02-01  3863
2013-03-01  5190
```

```
In [21]: train_len = int(len(data) * 0.8)
training = data[:train_len]
test = data.drop(training.index)
```

```
In [22]: ## 귀무가설 : 데이터가 정상성(stationary)을 갖지 않는다. (변동이 없다.)
## 대립가설 : 데이터가 정상성(stationary)을 갖는다
from statsmodels.tsa.stattools import adfuller
```

[가설검정]

04:58

귀무가설 : 데이터가 정상성(stationary)을 갖지 않는다. (변동이 있다.)

대립가설 : 데이터가 정상성(stationary)을 갖는다

The screenshot shows a Jupyter Notebook interface with the following content:

```
data.head(3)
```

```
Out[9]:
```

day	price
2013-01-01	3794
2013-02-01	3863
2013-03-01	5190

```
In [10]: train_len = int(len(data) * 0.8)
training = data[:train_len]
test = data.drop(training.index)
```

```
In [22]: ## 귀무가설 : 데이터가 정상성(stationary)을 갖지 않는다. (변동이 있다.)
## 대립가설 : 데이터가 정상성(stationary)을 갖는다

from statsmodels.tsa.stattools import adfuller

adf = adfuller(training, regression='ct')

print('ADF Statistic: {}'.format(adf[0]))
print('p-value: {}'.format(adf[1]))

ADF Statistic: -1.9997199341327796
p-value: 0.6015863303794179
```

단위근 검정에서 p-value가 0.05보다 크므로 비정상적인 것으로 판단됨

처분 후 단위근 검정 재실행시

[ARIMA 모델의 파라미터 정하는 방법]

08:30

- ACF: Lag에 따른 관측치들 사이의 관련성을 측정하는 함수
- PACF: k 이외의 모든 다른 시점 관측치의 영향력을 배제한 후의 두 관측치의 관련성을 측정하는 함수