

09

## ARIMA

### [ARIMA 개념]

# ARIMA 모델은 AR모형과 MA모형을 합친 데이터 분석 기법입니다.

그리고 I는 정상성을 만족시키게 하기 위한 차분의 횟수를 의미합니다.

00:01

The screenshot shows a Jupyter Notebook titled '9.4 ARIMA'. The content includes a definition of ARIMA, a list of parameters (AR, MA, I) with handwritten notes, and a code cell for data loading and splitting. Handwritten red notes are present:

- AR(자기상관): 이전의 값이 이후의 값에 영향을 미치고 있는 상황  $p=2$
- MA(이동평균): 랜덤 변수의 평균값이 지속적으로 증가하거나 감소하는 추세  $q=2$
- I: 비정상성을 정상성으로 만들기 위해, 관측값들의 차이를 계산  $d=1$

The code cell contains the following Python code:

```
in [4]: import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('../data/arima_data.csv', names = ['day', 'price'])
data['day'] = pd.to_datetime(data['day'], format="%Y-%m-%d")
data.set_index('day', inplace=True)

#Train, Test 데이터 구분
train_len = int(len(data) * 0.8)
training = data[:train_len]
test = data.drop(training.index)
```

Below the code, the section '9.4.1 [Plot] 시계열 데이터 확인하기' is visible.

### [ARIMA 모형 생성]

# 정상성이란 추세와 계절성을 갖지 않고 진동하는 데이터입니다.

ARIMA 모델에 적용시키기 위해서는 정상성을 갖춘 데이터가 필요합니다.

01:58

Jupyter 9.3 정상성 검증 및 확인 (unsaved changes)

```
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: 0.0615863303794179
```

단위근 검정에서 p-value가 0.05보다 크므로 비정상적인 것으로 판단됨

처분 후 단위근 검정 재실행시

[ARIMA 해석]

04:05

Jupyter 9.4 ARIMA Last Checkpoint: 15시간 전 (unsaved changes)

```
In [8]: from statsmodels.tsa.arima.model import ARIMA
import statsmodels.api as sm

model = ARIMA(training.price.values, order=(2,1,2))

res = model.fit()

res.summary()
```

C:\Users\Wjang\Anaconda3\envs\WADP\_Class\lib\site-packages\statsmodels\tsa\statespace\sarimax.py:963: UserWarning: Non-stationary starting autoregressive parameters found. Using zeros as starting parameters.  
warn('Non-stationary starting autoregressive parameters')  
C:\Users\Wjang\Anaconda3\envs\WADP\_Class\lib\site-packages\statsmodels\tsa\statespace\sarimax.py:975: UserWarning: Non-invertible starting MA parameters found. Using zeros as starting parameters.  
warn('Non-invertible starting MA parameters found.')

```
Out[8]: SARIMAX Results
```

Dep. Variable:	y	No. Observations:	48
Model:	ARIMA(2, 1, 2)	Log Likelihood:	-375.875
Date:	Wed, 08 Dec 2021	AIC:	761.750
Time:	23:50:59	BIC:	771.000
Sample:	0	HQIC:	765.231

```
In [9]: help(ARIMA)
```

Help on class ARIMA in module statsmodels.tsa.arima.model: