

해커톤 발표 21조

김재현 박용정 박재형 전지은(조장)

목차

 주제

 코딩

 기대효과

 총평

주제

코로나 일일 확진자 수에 따른 주가 변동 예측 모델 만들기





코딩

1.데이터수집

2.상관분석

3.선형회귀모델

데이터 수집

< 라이브러리 >

- **pandas**
- **pandas_datareader.data**
- **BeautifulSoup**
- **Requests**


```
import requests
url = 'http://openapi.data.go.kr/openapi/service/rest/Covid19/getCovid19InfStateJson?serviceKey=L6ZlvNPfK1wZXBgZpxgKY%2BMFvWcLTfDe0InKDeUU6br4uNL5yETyhCRyUp9fIVeQr7XGly7RIYwbcUlgedLnSA%3D%3D&pageNo=1&numOfRows=10'
print(url)
```

```
http://openapi.data.go.kr/openapi/service/rest/Covid19/getCovid19InfStateJson?serviceKey=L6ZlvNPfK1wZXBgZpxgKY%2BMFvWcLTfDe0InKDeUU6br4uNL5yETyhCRyUp9fIVeQr7XGly7RIYwbcUlgedLnSA%3D%3D&pageNo=1&numOfRows=10
```

```
[ ] # url 출력
response=requests.get(url)
print(response.text)
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?><response><header><resultCode>00</resultCode><resultMsg>NORMAL SERVICE.</resultMsg></header><body><items><item><accDefRate>1.0584553504</accDefRate>
```

```
[ ] # 라이브러리
from bs4 import BeautifulSoup
import pandas
```

```
[ ] soup = BeautifulSoup(response.text, 'html.parser')
ItemList=soup.findAll('item')
```

```
[ ] # item 단위(일 단위로 구분)
for item in ItemList:
    print(item)
    print('-----')
```

```
-----
<item><accdefrate>1.2126452689</accdefrate><accexamcnt>990960</accexamcnt><accexamcompnt>962194</accexamcompnt><carecnt>889</carecnt><clearcnt>10506</clearcnt><createdt>2020-06-05 10:56:50.891</creat
```

CSV파일로 저장!

주식 데이터

-삼성전자

-신라호텔

```
File Edit View Insert Cell Kernel Widgets Help
+ ✂ 📄 📄 ⬆ ⬇ ▶ Run ■ ↺ ▶ Code ▾ 📄
```

```
In [1]: import pandas as pd

In [1]: import pandas_datareader.data as dr

In [3]: import datetime

In [4]: start = datetime.datetime(2019,10,1)

In [5]: end = datetime.datetime(2020,9,5)

In [6]: sh = dr.DataReader('008770.KS', 'yahoo', start, end)

In [7]: sh

Out[7]:
```

	High	Low	Open	Close	Volume	Adj Close
Date						
2019-10-01	86000.0	82100.0	85800.0	82500.0	653709.0	82500.0
2019-10-02	82900.0	80700.0	82700.0	81000.0	541381.0	81000.0
2019-10-04	81500.0	79200.0	81000.0	79800.0	435223.0	79800.0
2019-10-07	81500.0	79800.0	80400.0	80200.0	236583.0	80200.0
2019-10-08	81300.0	80200.0	80400.0	80800.0	266694.0	80800.0
...
2020-08-31	71600.0	70300.0	71600.0	70300.0	348867.0	70300.0

상관분석

```
corr = result2_na.corr(method='pearson')
```

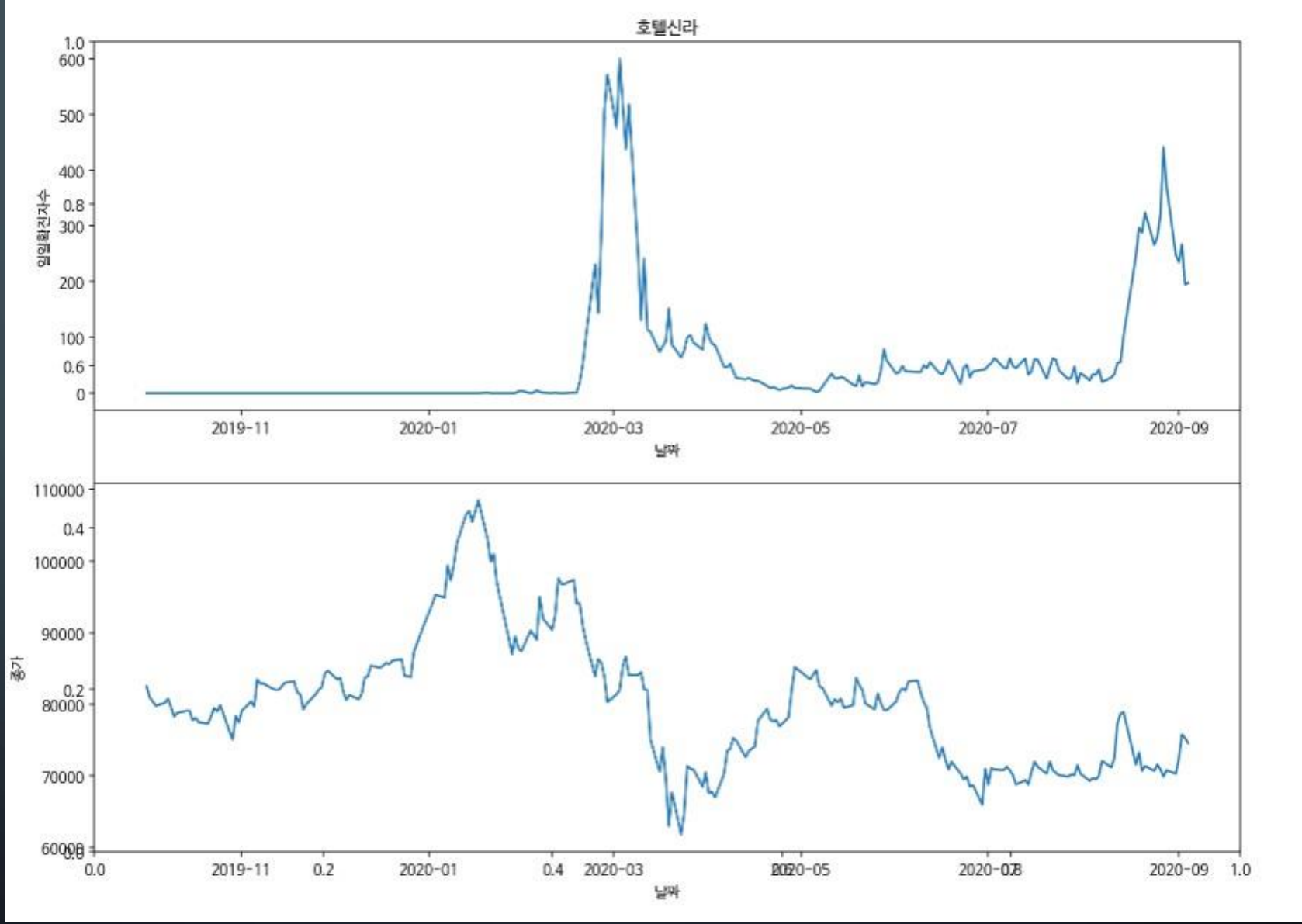
corr

	일일확진자수	Close
일일확진자수	1.000000	-0.206266
Close	-0.206266	1.000000

```
corr = result2_na.corr(method='pearson')
```

corr

	일일확진자수	Close
일일확진자수	1.000000	0.122976
Close	0.122976	1.000000



선형회귀모델

데이터 스케일링 :

```
from sklearn.preprocessing import RobustScaler
```

모델링&시각화 :

```
from sklearn.linear_model import LinearRegression  
import matplotlib  
import matplotlib.pyplot
```

선형회귀모델

```
In [28]: X = result2_na[['일일확진자수']]  
         Y = result2_na[['Close']]
```

```
In [29]: # 스케일링, 데이터가 부족해서 train, test set 나누지 않았음.  
         from sklearn.preprocessing import RobustScaler  
         scaler = RobustScaler()  
         scaler.fit(X)  
         scaler.fit(Y)  
         X = scaler.fit_transform(X)  
         Y = scaler.fit_transform(Y)
```

```
In [30]: # 모델링
import numpy as np
from sklearn.linear_model import LinearRegression
```

```
In [31]: model = LinearRegression()
```

```
In [32]: # 학습과 예측
model.fit(X, Y)
model.predict(X)
Y_pred = model.predict(X)
```

```
In [33]: # 시각화
import matplotlib as mpl
import matplotlib.pyplot as plt

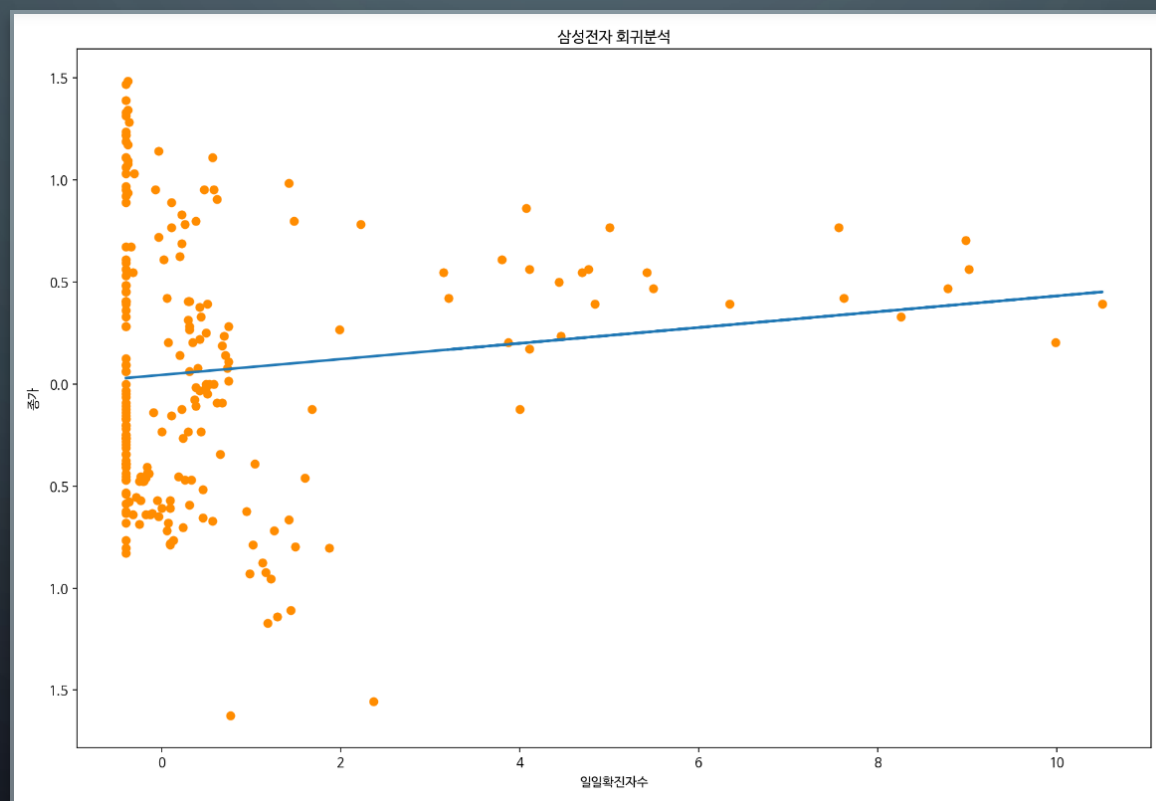
# 한글
%config InlineBackend.figure_format = 'retina'

!apt -qq -y install fonts-nanum

import matplotlib.font_manager as fm
fontpath = '/usr/share/fonts/truetype/nanum/NanumBarunGothic.ttf'
font = fm.FontProperties(fname=fontpath, size=9)
plt.rc('font', family='NanumBarunGothic')
mpl.font_manager._rebuild()
```

```
In [34]: # 시각화
plt.rcParams["figure.figsize"] = (15,10)
plt.scatter(X, Y, c="darkorange")
plt.plot(X, Y_pred, label="linear", linewidth=2)
plt.title('삼성전자 회귀분석')
plt.xlabel('일일확진자수')
plt.ylabel('종가')
plt.show()
```

```
/usr/local/lib/python3.6/dist-packages/matplotlib/backends/backend_agg.py:214: RuntimeWarning: Glyph 8722 missing from current font.
  font.set_text(s, 0.0, flags=flags)
/usr/local/lib/python3.6/dist-packages/matplotlib/backends/backend_agg.py:183: RuntimeWarning: Glyph 8722 missing from current font.
  font.set_text(s, 0, flags=flags)
```



```
In [28]: X = result2_na[['일일확진자수']]  
         Y = result2_na[['Close']]
```

```
In [29]: # 스케일링, 데이터가 부족해서 train, test set 나누지 않았음.  
         from sklearn.preprocessing import RobustScaler  
         scaler = RobustScaler()  
         scaler.fit(X)  
         scaler.fit(Y)  
         X = scaler.fit_transform(X)  
         Y = scaler.fit_transform(Y)
```



```
In [30]: # 모델링
import numpy as np
from sklearn.linear_model import LinearRegression
```

```
In [31]: model = LinearRegression()
```

```
In [32]: # 학습과 예측
model.fit(X, Y)
model.predict(X)
Y_pred = model.predict(X)
```

```
In [33]: # 시각화
import matplotlib as mpl
import matplotlib.pyplot as plt

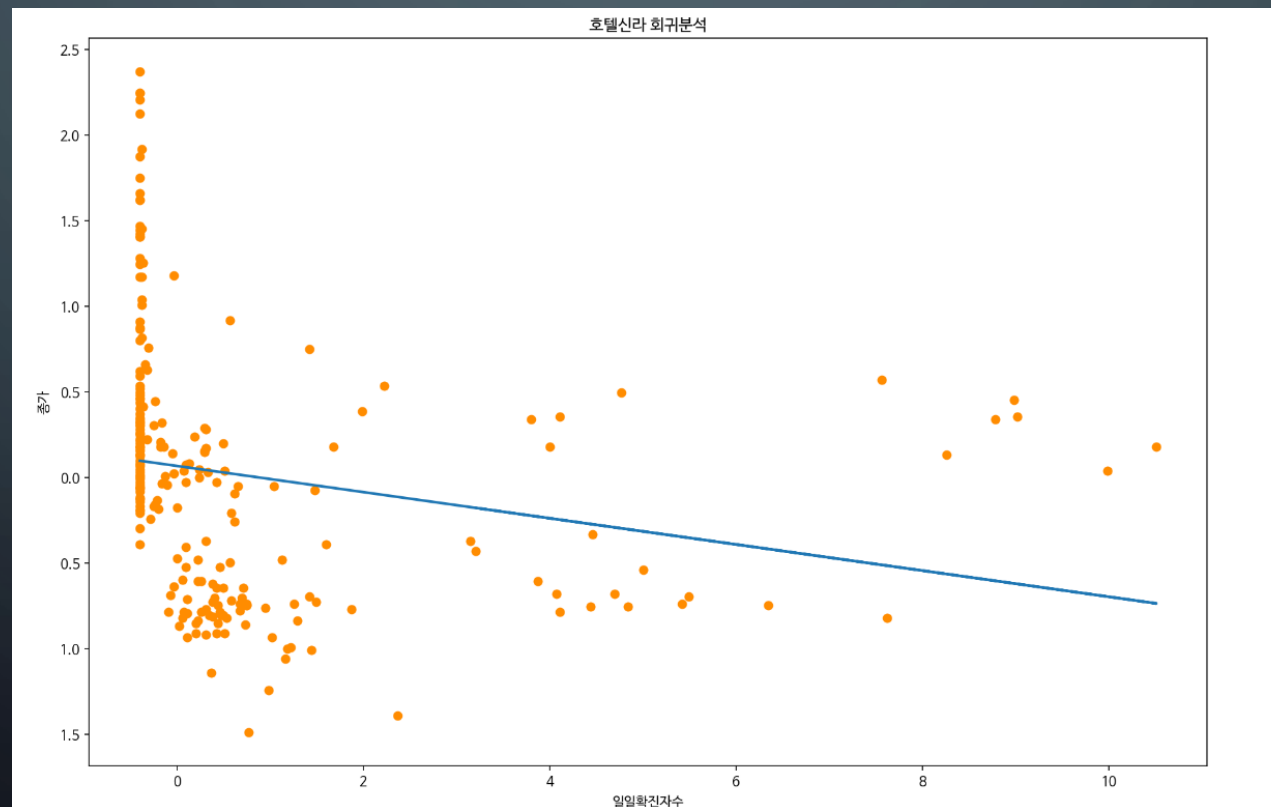
# 한글
%config InlineBackend.figure_format = 'retina'

!apt -qq -y install fonts-nanum

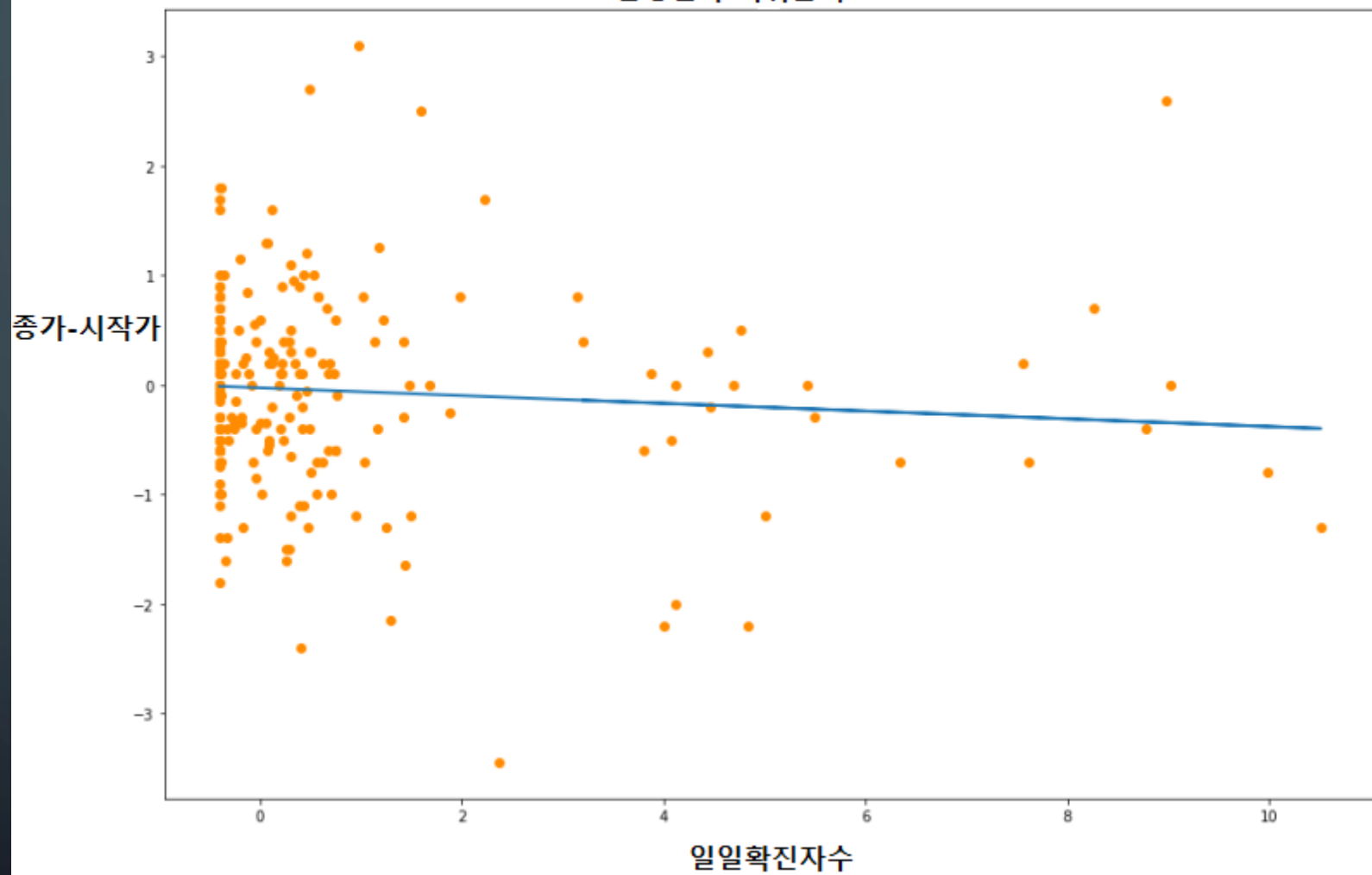
import matplotlib.font_manager as fm
fontpath = '/usr/share/fonts/truetype/nanum/NanumBarunGothic.ttf'
font = fm.FontProperties(fname=fontpath, size=9)
plt.rc('font', family='NanumBarunGothic')
mpl.font_manager._rebuild()
```

```
In [34]: # 시각화
plt.rcParams["figure.figsize"] = (15,10)
plt.scatter(X, Y, c="darkorange")
plt.plot(X, Y_pred, label="linear", linewidth=2)
plt.title('호텔신라 회귀분석')
plt.xlabel('일일확진자수')
plt.ylabel('종가')
plt.show()
```

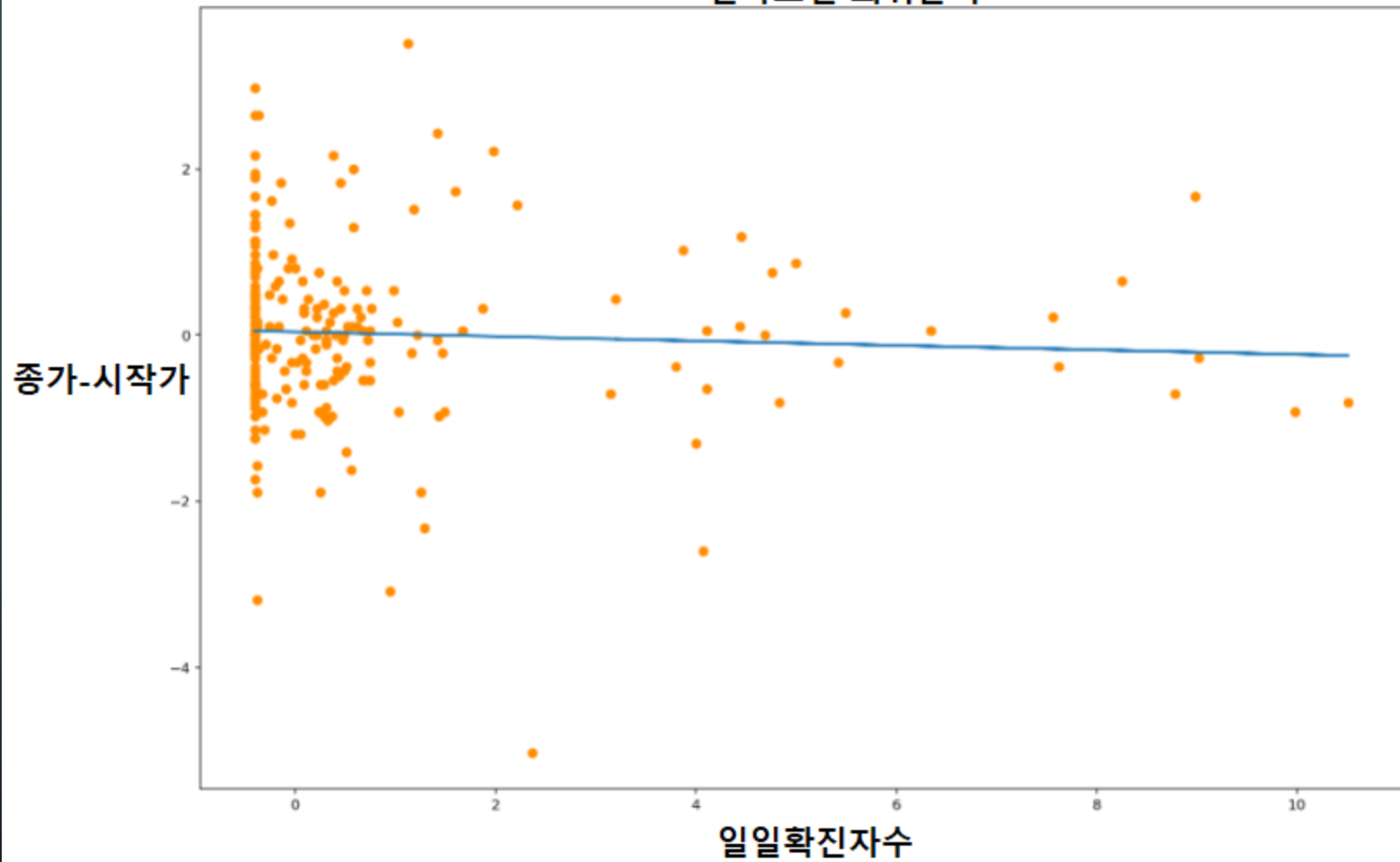
```
/usr/local/lib/python3.6/dist-packages/matplotlib/backends/backend_agg.py:214: RuntimeWarning: Glyph 8722 missing from current font.
  font.set_text(s, 0.0, flags=flags)
/usr/local/lib/python3.6/dist-packages/matplotlib/backends/backend_agg.py:183: RuntimeWarning: Glyph 8722 missing from current font.
  font.set_text(s, 0, flags=flags)
```



삼성전자 회귀분석



신라호텔 회귀분석



기대효과

1. 다른 주식에 대한 확진자 수의 영향력 시각화
2. 향후 확진자 수에 따른 주가 동향 통계적인 예측가능
3. 투자의 방향성 판단에 도움
ex. 안정성, 수익성, 장기성, 단기성

The background is a dark blue gradient with a faint, large circular pattern in the center. In the four corners, there are white line art illustrations of circuit boards or neural networks, featuring lines and small circles.

총평

The image features a dark blue gradient background with faint, stylized circuit board traces in the corners. These traces are composed of thin white lines and small circles, resembling electronic components or data paths. The central focus is the Korean text '감사합니다' (Thank you), rendered in a clean, white, sans-serif font.

감사합니다