

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

상관분석

#01. 작업준비

패키지 가져오기

```
import numpy as np
from pandas import read_excel
from scipy import stats

import sys
import seaborn as sb
from matplotlib import pyplot as plt
```

예제 (1)

언어영역 점수와 수리영역 점수의 상관관계

1. 데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E03/point.xlsx")
df
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

	언어영역	수리영역
0	392	428
1	404	518
2	346	376
3	382	426
4	368	409
5	449	490
6	391	448

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관정도 확인

```
plt.rcParams["font.family"] = 'AppleGothic' if sys.platform == 'darwin'
plt.rcParams["font.size"] = 12
plt.rcParams["figure.figsize"] = (10, 5)
plt.rcParams["axes.unicode_minus"] = False
```

```
#sb.scatterplot(data=df, x='언어영역', y='수리영역')
sb.regplot(data=df, x='언어영역', y='수리영역')
plt.grid()
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

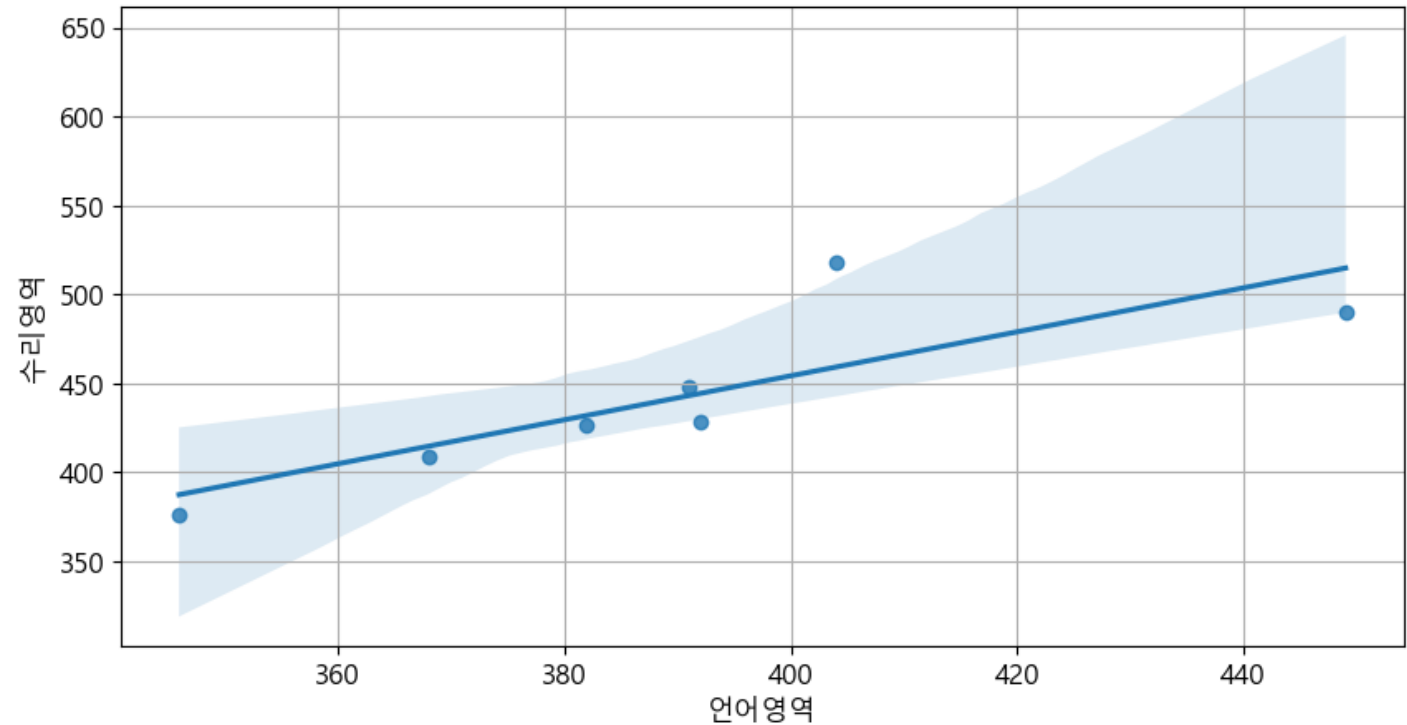
상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

```
plt.show()
plt.close()
```



3. 공분산

공분산은 상관관계의 상승 혹은 하강하는 경향을 이해할 수는 있으나 2개 변수의 측정 단위의 크기에 따라 값이 달라지므로 절대적 정도를 파악하기에는 한계가 있다.

결과값이 양수면 정비레, 음수면 반비례

그래서 잘 안쓴다.

```
np.cov(df['언어영역'], df['수리영역'])[0, 1]
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

1271.9523809523807

4. 상관분석

numpy를 통한 상관계수 구하기

p-value 는 알 수 없기 때문에 분석 결과를 검정할 수 없다.

기본적으로 상관행렬을 표시한다.

```
np.corrcoef(df['언어영역'], df['수리영역'])
```

```
array([[1.          , 0.82106421],
       [0.82106421, 1.          ]])
```

상관계수를 얻기 위해서는 0,1 혹은 1,0 번째 항목을 취한다.

```
np.corrcoef(df['언어영역'], df['수리영역'])[0,1]
```

0.821064207442306

피어슨 상관계수와 p-value 구하기

```
r, p = stats.pearsonr(df['언어영역'], df['수리영역'])
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기
2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산
4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

`"상관계수: %f, p-value: %f" % (r, p)``'상관계수: 0.821064, p-value: 0.023564'`

결과 해석

`p-value`가 유의수준 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 채택한다. 그러므로 언
어영역 점수와 수리영역 점수는 상관이 있다.

데이터프레임 객체를 통한 상관계수 구하기

numpy를 통해 구한 결과와 동일하다.

이와 같이 표시되는 형태를 상관행렬이라고 한다.

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

	언어영역	수리영역
언어영역	1.000000	0.821064
수리영역	0.821064	1.000000

상관분석 결과 시각화

`plt.rcParams["figure.figsize"] = (5,3)``sb.heatmap(corr,`

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

02_상관분석 (1).ipynb

```

annot = True, #실제 값 화면에 나타내기
cmap = 'Greens', #색상
vmin = -1, vmax=1, #컬러차트 영역 -1 ~ +1
linewidths=.5 #구분선

```

)

```

plt.show()
plt.close()

```



예제 (2) - mtcars 데이터를 통한 실린더수와 연비의 상관관계 분석

R 패키지인 ggplot2에 내장된 샘플 데이터로 자동차 32종의 11개 속성에 대한 정보를 담고 있는 데이터

필드	설명
mpg	연료 효율성 (마일 당 갤런)

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

필드	설명
cyl	실린더 수
disp	배기량 (cubic inches)
hp	마력 (horsepower)
drat	후륜축 비율 (rear axle ratio)
wt	차량 무게 (1000 파운드)
qsec	1/4 마일 달성 시간 (초)
vs	V/S (V 엔진 형태인지, S는 직렬 엔진 형태인지를 나타냄)
am	변속기 형태 (0: 자동, 1: 수동)
gear	전진 기어 수 (전송 기어)
carb	기화기 개수 (carburetors)

데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E03/mtcars.xlsx", index_col='name')
df
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
name											
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	ca
name											
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

	mpg	cyl	displacement	hp	drat	wt	qsec	vs	am	gear	car
name											
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	ca
name											
Firebird											
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

실린더수에 따른 연비에 대한 산점도 그래프

```
plt.rcParams["figure.figsize"] = (10, 5)

sb.regplot(data=df, x='cyl', y='mpg')
plt.grid()
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

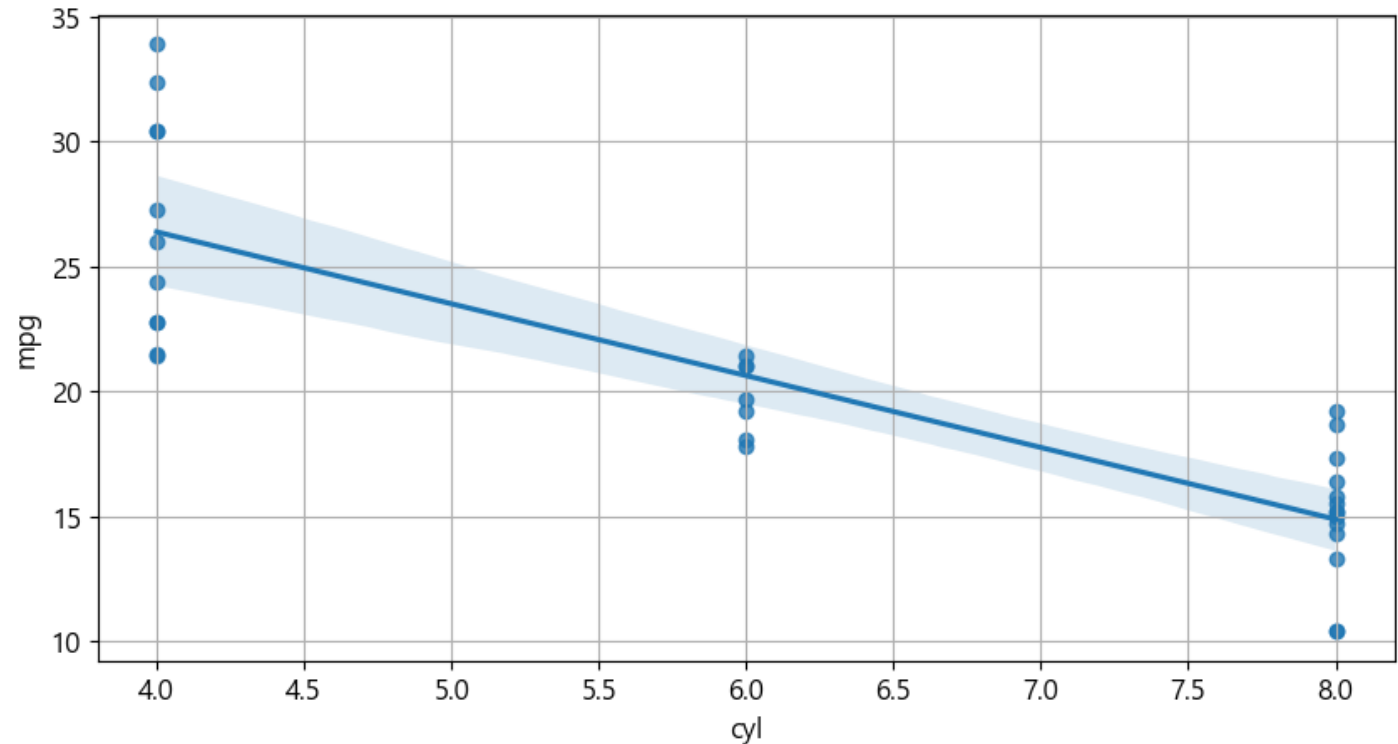
상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

```
plt.show()
plt.close()
```



실린더 수가 증가함에 따라 연비가 낮아지는 경향이 있는것으로 보임

상관분석

```
plt.rcParams["figure.figsize"] = (5,3)

corr = df.filter(['cyl', 'mpg']).corr(method='pearson')

sb.heatmap(corr,
            annot = True, #실제 값 화면에 나타내기
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

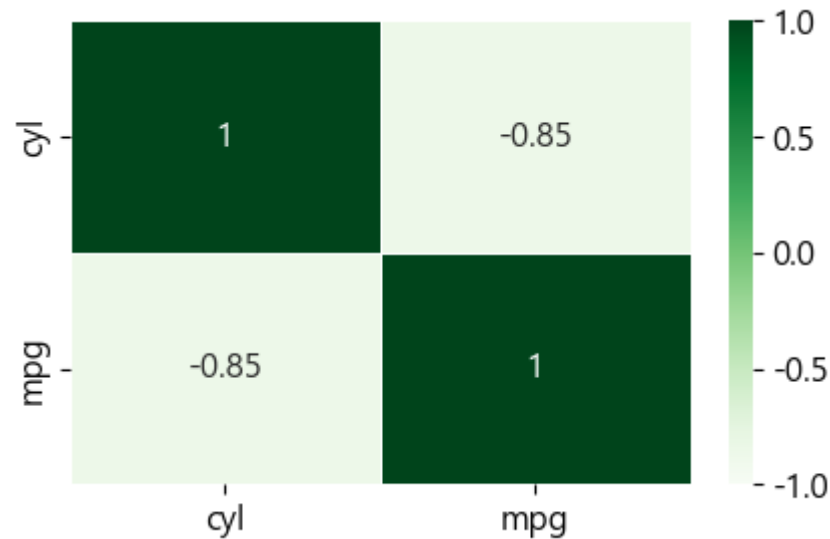
실린더수에 따른 연비에 대한

02_상관분석 (1).ipynb

```
cmap = 'Greens', #색상
vmin = -1, vmax=1, #컬러차트 영역 -1 ~ +1
linewidths=.5 #구분선
```

)

```
plt.show()
plt.close()
```



분석결과 상관계수가 **-0.85**로 나타나 실린더수와 연비는 강한 음의 상관관계가 있는 것으로 나타
남

분석결과 검정

```
stats.pearsonr(df['cyl'], df['mpg'])
```

상관분석

#01. 작업준비

패키지 가져오기

예제 (1)

1. 데이터 가져오기

2. 탐색적 데이터 분석

산점도 그래프를 통한 상관
정도 확인

3. 공분산

4. 상관분석

numpy를 통한 상관계수 구
하기피어슨 상관계수와 p-
value 구하기

결과 해석

데이터프레임 객체를 통한 상
관계수 구하기

상관분석 결과 시각화

예제 (2) - mtcars 데이터를 통한
실린더수와 연비의 상관관계 분
석

데이터 가져오기

실린더수에 따른 연비에 대한

```
PearsonRResult(statistic=-0.8521619594266132, pvalue=6.112687142580964e-
```

결과보고

32종의 자동차를 대상으로 실린더수에 따른 연비를 조사한 결과 실린더 수와 연비는 강한 음의 상관 관계를 갖는 것으로 나타났다. **p-value**가 유의수준 0.05보다 작으므로 이는 통계적으로 유의하다.