

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검토

스피어만 상관분석

스피어만 상관계수는 두 변수 간의 순위 순서를 기반으로 상관관계를 측정

순위 데이터 또는 서열 데이터에서 사용되며, 정확한 값보다 상대적인 순위가 중요한 경우에 사용

값 자체보다는 값의 상대적인 순위가 중요

자동차 성능 순위나 음악 차트 등의 데이터 유형

피어슨 상관계수보다는 이상치에 강건하며, 비선형적인 관계도 잘 캡처할 수 있음.

값의 범위는 -1에서 1까지이며, 피어슨 상관계수와 마찬가지로 1에 가까울수록 양의 상관관계, -1에 가까울수록 음의 상관관계를 의미하며, 0에 가까울수록 상관관계가 없음을 의미함

#01. 작업준비

패키지 참조

```
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
import sys
import os
```

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검증

```
sys.path.append(os.path.dirname(os.path.dirname(os.getcwd()))))
from helper import spearman_r
```

데이터 가져오기

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

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

데이터 전처리

종속변수(`mpg`)와 명목형 변수들을 추출함

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검정

```
df2 = df.filter(['mpg', 'vs', 'am', 'gear', 'carb'])
df2.head()
```

	mpg	vs	am	gear	carb
name					
Mazda RX4	21.0	0	1	4	4
Mazda RX4 Wag	21.0	0	1	4	4
Datsun 710	22.8	1	1	4	1
Hornet 4 Drive	21.4	1	0	3	1
Hornet Sportabout	18.7	0	0	3	2

#02. 산점도 행렬 시각화

그래프 초기화

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

산점도 행렬

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검정

명목형 변수에 따라 데이터가 몰려 있기 때문에 파악이 쉽지 않다.

```
sb.pairplot(df2)
plt.show()
plt.close()
```

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

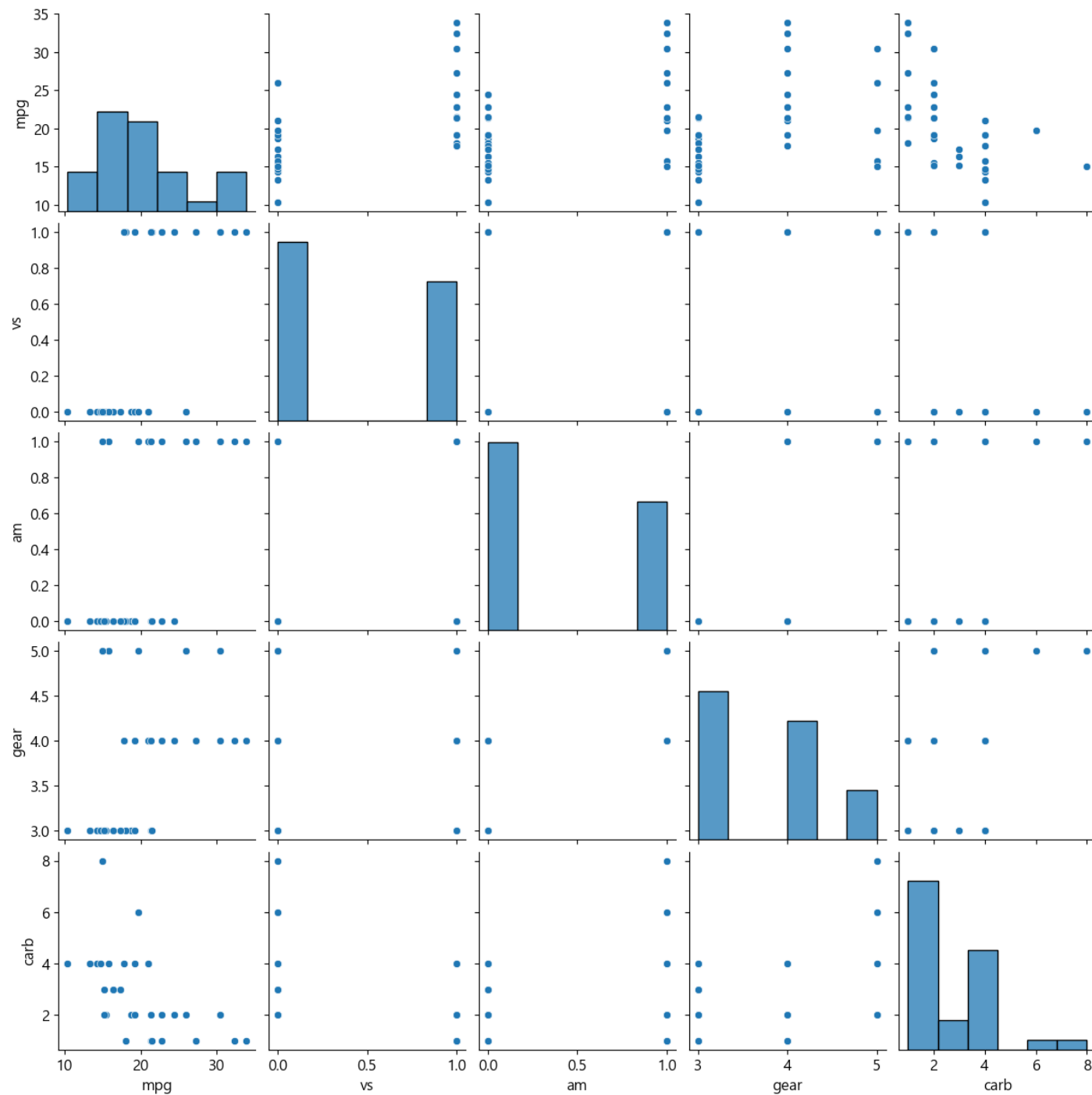
그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검정



스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검증

#02. 스피어만 상관분석

1. 상관행렬

```
df2.corr(method='spearman')
```

	mpg	vs	am	gear	carb
mpg	1.000000	0.706597	0.562006	0.542782	-0.657498
vs	0.706597	1.000000	0.168345	0.282662	-0.633695
am	0.562006	0.168345	1.000000	0.807688	-0.064365
gear	0.542782	0.282662	0.807688	1.000000	0.114887
carb	-0.657498	-0.633695	-0.064365	0.114887	1.000000

```
plt.rcParams["figure.figsize"] = (5,3)
sb.heatmap(df2.corr(method='spearman'), annot = True, cmap = 'Greens', v
plt.show()
plt.close()
```

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

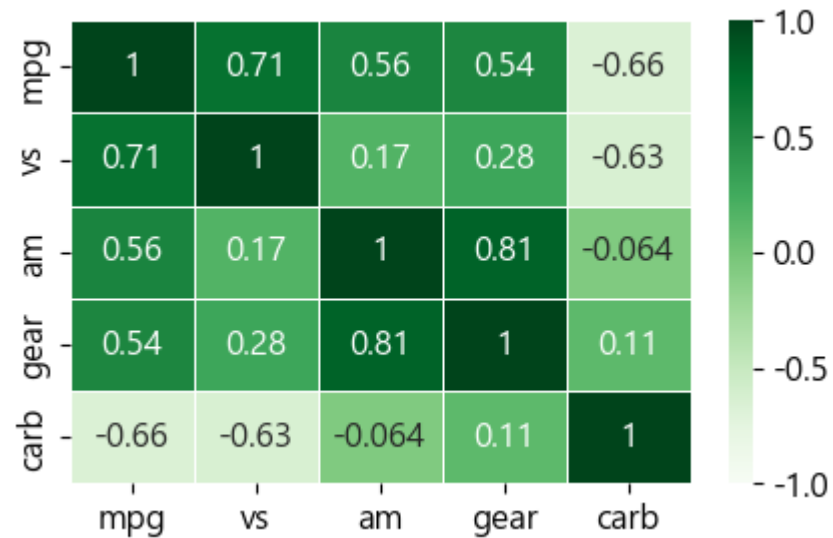
그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검정



2. 결과 검정

```
stats.spearmanr(df2['mpg'], df2['vs'])
```

```
SignificanceResult(statistic=0.7065967937356457, pvalue=6.19144980729084e-06)
```

```
spearman_r(df2)
```

	statistic	pvalue	result
fields			
mpg vs vs	0.706597	6.191450e-06	True

스피어만 상관분석

#01. 작업준비

패키지 참조

데이터 가져오기

데이터 전처리

#02. 산점도 행렬 시각화

그래프 초기화

산점도 행렬

#02. 스피어만 상관분석

1. 상관행렬

2. 결과 검정

	statistic	pvalue	result
fields			
vs vs am	0.168345	3.570439e-01	False
am vs gear	0.807688	2.304063e-08	True
gear vs carb	0.114887	5.312358e-01	False
carb vs mpg	-0.657498	4.337570e-05	True