

▼ Correlation

```
import warnings
warnings.filterwarnings('ignore')
```

▼ I. Load Data

- 키, 몸무게 데이터

```
import pandas as pd
```

```
url = 'https://raw.githubusercontent.com/takytaky/DATA/main/PII.csv'
```

```
DF = pd.read_csv(url)
```

```
DF.info()
```

```
DF.head()
```

▼ II. Covariance

▼ 1) 공분산

```
import numpy as np
```

```
np.cov(DF.Height, DF.Weight)[0][1]
```

▼ 2) Pearson 상관계수

```
np.cov(DF.Height, DF.Weight)[0][1] / (np.std(DF.Height, ddof = 1) * np.std(DF.Weight, ddof = 1))
```

▼ III. scipy

```
from scipy import stats
```

▼ 1) Karl Pearson 상관계수

- 기본적으로 등간척도/비율척도 변수에만 적용가능

```
stats.pearsonr(DF.Height, DF.Weight)[0]
```

▼ 2) spearman 상관계수

- 서열척도 변수가 포함되어도 적용가능
- 등간척도/비율척도 두 변수 간의 관계가 비선형적 일 때 적용

```
stats.spearmanr(DF.Height, DF.Weight)[0]
```

▼ 3) kendall tau

- spearman 상관계수와 같은 경우 적용가능
- 표본이 작을 때 spearman 상관계수보다 신뢰할 수 있음

```
stats.kendalltau(DF.Height, DF.Weight)[0]
```

▼ IV. pandas

▼ 1) Pearson 상관계수

- method('pearson', 'spearman', 'kendall')

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

▼ 2) Heat Map

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize = (9, 7))
sns.heatmap(DF.corr(),
            cbar = True,
            annot = True,
            annot_kws={'size' : 20},
            fmt = '.2f',
```

```
square = True,  
cmap = 'Blues')  
plt.show()
```

▼ V. numpy

```
import numpy as np
```

▼ 1) Pearson 상관계수

```
np.corrcoef(DF.Height, DF.Weight)
```

▼ VI. Visualization

```
plt.figure(figsize = (7, 5))  
sns.scatterplot(x = DF.Height, y = DF.Weight, s = 100)  
plt.axvline(DF.Height.mean(), color = 'r', linestyle = 'dashed', linewidth = 2)  
plt.axhline(DF.Weight.mean(), color = 'b', linestyle = 'dashed', linewidth = 2)  
plt.show()
```

#

#

#

The End

#

#

#

