

```
import numpy as np
from scipy.stats import *
import pandas as pd
```

▼ 산술 평균 계산

```
x = [1, 2, 3, 4, 5]
print(np.mean(x))
print(np.array(x).mean())
print(pd.Series(x).mean())
```

```
3.0
3.0
3.0
```

이진 변수에 대한 평균

```
x = [1, 0, 0, 0, 1]
print(np.mean(x))
```

```
0.4
```

▼ 조화 평균

```
x = np.array([0.1, 0.2, 0.3, 0.4, 0.5])
print(len(x) / np.sum(1/x))
print(hmean(x))
```

```
0.21897810218978103
0.21897810218978103
```

▼ 절사 평균이 필요한 이유

```
income = np.random.normal(2000000, 500000, 100) # 평균이 2백만원이고 표준편차가 50만원
print(np.mean(income))
```

```
1931382.8277989388
```

```
income = np.append(income, 10**9) # 소득이 10억원인 사람의 등장
print(np.mean(income)) # 100명의 소득이 200만원에 가까운데, 한 명 때문에 대표값인 평균
```

```
11813250.324553404
```

▼ 절사 평균 계산

```
trim_mean(income, 0.2) # [20% ~ 80%]
```

▼ 최빈값 계산

```
x = np.random.choice(['A', 'B', 'C'], 1000) # A, B, C라는 요소로 구성
x[:10]
```

```
array(['B', 'C', 'A', 'C', 'C', 'B', 'B', 'A', 'B', 'A'], dtype='<U1')
```

```
#stats.mode(x) # [0]: 최빈값, [1]: 빈도
mode(x) # [0]: 최빈값, [1]: 빈도
```

```
ModeResult(mode=array(['B'], dtype='<U1'), count=array([358]))
```

```
pd.Series(x).value_counts()
```

```
B    358
C    331
A    311
dtype: int64
```

```
pd.Series(x).value_counts().index[0] # 맨 앞에 있는 것이 최빈 값
```

```
'B'
```

▼ 데이터의 대푯값

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

```
#np.random.seed(0)
data = np.random.randint(0, 5, 10)
print(data)
```

```
mean = np.mean(data); print("평균값 :", mean.round(2))
median = np.median(data); print("중앙값 :", median)
mode = stats.mode(data); print("최빈값 : {} ({}).format(mode[0][0], mode[1][0]))
```

```
[1 1 3 2 4 3 4 4 1 0]
평균값 : 2.3
중앙값 : 2.5
최빈값 : 1 (3)
```

▼ pandas로 대푯값 구하기

```
<class 'float'>  
평균값 : 49.82  
중앙값 : 50.0  
최빈값 : 0    90  
dtype: int64
```