

# 통계학실험 Chapter 2 Python 과제 및 답안

## 문제 1

pandas 패키지를 이용하여 cars93.csv 파일을 DataFrame으로 읽어들이고,

1. 차량의 Type 이 Midsize 인 데이터의 개수를 구하시오.
2. 휠베이스( Wheelbase )의 평균을 소수점 셋째자리까지 구하시오(Hint : round 함수를 이용).
3. Origin 이 USA 인 차량의 가격( Price )와 non-USA 인 차량의 가격( Price )의 평균을 구하고 비교하시오.

```
In [7]: import pandas as pd
import numpy as np

data = pd.read_csv('cars93.csv', index_col = 0)

# Q1
midsize = data.loc[data.Type == 'Midsize',:]
print(len(midsize))

# Q2
print(np.mean(data.Wheelbase).round(3))

# Q3
usa = data.loc[data.Origin=='USA', :]
non_usa = data.loc[data.Origin=='non-USA', :]

print(np.mean(usa.Price))
print(np.mean(non_usa.Price))

22
103.946
18.572916666666668
20.508888888888889
```

## 문제 2

seaborn 패키지의 iris 데이터프레임을 로드하자. for 문을 이용해 각 종( species )별 sepal\_length 의 평균과 표준편차를 round 함수를 이용하여 소수점 셋째자리까지 텍스트로 출력해보시오.

```
In [10]: import seaborn as sns
import numpy as np

data = sns.load_dataset('iris')
species = data.species.unique()

for s in species:
    mu = np.mean(data.loc[data.species == s, 'sepal_length'])
    sd = np.std(data.loc[data.species == s, 'sepal_length'])
    print(f"{s} : 평균 = {round(mu,3)}, 표준편차 = {round(sd,3)}")

setosa : 평균 = 5.006, 표준편차 = 0.349
versicolor : 평균 = 5.936, 표준편차 = 0.511
virginica : 평균 = 6.588, 표준편차 = 0.629
```

## 문제3

1. `seed` 를 10로 지정하고 1 이상 100 이하의 자연수에서 50개를 임의로 선택하여  $5 \times 10$  크기의 행렬을 생성하시오.
2. 1에서 생성된 행렬에서 2번째 행부터 5번째 행까지 추출하여 저장하시오.(\* 인덱스 주의!!!)
3. 1에서 생성된 행렬의 7번째 열의 평균과 분산을 계산하시오.

```
In [12]: import numpy as np

np.random.seed(10)
mat = np.random.randint(1, 101, size=(5,10))
mat

mat_24 = mat[1:5, :]
print(mat_24)

mat_7 = mat[:, 6]
print(np.mean(mat_7)) # 평균
print(np.var(mat_7)) # 분산

[[41 37 17 12 55 89 63 34 73 79]
 [50 52 55 78 70 14 26 14 93 87]
 [31 31 90 13 66 32 58 37 28 19]
 [94 78 23 24 95 12 29 75 89 10]]
41.2
252.56000000000003
```

## 문제 4

1. `seed` 를 30으로 고정하고 1 이상 1000 이하의 자연수에서 100개의 자연수를 복원추출하여 1차원 `array` 로 저장하자.
2. 1에서 저장한 `array` 의 각 원소에 대해 5의 배수이면 1, 그렇지 않으면 0을 가지는 새로운 `array` 를 만들고, 이 `array` 의 평균과 표준편차를 구하시오.

```
In [10]: import numpy as np
np.random.seed(30)
arr = np.random.randint(1, 1001, size=100)
arr2 = np.array([])
for i in arr:
    if i%3 == 0:
        arr2 = np.append(arr2, np.array([1]))
    else:
        arr2 = np.append(arr2, np.array([0]))

print(np.mean(arr2))
print(np.std(arr2))

0.33
0.47021271782034985
```

## 문제 5

다음의 지시에 따라 반복문(for), 조건문(if)을 이용하여 리스트 x와 x의 평균이 출력되는 코드를 작성하시오.

1. 빈 리스트 x를 생성하자.
2. 1부터 100까지 정수 중에서 10개를 랜덤하게 추출하여 a와 b에 저장한다.
3. a와 b의 값들의 크기를 서로 비교하여 크기가 큰 값들을 리스트 x에 저장(x.append() 이용)하고 그렇지 않으면 리스트 x에 0을 저장한다.
4. 리스트 x에 저장된 값들을 출력하고 이 값들의 평균을 구하여 출력한다.

```
In [1]: import numpy as np

x = []

a = np.random.choice(101, 10)
b = np.random.choice(101, 10)

for i in range(0, 10):
    if a[i]<b[i]:
        x.append(b[i])
    elif b[i] < a[i]:
        x.append(a[i])
    else:
        x.append(0)

print(x)
print(np.mean(x))

[82, 82, 77, 75, 65, 77, 0, 68, 89, 54]
66.9
```