

Chapter 3 - Exercise 4: Thao tác dữ liệu mảng dữ liệu baseball

Cho tập tin *baseball_2D.txt* => chép dữ liệu từ tập tin vào list là *baseball*

Dữ liệu *baseball* cho biết chiều cao (cột 1) tính theo inch và cân nặng (cột 2) tính theo pounds của các cầu thủ

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

```
In [2]: # dữ liệu baseball
baseball = [[74, 180], [74, 215], [72, 210], [72, 210], [73, 188], [69, 176], [69, 209],
```

```
In [3]: # Câu 1: Tạo một 2D numpy array tên np_baseball từ baseball.
np_baseball = np.array(baseball)
# Xem kiểu dữ liệu (type) của np_baseball
print(type(np_baseball))
# Xem kích thước (shape) của np_baseball
print(np_baseball.shape)
```

```
<class 'numpy.ndarray'>
(1015, 2)
```

```
In [4]: # Câu 2: In các giá trị của dòng thứ 50 trong np_baseball.
print(np_baseball[50-1])
```

```
[ 70 195]
```

```
In [5]: # Câu 3: Tạo một numpy array np_weight với dữ liệu được lấy từ cột hai của np_baseball.
np_weight = np_baseball[:,1]
# In danh sách các phần tử của np_weight.
print(np_weight)
```

```
[180 215 210 ... 205 190 195]
```

```
In [6]: # Câu 4: Cho biết chiều cao của vận động viên thứ 124, và in ra kết quả
print(np_baseball[124-1,0])
```

```
75
```

```
In [7]: # Câu 5: Cho biết chiều cao trung bình, cân nặng trung bình của các cầu thủ
# code here
```

Nhấn vào đây để xem kết quả!

Chiều cao trung bình: 73.6896551724138

Cân nặng trung bình: 201.34876847290641


```
In [8]: # Câu 6: Bạn nhận xét gì về mối tương quan giữa chiều cao và cân nặng của các cầu thủ:  
# Có/ không có tương quan, tương quan thuận/nghịch  
# Tính hệ số tương quan  
x = np_baseball[:,0]  
y = np_baseball[:,1]  
np.corrcoef(x,y)
```

```
Out[8]: array([[1.          , 0.53153932],  
               [0.53153932, 1.          ]])
```

Kết luận: Cân nặng và chiều cao của các cầu thủ có mối tương quan thuận. (hệ số tương quan: 0.53153932)

In []:

