

Pyladies - Numpy Pandas - odpowiedzi

April 3, 2017

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

```
In [15]: # array([[ 1,  2,  3,  4,  5], [ 6,  7,  8,  9, 10], [11, 12, 13, 14, 15], [16, 17, 18, 19, 20], [21, 22, 23, 24, 25]])

a = np.arange(1,26).reshape(5,5)
a
```

```
Out[15]: array([[ 1,  2,  3,  4,  5],
                [ 6,  7,  8,  9, 10],
                [11, 12, 13, 14, 15],
                [16, 17, 18, 19, 20],
                [21, 22, 23, 24, 25]])
```

```
In [7]: # 1) array([[12, 13, 14, 15], [17, 18, 19, 20], [22, 23, 24, 25]])
a[2:,1:]
```

```
Out[7]: array([[12, 13, 14, 15],
                [17, 18, 19, 20],
                [22, 23, 24, 25]])
```

```
In [8]: # 2) 20
a[3, 4]
```

```
Out[8]: 20
```

```
In [11]: # 3) array([[ 2], [ 7], [12]])
a[:3, 1]
```

```
Out[11]: array([ 2,  7, 12])
```

```
In [12]: # 4) sum wszystkich elementów w macierzy
a.sum()
```

```
Out[12]: 325
```

```
In [14]: # 5) sum dla każdej kolumny w macierzy
a[:].sum(axis=0)
```

```
Out[14]: array([55, 60, 65, 70, 75])
```

```
In [19]: # wiczenie 2
# zaimportuj dane z pliku
# https://mdcune.psych.ucla.edu/modules/bioinformatics/extras/QTL_Sample_data.xls/view
import pandas as pd

data = pd.read_excel('QTL_Sample_data.xls', sheet_name='Sheet1')

In [20]: # sprawdź struktur tabeli
data.head()
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 135 entries, 0 to 134
Data columns (total 13 columns):
ID                135 non-null int64
Point             135 non-null float64
Strain            135 non-null object
sex              135 non-null object
sex#             135 non-null int64
age              135 non-null int64
bodywt           135 non-null float64
brainwt          135 non-null float64
MedUNshOB        135 non-null float64
Res1_sex         135 non-null float64
Res2_sex-age     135 non-null float64
Res3_sex-age-bw  135 non-null float64
Res4_sex-age-bw-brnw 135 non-null float64
dtypes: float64(8), int64(3), object(2)
memory usage: 13.8+ KB
```

```
In [21]: # sprawdź jaki jest redni wiek (age)
data['age'].mean()
```

```
Out[21]: 87.362962962962968
```

```
In [22]: # sprawdź jaka jest najwysza warto brainwt
data['brainwt'].max()
```

```
Out[22]: 551.5
```

```
In [24]: # sprawdź jaka jest płeć (sex) myszy o ID 1709
data[data['ID']==1709]['sex']
```

```
Out[24]: 3      M
Name: sex, dtype: object
```

```
In [28]: # sprawdź jakie ID ma mysz o najwyższej wartości bodywt
data[data['bodywt']== data['bodywt'].max()]['ID']
# lub
# data.loc[data['bodywt'].idxmax()]['ID']
```

```
Out[28]: 232
```

```
In [33]: # sprawdź ile jest myszy pci mskiej i eskiej
data.groupby('sex').count()
```

```
Out[33]:
```

	ID	Point	Strain	sex#	age	bodywt	brainwt	MedUNshOB	Res1_sex	\
sex										
F	37	37	37	37	37	37	37	37	37	
F	27	27	27	27	27	27	27	27	27	
M	40	40	40	40	40	40	40	40	40	
M	31	31	31	31	31	31	31	31	31	

	Res2_sex-age	Res3_sex-age-bw	Res4_sex-age-bw-brnw
sex			
F	37	37	37
F	27	27	27
M	40	40	40
M	31	31	31

```
In [36]: data['sex'] = data['sex'].str.strip()
data.groupby('sex').count()
```

```
Out[36]:
```

	ID	Point	Strain	sex#	age	bodywt	brainwt	MedUNshOB	Res1_sex	\
sex										
F	64	64	64	64	64	64	64	64	64	
M	71	71	71	71	71	71	71	71	71	

	Res2_sex-age	Res3_sex-age-bw	Res4_sex-age-bw-brnw
sex			
F	64	64	64
M	71	71	71

```
In [41]: # Sprawdź ile jest rodzajów linii myszy (Strain)
len(data['Strain'].unique())
```

```
Out[41]: 38
```

```
In [42]: #jest to ta funkcja
data['Strain'].nunique()
```

```
Out[42]: 38
```

```
In [38]: # Sprawdź jaka jest średnia wartość brainwt i bodywt odpowiednio dla myszy eskich i mskich
data.groupby('sex').mean()[['bodywt', 'brainwt']]
```

```
Out[38]:
```

	bodywt	brainwt
sex		
F	20.357187	422.742188
M	22.194225	420.369014

```
In [39]: # czy różnice s istotne statystyczne?  
from scipy.stats import ttest_ind
```

```
female = data[data['sex']=='F']  
male = data[data['sex']=='M']
```

```
ttest_ind(female['bodywt'], male['bodywt'])
```

```
Out[39]: Ttest_indResult(statistic=-2.0311889534264949, pvalue=0.044228322810162403)
```

```
In [40]: ttest_ind(female['brainwt'], male['brainwt'])
```

```
Out[40]: Ttest_indResult(statistic=0.34546988237361048, pvalue=0.73028695097011398)
```

```
In [ ]:
```