

/opt/pyenv-3.7.5/bin/jupyter notebook

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

Δημιουργία τυχαίου δείγματος 1024 στοιχείων από την $\mathcal{N}(2, 5)$

```
In [ ]: dataset = pd.DataFrame(5*np.random.randn(1024)+2, columns=['x'])
```

```
In [ ]: dataset.head()
```

Υπολογισμός του συντελεστή συμμετρίας Fisher-Pearson

$$G_1 = \frac{N^2}{(N-1)(N-2)} \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \bar{X})^3}{s^3}$$

```
In [ ]: dataset.skew()
```

Με τη len() λαμβάνουμε πόσα στοιχεία έχει το σύνολο δεδομένων

```
In [ ]: N = len(dataset)
```

Υπολογισμός του $\sum_{n=1}^N (x_n - \bar{X})^3$

```
In [ ]: dataset - dataset.mean()
```

```
In [ ]: sum3 = ((dataset - dataset.mean())**3).sum()
```

```
In [ ]: s = dataset.std()
```

```
In [ ]: G1 = N**2 / ((N-1)*(N-2)) * (1/N) * sum3 / s**3
```

```
In [ ]: G1
```

Υπολογισμός του μέτρου κύρτωσης

$$\text{kurt} = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \bar{X})^4}{s^4} - 3$$

```
In [ ]: dataset.kurt()
```

Άσκηση : Υπολογισμός του $\sum_{n=1}^N (x_n - \bar{X})^4$

```
In [ ]: sum4 = [?]??
```

```
In [ ]: sum4
```

```
In [ ]: kurt = (1/N * sum4)/s**4 - 3
```

```
In [ ]: kurt
```

Άσκηση : Υπολογίστε τη μέση τιμή, τη διάμεσο, τα ποσοστημόρια για το dataset

Άσκηση : Δημιουργήστε το Box-and-Whisker plot για το dataset

scipy.stats

```
In [ ]: import scipy.stats as st
```

z-scores

$$z = \frac{x - \bar{X}}{s}$$

```
In [ ]: x1 = 17.0
```

```
In [ ]: z1 = (x1 - dataset.mean())/dataset.std()
```

```
In [ ]: z1
```

Υπολογισμός του $P(Z \leq z_1)$ με την scipy.stats

```
In [ ]: P1 = st.norm.cdf(z1)
```

```
In [ ]: P1
```

Άσκηση : Επιλύστε με χρήση των pandas και scipy.stats

Σε ένα δημοψήφισμα με απάντηση ΝΑΙ / ΟΧΙ, το ΟΧΙ συγκέντρωσε το 60% των συνολικών ψήφων. Ποιά η πιθανότητα σε ένα τυχαίο δείγμα 500 ψηφοφόρων το ΝΑΙ να συγκεντρώνει τουλάχιστον το 50%;

In []: