

Entrée [1]:

```
import numpy as np
import scipy.stats as st

import matplotlib.pyplot as plt
```

Exercice 1

Entrée [2]:

```
sommeil = [7.5, 8.2, 6.8, 7.1, 6.5, 7.9, 8.0, 6.3, 7.7, 7.0, 6.6, 8.5, 7.2, 6.9, 7.8, 8.
```

Entrée [3]:

```
mean = np.mean(sommeil)
median = np.median(sommeil)
std = np.std(sommeil)
variance = std**2
rang = np.max(sommeil) - np.min(sommeil)
q1, q3 = np.quantile(sommeil, [0.25, 0.75])
```

Entrée [4]:

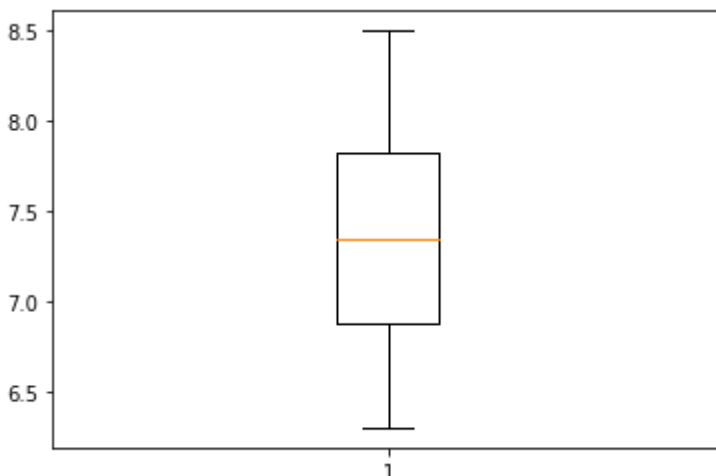
```
print(f"Moyenne : {mean:.3}, \nMédiane : {median}, \nEcart-type : {std:.3}, \nVariance :
```

```

Moyenne : 7.35,
Médiane : 7.35,
Ecart-type : 0.603,
Variance : 0.363,
Etendue : 2.2,
1er quartile : 6.875,
3e quartile : 7.825
```

Entrée [5]:

```
plt.boxplot(sommeil)
plt.show()
```



Exercice 2

Situation 1 : Courbe / Lineplot

Situation 2 : Diagramme en bâtons / Barchart

Situation 3 : Histogramme

Situation 4 : Nuage de points / Scatterplot

Situation 5 : Histogramme

Exercice 3

Situation 1 : On privilégie la médiane car quelques valeurs extrêmes risquent de biaiser la perception.

Situation 2 :

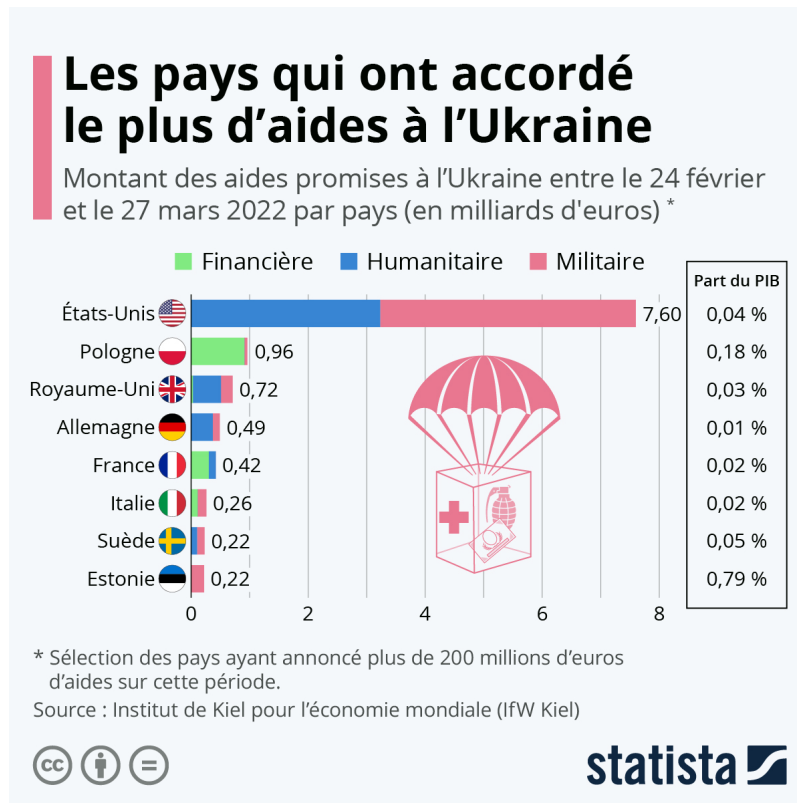
- Q1 : Moyenne car la distribution a toutes les chances d'être symétrique
- Q2 : Asymétrie à droite (right-skewed) : la moyenne va surestimer le temps de réaction de la plupart des athlètes
- Q3 : Dispersion des données (variance, IQR), corrélation avec autres variables

Exercice 4

Q1 : Biais de référence. La recherche peut représenter 20 ou 30% du budget d'un département sans que cela ne représente un montant conséquent s'il s'agit d'un département à faible budget total. Ex : 10% du budget de Paris représentera bien plus que 90% du budget de la Creuse.

Q2 : Biais d'échelle. Le département qui fournit 500k€ à la recherche consacre peut-être un montant bien plus important au regard de son budget total.

→ Parallèle Pologne/USA dans l'aide fournie à l'Ukraine



Exercice 5

Q1 :

- ☒ Corrélation entre temps d'étude et performances académiques
- ☒ Temps d'étude antérieur aux performances académiques (examens en fin d'année)
- ☒ Toutes choses égales par ailleurs → à démontrer

Les informations actuelles ne sont pas suffisantes pour établir la causalité.

Q2 : autres facteurs

- heures de colle, parents pris donc étude le soir, QI, redoublant, CSP des parents, temps de sommeil ...

Q3 : faire des paires d'étudiants ayant exactement le même profil hormis leur temps d'étude et analyser l'influence sur la réussite académique

Exercice 6

Il y a deux hôpitaux dans une ville, Hôpital A et Hôpital B.

Hôpital A :

- Département enfant : taux de survivabilité : 30 % (sur 100 patients)
- Département adulte : taux de survivabilité : 96.66668 % (sur 900 patients)

Hôpital B :

- Département enfant : taux de survivabilité : 52.5% (sur 400 patients)
- Département adulte : taux de survivabilité : 98.3 % (sur 600 patients)

Question 1 : Si tu analyses les taux de survie de chaque hôpital séparément, lequel semble avoir de meilleurs résultats dans chaque groupe ?

Question 2 : Si tu combines les résultats des deux départements dans chaque hôpital, qu'observe-tu concernant les taux de survie globaux dans les groupes A et B ?

Question 3 : Comment le paradoxe de Simpson se manifeste-t-il dans cette situation ? Quelle conclusion importante peut-on en tirer ?

Q1:

L'hôpital B est le meilleur hôpital, avec un taux de réussite plus élevé dans chaque département par rapport à l'hôpital A.

Q2 :

- Hôpital A :

$$\frac{0.3 \times 100 + 0.96666668 \times 900}{100 + 900} = 90\%$$

- Hôpital B :

$$\frac{0.525 \times 400 + 0.983 \times 600}{400 + 600} = 80\%$$

L'hôpital A a un meilleur taux de réussite global et semble mieux que l'hôpital B.

Entrée []:

Entrée []: