

Feuille_info_20 : Statistiques univariées.

Trois fichiers Python vous sont fournis :

- `feuille_20_1.py`
- `feuille_20_2.py`
- `feuille_20_3.py`

et des fichiers de données réelles :

- `diagonale_des_fous_temps_2022.txt`
- `diagonale_des_fous_temps_2023.txt`
- `diagonale_des_fous_temps_2024.txt`
- `diagonale_des_fous_temps_2025.txt`

Commencer par télécharger tous ces fichiers dans un même dossier.

Vous devez :

- compléter certaines parties de code ,
- commenter ce qui vous semble nécessaire, préciser ce que font les différentes fonctions,
- exécuter les programmes et reporter vos observations dans les tableaux ci-dessous.

L'objectif de ce TD n'est pas de programmer beaucoup, mais de :

- savoir décrire une série statistique avec la moyenne, l'écart-type et la courbe des fréquences cumulées.
- déterminer un intervalle couvrant 95% des données.
- observer le comportement de la moyenne d'un grand nombre d'échantillons et mettre en évidence l'apparition du facteur $\frac{1}{\sqrt{n}}$

1) Lois simulées (`feuille_20_1.py`)

- a) Compléter les premières fonctions.
- b) Exécuter le programme et compléter le tableau ci-dessous.

Loi simulée	Moyenne $\bar{x}$	Écart-type s	Proportion dans $[\bar{x} \pm 2s]$
Normale $\mathcal{N}(0, 1)$			
Uniforme $[-1, 1]$			
Binomiale $\mathcal{B}(100, 0,2)$			
baroque			

A-t-on toujours 95% des résultats dans $[\bar{x} \pm 2s]$?

2) Données réelles (`feuille_20_2.py`)

- a) Pour chaque année, observer la figure obtenue et compléter les tableaux.

Année	$\bar{x}$ (h)	s (h)	Proportion dans $[\bar{x} \pm 2s]$
2022			
2023			
2024			
2025			

Année	Longueur $[\bar{x} \pm 2s]$ (h)	Longueur intervalle minimal 95% (h)
2022		
2023		
2024		
2025		

- b) Quel intervalle est le plus court ?
- c) La moyenne semble-t-elle être au « centre » des données ?
- d) Que peut-on dire de l'utilisation de $\bar{x}$ et s pour décrire ces données ?

e) À l'aide de la courbe des fréquences cumulées de chaque année, compléter le tableau suivant.

Sans programmer juste par lecture graphique.

	2022	2023	2024	2025
Proportion de coureurs en moins de 45 h				
Temps médian (50%)				
Temps le quel 25% des coureurs ont déjà terminé				
Temps le quel 75% des coureurs ont déjà terminé				
Proportion de coureurs ayant fait entre 50h et 60h				

3) Moyennes d'échantillons (feuille_20_3.py)

On fixe $n = 100$ et $N = 10000$. Pour chaque loi, on génère N échantillons de taille n . Pour chaque échantillon k , on calcule sa moyenne $\bar{x}_k$ et son écart-type s_k , puis on forme l'intervalle

$$I_k = \left[\bar{x}_k - 2 \frac{s_k}{\sqrt{n}}, \bar{x}_k + 2 \frac{s_k}{\sqrt{n}} \right].$$

a) Pour chaque loi, compléter le tableau suivant (indiquer la valeur de la moyenne théorique μ et la proportion d'intervalles I_k contenant μ) :

Loi	Moyenne théorique μ	Proportion des intervalles I_k contenant μ
Normale $\mathcal{N}(0, 1)$		
Uniforme $[-1, 1]$		
Binomiale $\mathcal{B}(100, 0.2)$		
Baroque		

b) Qu'observe-t-on ? (Comparer les valeurs obtenues entre les différentes lois.)

c) Donner une situation pratique où on peut utiliser ce type d'intervalle.

d) Application aux données de la Diagonale des Fous

On considère les temps de l'année 2022 stockés dans la liste `T_2022`.

On fixe $n = 100$ et $N = 10000$.

i. À l'aide de `[rd.choice(T_2022) for _ in range(n)]`, tirer N échantillons aléatoires de taille n dans la liste `T_2022`.

ii. Pour chaque échantillon k , calculer sa moyenne $\bar{x}_k$ et son écart-type s_k , puis former l'intervalle

$$I_k = \left[\bar{x}_k - 2 \frac{s_k}{\sqrt{n}}, \bar{x}_k + 2 \frac{s_k}{\sqrt{n}} \right].$$

iii. Calculer la proportion de ces intervalles contenant la moyenne globale

$$\mu = \text{moyenne de } T_{2022}.$$