

TP7 - STATISTIQUES À UNE VARIABLE

Dans tout le TP, on importe les modules suivants :

```
1 import numpy as np
import numpy.random as rd
import matplotlib.pyplot as plt
```

1 Statistiques à une variable

Exercice 1

★

Faire 10 000 tirages selon une loi uniforme sur $[1, 30]$. En utilisant la fonction `np.unique`, déterminer le nombre de modalités. Faire de même avec une loi de Poisson de paramètre 30 et une loi normale $\mathcal{N}(3, 4)$. Que constate-t-on ?

Exercice 2

★★

1. Récupérer le fichier `TP07_pop_communes_2019.csv` sur :
<https://cahier-de-prepa.fr/ecg2-fauriel/download?id=2095>
Enregistrer le fichier dans un dossier identifiable.
2. Utiliser la commande suivante :

```
1 stats = np.loadtxt("nom-du-fichier.csv")
```

où vous remplacerez par le chemin et le nom du fichier que vous avez choisi. Le contenu du fichier est ainsi chargé dans un tableau `numpy`. Ce fichier contient les résultats du dernier recensement en 2019. Chaque ligne du tableau correspond au nombre d'habitants d'une commune française. Il y a 34 990 lignes correspondant aux 34 990 communes en France métropolitaine lors du recensement.

3. (a) Quelle est la population totale lors du recensement ?
(b) En moyenne, combien y a-t-il d'habitant par communes ?
(c) Quelle est le plus petit nombre d'habitants d'une commune ?
(d) **Plus dur** : Combien de commune ont ce nombre d'habitants ?

Exercice 3

★★★

Tant que l'on pose des questions sur une seule modalité, on peut s'en sortir avec les techniques vues jusqu'à présent. Nous allons voir comment poser des questions plus générales.

1. **Un cas (plus) simple**

Exécuter le code suivant :

```
1 vals = np.array([1, 42, 112])
filtre = np.array([False, True, True])
print(vals[filtre])
```

Que se passe-t-il ? Que font les lignes précédentes ?

2. On continue de travailler sur la tableau `stats` de l'exercice précédent. Exécuter le code suivant :
`print(stats <= 100)`. À votre avis, à quoi correspond ce qui s'affiche ?
3. Que fait le code suivant :

```
1 filtre = stats <= 100
  stats_filtrees = stats[filtre]
  print(len(stats_filtrees))
```

- Combien de villes ont moins de 200 habitants ? Combien d'habitants totalisent-elles ?
- Plus dur** : Comment pourrait-on déterminer le nombre de villes ayant entre 100 et 200 habitants ?

2 Graphiques en Python

Exercice 4

★★

- Utiliser `rd.binomial` pour simuler 100 tirages d'une variables aléatoires selon une loi $\mathcal{B}(10, 1/2)$.
- Utiliser la commande `np.unique` pour en tirer les valeurs et les effectifs.
- À l'aide de `plt.plot`, tracer la courbe des fréquences cumulées.
- On suppose que les tirages ont été stockées dans la variable `tirages`. Exécuter le code suivant :

```
1 plt.hist(tirages)
  plt.show()
```

Que se passe-t-il ? Remplacer la première ligne par `plt.hist(tirages, 5)`. Que se passe-t-il ?

3 Divers

Exercice 5

★★★

Un résultat classique mentionne que si X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$ alors $P(\mu - \sigma \leq X \leq \mu) = P(\mu \leq X \leq \mu + \sigma) \simeq 0,34$. De même $P(\mu - 2\sigma \leq X \leq \mu - \sigma) = P(\mu + \sigma \leq X \leq \mu + 2\sigma) \simeq 0,135$.

- Faire une simulation avec 10 000 tirages selon une loi normale centrée réduite. Stocker le résultat dans un tableau.
- Utiliser la simulation pour estimer l'effectif des modalités entre -1 et 0 . A-t-on la valeur attendue ? Vérifier de même les autres valeurs données.
- Utiliser la simulation pour estimer $P(\mu - (k+1)\sigma \leq X \leq \mu - k\sigma)$ de manière générale.

Exercice 6

★★★

- Écrire une fonction qui renvoie la valeur de la probabilité que $X = k$ si X suit la loi $\mathcal{B}(10, 1/2)$.
- Remplir un tableau `numpy` avec les valeurs possibles de X .
- Remplir un second tableau avec le nombre de tirages attendus pour chacune des valeurs possibles de X si on fait 100 tirages.
- En réutilisant le tirage de l'exercice 4, superposer l'histogramme et la courbe théorique. Que constate-t-on ?

Exercice 7

★★

- En utilisant le fichier `TP07_pop_communes_2019.csv`, faire la courbe des fréquences cumulées du nombre d'habitants par villes.
- Tracer également un histogramme des fréquences suffisamment fin pour voir les tranches $[0, 99]$, $[100, 199]$, etc.