

T.D. n°5 : Statistiques univariées

Dans tout ce T.D., on utilisera :

```
import numpy as np
import numpy.random as rd
import matplotlib.pyplot as plt
import pandas as pd
```

1. En Python, on considère le code suivant :

```
p=rd.random()
x=rd.binomial(100,p,1000)
```

Avant d'écrire le code, expliquer ce que contient x. Vérifier ensuite.

2. a) Déterminer pour cette série statistique :

Minimum :	Médiane :	Maximum :
Etendue :	Moyenne : $\bar{x}$	Ecart-type :

b) On ajoute la ligne :

```
dfx=pd.DataFrame(x) #Permet de transformer la liste x en 'DataFrame' pour pandas
```

A l'aide des fonctions de la bibliothèque pandas, vérifier les résultats précédents et déterminer :

1 ^{er} quartile : $\bar{x}$	3 ^{ème} quartile : $\bar{x}$	Ecart-inter-quartile : $\bar{x}$
--------------------------------------	---------------------------------------	----------------------------------

c) Calculer $\frac{\bar{x}}{100}$ et comparer le à p. Conclusion ? (on dira que $\frac{\bar{x}}{100}$ est un estimateur de p)

d) A l'aide de la fonction np.sum, vérifiez le résultat de la moyenne (en plus : et de l'écart-type).

e) Représenter la boîte de Tukey de x.

3. a) On considère le script suivant :

```
effectifs=np.zeros(10)
for i in range(9):
    a=10*i
    b=10*(i+1)
    nb=sum((x >= a)and(x < b))
    effectifs[i]=nb
effectifs[9]=sum(x>=90)
```

Avant d'écrire ce script, expliquer précisément que contient le tableau effectifs.

b) A l'aide de np.cumsum, remplir le tableau suivant :

Valeurs de x	[0;10[	[10;20[	[20;30[	[30;40[	[40;50[	[50;60[	[60;70[	[70;80[	[80;90[	[90;100]
Nombre d'occurrences										
Effectifs cumulés										
Fréquences										
Fréquences cumulées										

c) Tracer l'histogramme correspondant à ce regroupement en classes.

Question en plus : Représenter les fréquences cumulées sur un graphique. Comment pourrait-on définir graphiquement la médiane ?