

P7 - Statistiques

Partie 1 - Statistiques "à la main"

On va commencer par analyser cinq séries de résultats à un examen (noté sur 100). Les résultats sont donnés dans les listes suivantes.

```
In [1]: groupeA = [47,78,50,73,55,39,62,41,80,50,58,54,51,57,62,77,72,68,47,46,81,69,28,63,49,44,
groupeB = [72,90,71,88,40,38,65,18,65,69,63,57,53,59,65,67,61,44,47,49,79,65,79,60,52,70,
groupeC = [69,76,58,69,88,46,54,69,67,70,55,58,66,82,53,71,33,0,69,60,39,61,49,73,76,65,2
groupeD = [64,40,66,44,74,62,69,40,74,50,75,57,88,52,58,59,45,58,63,61,42,71,73,58,65,68,
groupeE = [56,97,81,50,53,77,82,79,30,62,72,63,66,99,83,87,70,42,100,66,75,81,94,50,80,66
```

1. Moyenne - Écart-type

1. Créer une fonction moyenne(liste) calculant la moyenne des valeurs d'une liste, puis calculer la moyenne de chacun des cinq groupes.

```
In [2]: def moyenne(liste):
S = 0
for x in liste:
S = S + x
return S/len(liste)
```

```
In [3]: for liste in [groupeA, groupeB, groupeC, groupeD, groupeE]:
print(moyenne(liste))
```

```
60.98876404494382
63.45263157894737
64.46394984326018
67.36259541984732
73.82142857142857
```

2. Calculer l'écart-type des groupes.

```
In [4]: import numpy as np
def ecarttype(liste):
m = moyenne(liste)
v = 0
for x in liste:
v = v + x**2
v = v/len(liste) - m**2
return np.sqrt(v)
```

```
In [5]: for liste in [groupeA, groupeB, groupeC, groupeD, groupeE]:
print(ecarttype(liste))
```

```
15.584848866856293
15.427431768428896
14.829367565492111
13.743083400583721
15.478680202663499
```

3. Comparer les séries selon l'indicateur (moyenne, écart-type).

2. Médiane - Étendue

4. Écrire une fonction qui détermine l'étendue d'une liste, puis calculer celle-ci pour chacune des cinq séries.

```
In [6]: def etendue(liste):
mini = liste[0]
maxi = liste[0]
for x in liste :
if x > maxi:
maxi = x
if x < mini:
mini = x
return maxi-mini
```

```
In [7]: for liste in [groupeA, groupeB, groupeC, groupeD, groupeE]:
print(etendue(liste))
```

```
94
89
98
74
70
```

5. Reste la médiane. La définition vue en cours nécessite de classer les données dans l'ordre croissante. La méthode liste.sort() permet de faire cela.

```
In [8]: def mediane(liste):
#il faut trier les listes dans l'ordre croissant
liste.sort()
N = len(liste)
if N%2 == 1 : # si N est impair
return liste[(N+1)//2]
else:
m1 = N//2
m2 = N//2+1
return (liste[m1]+liste[m2])/2
```

```
In [9]: for liste in [groupeA, groupeB, groupeC, groupeD, groupeE]:
print(mediane(liste))
```

```
61
63.5
65
69.0
75.0
```

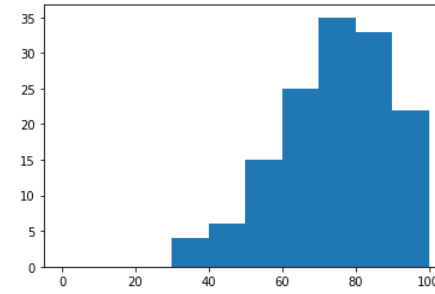
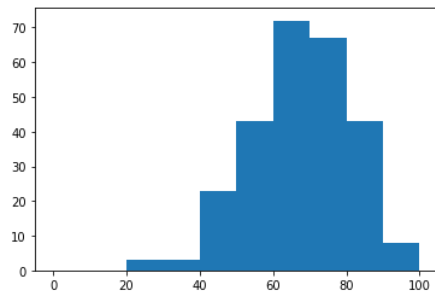
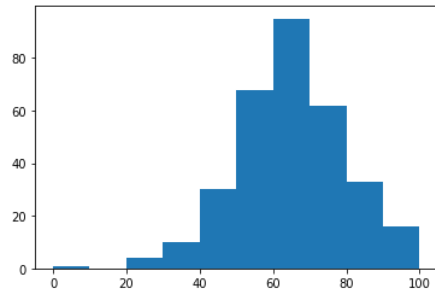
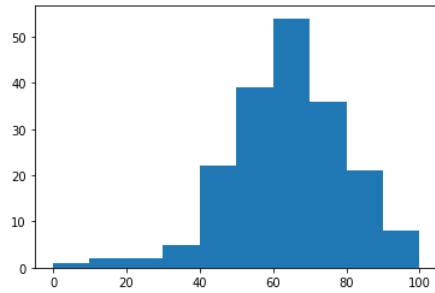
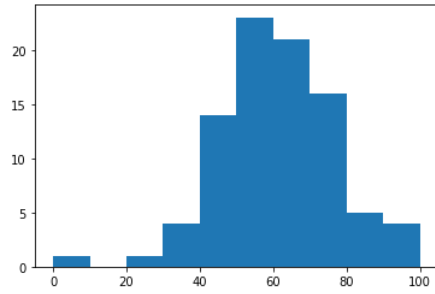
6. Comparer selon l'indicateur (médiane, étendue).

3. Représentations graphiques

7. On va tracer des histogrammes représentant les groupes. Pour cela on va utiliser la bibliothèque matplotlib.pyplot qui possède une fonction plt.hist qui fait cela toute seule.

```
In [14]: import matplotlib.pyplot as plt
```

```
for liste in [groupeA, groupeB, groupeC, groupeD, groupeE]:
plt.hist(liste, bins=[0,10,20,30,40,50,60,70,80,90,100])
plt.show()
```



8. De même, on peut tracer les boîtes à moustaches.

```
In [23]: plt.boxplot([groupeA, groupeB, groupeC, groupeD, groupeE], vert = False, labels = ['A', 'B', 'C', 'D', 'E'], show())
```

