Statistique descriptive

1. Indicateurs de position

❖ La moyenne

Définition. *La* <u>*moyenne*</u> *d'une série statistique est la somme de toutes les valeurs divisée par l'effectif total. On la note* $\bar{x}$.

<u>Exemple A</u>

Voici les âges des élèves d'une classe ayant eu mention très bien au brevet :

14 – 15 – 15 – 14 – 13 – 13 – 14 – 14 – 14 – 13.

La somme des notes est de $14 + 15 + 15 + 14 + 13 + 13 + 14 + 14 + 14 + 13 = 139$.

Comme l'effectif est de 10 élèves, la moyenne des âges est $\frac{139}{10} = 13{,}9$ ans.

Considérons une série statistique prenant p valeurs x_1, x_2, …, x_p d'effectifs respectifs n_1, n_2, …, n_p comme dans le tableau ci-dessous

Valeur du caractère x_i	x_1	x_2	x_3	…	x_p
Effectif n_i	x_1	n_2	x_3	…	n_p

On calculera la moyenne avec la formule suivante $\frac{n_1x_1+n_2x_2+\cdots+n_px_p}{n_1+n_2+\cdots+n_p}$.

<u>Exemple A</u>

Les données précédentes peuvent se regrouper dans ce tableau.

Âge	13	14	15
Effectif	3	5	2

La moyenne peut dont se calculer ainsi : $\frac{3\times13+5\times14+2\times15}{3+5+2} = \frac{139}{10} = 13{,}9$.

Théorème (linéarité de la moyenne).

- Si on ajoute à toutes les valeurs d'un caractère quantitatif un nombre a, alors la moyenne est augmentée de a.
- Si on multiplie toutes les valeurs d'un caractère quantitatif par un nombre a, alors la moyenne est multipliée par a.

<u>Exemple</u>

Dans une entreprise, le salaire moyen des femmes est de 1 206 € par mois. Au bout de six mois, les salaires augmentent de 100 €, par conséquent le nouveau salaire moyen des femmes est : 1206 + 100 = 1306 €.

Six mois plus tard, les salaires augmentent de 5 %. Puisque une augmentation de 5 % se traduit par une multiplication par 1,05, le salaire moyen des femmes devient :

$$1306 \times 1{,}05 = 1371{,}3 \text{ €}.$$

Théorème (moyenne par paquet). Si une population est partagée en deux sous-groupes disjoints, d'effectifs p et q, et si les moyennes de ces sous-groupes sont respectivement M_1 et M_2, on peut calculer la moyenne de la série par la formule

$$\frac{pM_1 + qM_2}{p+q}.$$

<u>Exemple</u>

Le salaire moyen des 47 hommes d'une entreprise est de 1380 € et celui des 35 femmes est de 1206 €. Le salaire moyen dans cette entreprise est donc égal à :

$$\frac{47 \times 1380 + 35 \times 1206}{47 + 35} \approx 1306 \text{ €}.$$

❖ Médiane

Définition. *La <u>médiane</u> est une valeur qui partage la population en deux moitiés : celle dont les individus ont une valeur du caractère inférieure à cette médiane, et l'autre, dont les individus ont une valeur du caractère supérieure à cette médiane.*

En pratique, pour déterminer la médiane, on range les N valeurs par ordre croissant, et on distingue deux cas :

- Si N est impair, on prend la valeur centrale, c'est-à-dire la valeur de rang $\frac{N+1}{2}$.
- Si N est pair, on prend la moyenne des deux valeurs centrales, c'est-à-dire des valeurs de rang $\frac{N}{2}$ et $\frac{N}{2} + 1$.

<u>Exemple B</u>

Imaginons un devoir où les notes ont été

3 – 5 – 6 – 7 – 7 – 10 – 11 – 12 – 13 – 13 – 14 – 15 – 18 – 19.

Cette série compte 14 valeurs, donc la médiane est la moyenne des 7^{e} et 8^{e} valeurs, elle est donc égale à $\frac{11+12}{2} = 11{,}5$.

Cela signifie que 50 % des élèves ont eu moins de 11,5 et 50 % des élèves ont eu plus de 11,5.

❖ Quartiles

Définition. *Le <u>premier quartile</u> Q_1 est la plus petite valeur du caractère telle que 25 % de l'effectif ait une valeur du caractère inférieure ou égale à Q_1.*
Le <u>troisième quartile</u> Q_3 est la plus petite valeur telle que 75 % de l'effectif ait une valeur du caractère inférieure ou égale à Q_3.

S'il y a N valeurs, Q_1 est la valeur dont le rang est **le premier entier supérieur ou égal** à $\frac{N}{4}$ et Q_3 est la valeur dont le rang est le **premier entier supérieur ou égal** à $\frac{3}{4}N$.

<u>Exemple B</u>

Puisque $\frac{14}{4} = 3{,}5$, le premier quartile est la 4^{e} valeur de la série, donc $Q_1 = 7$.

Puisque $\frac{3}{4} \times 14 = 10{,}5$, le troisième quartile est la 11^{e} valeur de la série, donc $Q_1 = 14$.

2. Indicateurs de dispersion

Définition. *L'<u>étendue</u> d'une série statistique est la différence entre la plus grande et la plus petite valeur.*

Définition. On appelle <u>écart interquartile</u> le nombre $Q_3 - Q_1$.

Définition. L'<u>écart-type</u> de la série statistique de x_1, x_2, ..., x_p et d'effectifs n_1, n_2, ... n_p est le nombre

$$\sigma = \sqrt{\frac{n_1(x_1-\bar{x})^2 + n_2(x_2-\bar{x})^2 + \cdots n_p\left(x_p-\bar{x}\right)^2}{n_1+n_2+\cdots+n_p}}$$

L'écart-type est la racine carrée des moyennes des carrées des écarts à la moyenne.
C'est un indicateur de dispersion associé à la moyenne. Plus l'écart-type est grand, plus les valeurs sont dispersées autour de la moyenne. Il est toujours positif.

<u>Exemple A</u>

Les données précédentes peuvent se regrouper dans ce tableau.

Âge	13	14	15
Effectif	3	5	2

La moyenne peut dont se calculer ainsi : $\frac{3\times13+5\times14+2\times15}{3+5+2} = \frac{139}{10} = 13{,}9$.

$$\sigma = \sqrt{\frac{3(13-13{,}9)^2 + 5(14-13{,}9)^2 + 2(15-13{,}9)^2}{3+5+2}} = 0{,}7.$$

Voici un exemple « caricatural » qui permet de bien saisir l'intérêt de l'écart-type.

<u>Exemple</u>

Un devoir a été donné à deux classes de Seconde de 30 élèves chacune. Les notes sont données ci-dessous

Notes	0	20
Effectifs	15	15

Seconde A

Notes	10
Effectifs	30

Seconde B

Bien sûr, la moyenne est de 10 dans chaque classe. En revanche les écarts-types sont en

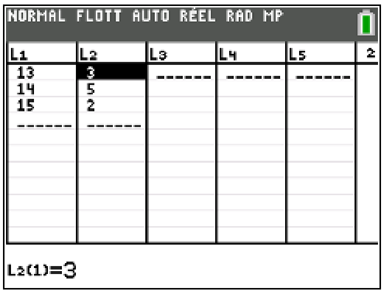
- seconde A : $\sigma_A = \sqrt{\frac{15(0-10)^2+15(20-10)^2}{30}} = 10$;
- seconde B : $\sigma_B = \sqrt{\frac{30(10-10)^2}{30}} = 0$.

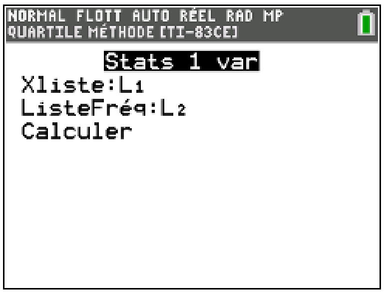
Cela traduit que les notes en seconde B sont peu (pas) dispersées.

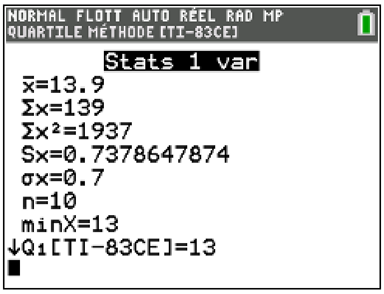
3. Calculs statistiques avec la calculatrice et Python

❖ Texas Instruments TI-83

Appuyer sur stats, *ÉDIT*, puis *Modifier...* Dans une des listes rentrer les âges, et dans une autre les effectifs.
Pour lancer les calculs, appuyer sur stats, *CALC* puis *Stats 1 Var*. Renseigner le nom de liste contenant les valeurs de la série, puis cette contenant les effectuer. Choisir *Calculer*.

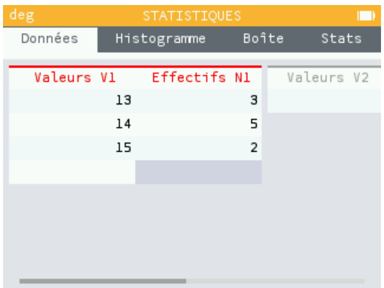


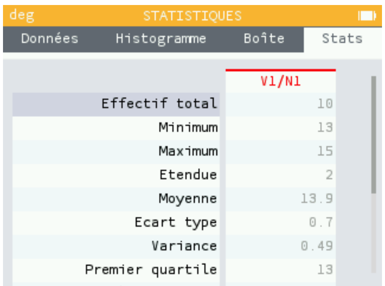




❖ Numworks

Dans le menu *Statistiques*, compléter les valeurs de la série (colonne V1) puis les effectifs (colonnes N1).
Le menu *Stats* affiche les indicateurs.





❖ Python

Le module statistics de Python permet d'effectuer les calculs de moyenne et écart-type sur Python.

```
>>> from statistics import *

>>> L= 3*[13]+5*[14]+2*[15]

>>> L
[13, 13, 13, 14, 14, 14, 14, 14, 15, 15]

>>> mean(L) # moyenne de la série L
13.9

>>> pstdev(L) # écart-type de la série L
0.7
```