Statistique descriptive

Faire des **statistiques**, c'est recueillir, organiser, synthétiser, représenter et exploiter des données, numériques ou non, dans un but de comparaison, de prévision, de constat...
Les plus gros « consommateurs » de statistiques sont les **assureurs** (risques d'accidents, de maladie des assurés), les **médecins** (épidémiologie), les **démographes** (populations et leur dynamique), les **économistes** (emploi, conjoncture économique), les **météorologues**, l'**industrie** (prévisions de défaillance de machines), …

I – Rappels

En statistique, nous pouvons étudier :
– les **valeurs** d'une quantité (exemple : l'âge des élèves d'une classe)
– les **effectifs** de chaque valeur (exemple : le nombre d'individus ayant un âge donné) ;
– les **fréquences** de chaque valeur (exemple : la proportion d'individu ayant un âge donné par rapport à la totalité de la population). Une fréquence peut s'exprimer sous la forme :

- d'un nombre décimal entre 0 et 1 (exemple : 0,25)
- d'un pourcentage (exemple : 25 %)
- d'une fraction (exemple : 1/4)

Exemple 1 : Étude des âges des élèves d'une classe de seconde.

Âge	14	15	16	17	**TOTAL**
Effectif	1	27	5	2	$N = 35$
Fréquence (fraction)					
Fréquence (décimal)					
Fréquence (%)					

Définitions / Notations

Les **valeurs** sont souvent notées $x_1, x_2, x_3, \dots, x_p$.

Les **effectifs des valeurs** sont souvent notés $n_1, n_2, n_3, \dots, n_p$.

L'effectif total N est : $N = n_1 + n_2 + n_3 + \dots + n_p$.

Les **effectifs des valeurs** sont souvent notés $f_1, f_2, f_3, \dots, f_p$ et se calculent ainsi : $\boxed{f_i = \frac{n_i}{N}}$.

La somme de toutes les fréquences est égale à 1.

II – Indicateurs de tendance centrale

1) Moyenne pondérée

La **moyenne pondérée** est définie par : $\boxed{m = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \dots + n_p x_p}{N} = \frac{\sum (n_i x_i)}{N}}$

Exemple 1 : calculez la moyenne des âges.
Réponse : m = ……………………………………………….

Remarque : m peut se calculer aussi avec : $m = f_1 x_1 + f_2 x_2 + f_3 x_3 + \dots + f_p x_p$.

Exemple 1 : calculez la moyenne des âges avec la seconde formule.
Réponse : m = ……………………………………………….

Propriété : linéarité de la moyenne
Soient a et b deux nombres.
Si une série de valeurs (x_i) a pour moyenne m alors la série de valeurs ($a\,x_i + b$) a pour moyenne $a\,m + b$.

Exemple 2 : Étude de mesures d'une distance entre deux points sur un terrain.

Mesures (en cm)	122	119	121	118	123	124
Effectifs	13	18	21	15	7	6

Ici, toutes les mesures sont proches de 120, elles peuvent donc s'écrire $120 + x_i$:

Écarts x_i avec 120	2	– 1	…………	…………	…………	…………
Effectifs	13	18	…………	…………	…………	…………

Calculez maintenant la moyenne de la série des (x_i) :

m = ………………………………………………………………

il suffit enfin d'ajouter 120 pour trouver la moyenne des mesures : ………………………………

2) Médiane et quartiles

La **médiane** est la valeur de la série qui partage l'effectif en deux sous-effectifs égaux : au moins 50 % des valeurs sont inférieures ou égales à la médiane.

Remarque : il faut que les valeurs soient rangées dans l'ordre croissant.

Exemple 1 : donnez l'âge médian.
Réponse : M_e = ……

Le **premier quartile** Q_1 est la première valeur de la série telle qu'au moins 25 % des valeurs sont inférieures ou égales à Q_1.
Le **deuxième quartile** Q_2 est la médiane.
Le **troisième quartile** Q_3 est la première valeur de la série telle qu'au moins 75 % des valeurs sont inférieures ou égales à Q_3.

Exemple 1 : donnez les quartiles.
Réponse : Q_1 = …… Q_2 = …… Q_3 = ……

III – Indicateurs de dispersion

1) Étendue

L'**étendue** est la différence entre la plus grand valeur et la plus petite valeur.

Exemple 1 : l'étendue des âges est …………

2) Écart inter-quartiles

L'**écart inter-quartiles** est égal à $Q_3 – Q_1$.

Exemple 1 : donnez l'écart inter-quartiles :
Réponse : $Q_3 – Q_1$ = …………

3) Écart-type

La **variance** se calcule ainsi :

$$\boxed{V=\frac{n_1(x_1-m)^2+n_2(x_2-m)^2+n_3(x_3-m)^2+\ldots+n_p(x_p-m)^2}{N}}$$

L'**écart-type** est la racine carrée de la variance : $\boxed{s=\sqrt{V}}$ (noté parfois σ).

Exemple 1 : complétez le tableau ci-dessous (arrondissez à 0,01 près) :

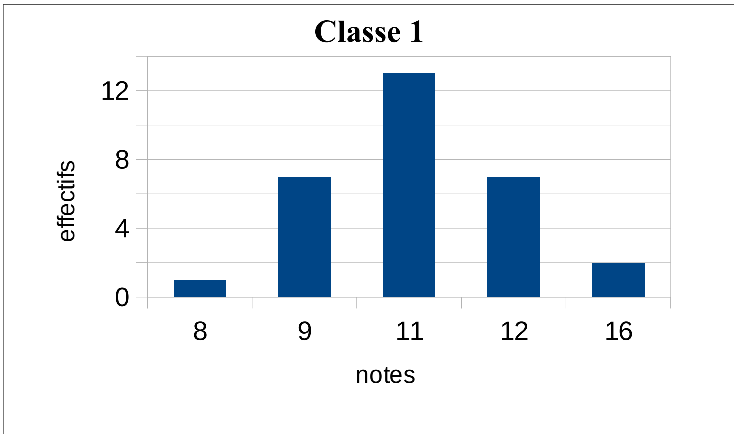
Âges x_i	14	15	16	17
Écarts à la moyenne $x_i - m$				
Carrés des écarts $(x_i - m)^2$				
Effectifs n_i	1	27	5	2

donc V = ………………………………………… et s = ………………………

III - Comparaisons de deux séries

Il est fréquent de comparer deux séries de valeurs en les résumant chacune à un indicateur de position et un indicateur de dispersion.

Exemple 3 : voici les résultats à un même devoir donné à deux classes différentes :



Pour chacune des deux classes, calculez la moyenne :

m_1 = …………………………………………

m_2 = …………………………………………

Calculez maintenant l'écart-type pour chacune des classes :

V_1 = ……………………………………………………………… donc $s_1=\sqrt{V_1}$ ≈ …………………………

V_2 = ……………………………………………………………… donc $s_2=\sqrt{V_2}$ ≈ …………………………

Que remarquez-vous ? Est-ce cohérent avec les graphiques ?

……
……
……
……