Statistiques : tendance centrale et mesure de dispersion.

Le caractère étudié est quantitatif, on cherche à résumer la répartition des termes x_i de la série statistique.

I Médiane et quartiles.

Une mesure de la tendance centrale : la médiane. la médiane est un nombre souvent noté M_e tel que les effectifs des classes $]-\infty; M_e]$ et $[M_e;+\infty[$ soient au moins égaux à la moitié de la population totale.
On peut retenir : "La médiane est une valeur du caractère qui partage l'effectif en deux parties d'effectifs égaux"
Méthodes : on ordonne tous les termes de la série par ordre croissant (chaque valeur est répétée un nombre de fois égal à son effectif):

Effectif total impair : N=2p+1	Effectif total pair : N=2p
$x_1 \leqslant \ldots \leqslant x_p \leqslant x_{p+1} \leqslant x_{p+2} \leqslant \ldots \leqslant x_{2p+1}$	$x_1 \leqslant \ldots \leqslant x_p \leqslant x_{p+1} \leqslant \ldots \leqslant x_{2p}$
$\boxed{M_e = x_{p+1}}$	$\boxed{M_e = \frac{x_p + x_{p+1}}{2}}$

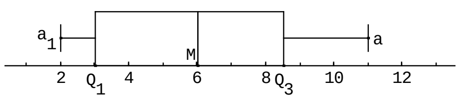
Si les valeurs sont regroupées par classes : on trace le polygone des effectifs cumulés croissants. L'effectif de la classe $]-\infty;M_e]$ est égal à 50% de l'effectif total

Une mesure de dispersion : les quartiles.
Le premier quartile, noté Q_1, est la plus petite valeur x_q des termes de la série, tel que l'effectif de la classe $]-\infty;x_q]$ soit au moins égal à 25% de l'effectif total.
Le troisième quartile, noté Q_3, est la plus petite valeur $x_{q'}$ des termes de la série, tel que l'effectif de la classe $]-\infty;x_{q'}]$ soit au moins égal à 75% de l'effectif total.
$]\, Q_1 ; Q_3 \,[$ est appelé **intervalle inter-quartile** (environ 50% de la population totale).
$Q_3 - Q_1$ est appelé **écart inter-quartile.**
Méthodes : on ordonne tous les termes de la série par ordre croissant (chaque valeur est répétée un nombre de fois égal à son effectif):

Effectif total multiple de 4 : N=4q	Effectif total non multiple de 4 : $q = E\left(\frac{N}{4}\right)+1$ et
$x_1 \leqslant \ldots \leqslant x_q \leqslant \ldots \leqslant x_{3q} \leqslant \ldots \leqslant x_{4q}$	$q' = E\left(\frac{3N}{4}\right)+1$: $x_1 \leqslant \ldots \leqslant x_q \leqslant \ldots \leqslant x_{q'} \leqslant \ldots \leqslant x_N$
$\boxed{Q_1 = x_q \,;\, Q_3 = x_{3q}}$	$\boxed{Q_1 = x_q \,;\, Q_3 = x_{q'}}$

Si les valeurs sont regroupées par classes : on trace le polygone des effectifs cumulés croissants. L'effectif de la classe $]-\infty;Q_1]$ est égal à 25% de l'effectif total et l'effectif de la classe $]-\infty;Q_3]$ est égal à 75% de l'effectif total.

L'ensemble de ces paramètres peut-être présenté dans un **diagramme en boîtes** ou "boîtes à moustaches": ci-contre les valeurs extrêmes sont 2 et 11, la médiane vaut 6, le premier quartile vaut 3 et le troisième quartile vaut 8,5.



Remarque : on définit, d'une façon analogue au premier et troisième quartiles, les premiers et neuvièmes déciles isolant respectivement 10% et 90% des données. Les extrémités du diagramme en boîtes, au lieu de représenter les valeurs extrêmes, peuvent alors correspondre au premier et neuvième déciles de la série statistique.

Influence d'un changement de variable affine : soient $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$ et la série statistique de termes y_i tels que : $y_i = ax_i + b$, de médiane M_e' de premier quartile Q_1' et de troisième quartile Q_3' alors : $\boxed{M_e' = aM_e + b}$
Si $a>0$ alors $Q_1' = aQ_1 + b$ et $Q_3' = aQ_3 + b$
Si $a<0$ alors $Q_1' = aQ_3 + b$ et $Q_3' = aQ_1 + b$

II. Moyenne et écart-type

Notations : les termes de la série statistique prennent k valeurs différentes notées x_i
L'effectif de la valeur x_i est noté n_i. (la valeur x_i est prises n_i fois par le caractère)
La série statistique est alors notée $(x_i;n_i)$

Ainsi l'effectif total est : $N = n_1 + n_2 + \ldots + n_k$ i.e. $\boxed{N = \sum_{i=1}^{k} n_i}$

La fréquence de chaque valeur est : $f_i = \frac{n_i}{N}$

Une mesure de la tendance centrale, la moyenne :

$$\overline{x} = x_1 f_1 + x_2 f_2 + \ldots + x_k f_k \text{ i.e. } \boxed{\overline{x} = \sum_{i=1}^{k} x_i f_i}$$

Propriété : la fonction f définie sur $\mathbb{R}$ par $f(x) = (x - x_1)^2 f_1 + \ldots + (x - x_k)^2 f_k$ atteint son minimum pour $x = \overline{x}$
Une mesure de la dispersion, la variance :

$$V = (\overline{x} - x_1)^2 f_1 + \ldots + (\overline{x} - x_k)^2 f_k \text{ i.e. } \boxed{V = \sum_{i=1}^{k} (\overline{x} - x_i)^2 f_i}$$

"la variance est la moyenne des carrés des écarts à la moyenne"

Propriété : $V = x_1^2 f_1 + \ldots + x_k^2 f_k - (\overline{x})^2$ i.e. $\boxed{V = \left(\sum_{i=1}^{k} x_i^2 f_i\right) - (\overline{x})^2}$

"la variance est la moyenne des carrés moins le carré de la moyenne"

Ecart-type : $\boxed{\sigma = \sqrt{V}}$
Remarque : V est positif ou nul car c'est la moyenne de carrés de nombres réels
L'écart-type σ est homogène aux valeurs de la série statistique.

Influence d'un changement de variable affine : soient $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$ et la série statistique $(y_i;n_i)$ définie par $y_i = ax_i + b$ pour tout $i \in \{1;\ldots;k\}$ alors:

$$\boxed{V_y = a^2 V_x} \text{ et } \boxed{\sigma_y = |a| \sigma_x}$$