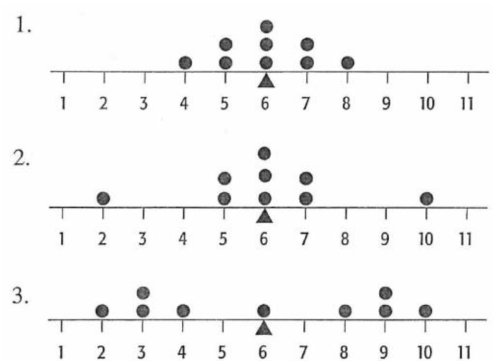
4. Les mesures de dispersion

Les mesures de dispersion permettent de mesurer la dispersion des données d'une série statistique. Nous étudierons plus particulièrement l'étendue et l'écart-type.

A. Étendue $= MAX - min$

Analysons les trois séries statistiques suivantes, représentées par un pictogramme (il s'agit du nom de ce type de graphique).

(après p.26)



$\sigma^2 = 1{,}333...$
$\sigma = 1{,}15$

$\sigma^2 = 4$
$\sigma = 2$

$\sigma^2 = 8{,}444...$
$\sigma = 2{,}91$

Ces trois séries statistiques présentent des moyennes identiques, mais la dispersion des données autour de ces moyennes diffère d'une série à l'autre. Les données de la série 1 sont assez concentrées autour de la moyenne, alors que la série 2 comporte deux données qui sont passablement éloignées de la moyenne. Les données de la série 3 sont encore plus dispersées par rapport à la moyenne que celles de la série 2. Comment peut-on mesurer mathématiquement cette dispersion?

L'étendue de la série pourrait peut-être nous permettre de la mesurer. Calculons l'étendue de chacune des séries :

	Série 1	Série 2	Série 3
Étendue	4	8	8

Comme l'étendue ne tient compte que de la plus grande et de la plus petite valeur de la série, on ne peut mesurer la différence de dispersion entre les données de la série 2 et de la série 3.

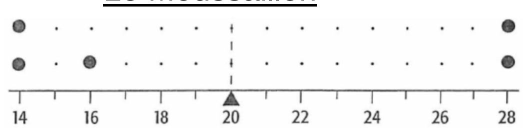
En fait, l'étendue sert très peu comme mesure de dispersion. On l'utilise surtout en contrôle de la qualité où les échantillons sont souvent de petites tailles (4 ou 5 éléments); dans ce cas, l'étendue donne une bonne idée de la dispersion des résultats de l'échantillon.

B. Variance et écart type

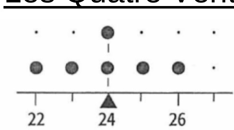
L'écart type est une mesure de dispersion qui, contrairement à l'étendue, tient compte de toutes les valeurs de la série de données. La variance est une donnée intermédiaire qui permet de calculer l'écart type.

Exemple 1: Supposons qu'un étudiant de cégep offre ses services à titre de matelot pendant les vacances. Il a reçu deux offres d'emploi avec des conditions de travail semblables. Il décide de faire son choix en se basant sur l'âge moyen de ses compagnons de voyage. Il a donc préféré l'offre d'emploi du voilier *Le Moussaillon*, où l'âge moyen est de 20 ans, à celui du voilier *Les Quatre Vents*, où il est de 24 ans. Les pictogrammes ci-dessous indiquent la distribution de l'âge pour chaque voilier. Notre cégépien a-t-il fait le bon choix?

Le Moussaillon

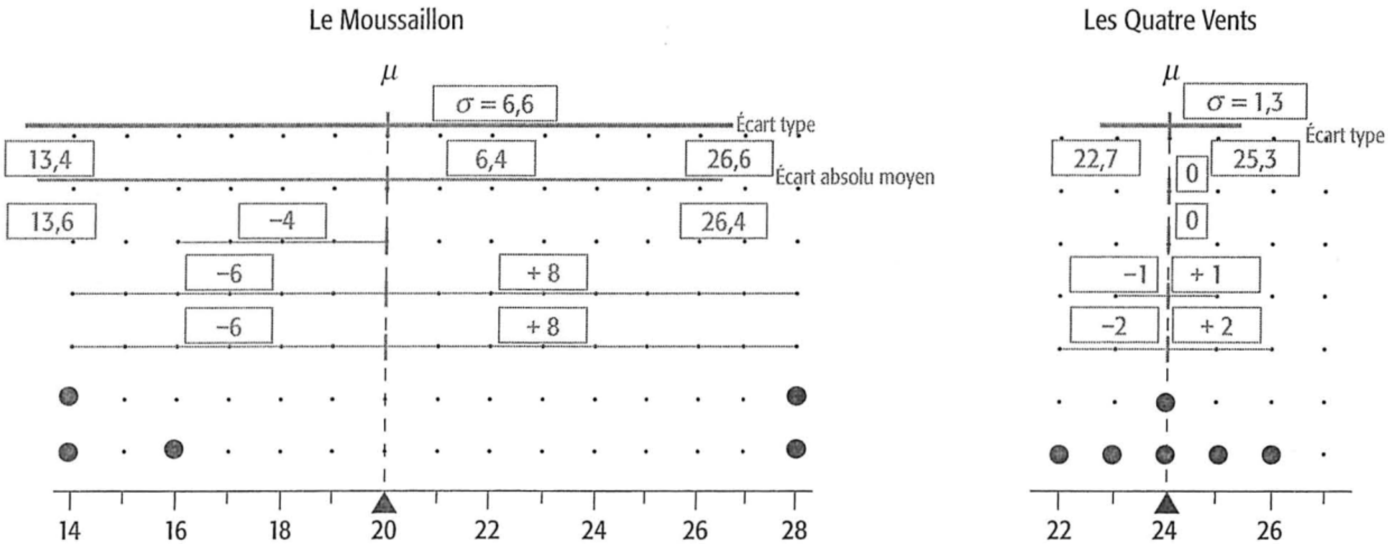


Les Quatre Vents



Bien sûr que non : il devra passer ses vacances avec M. et Mme Tremblay et leurs trois neveux, alors que l'autre voilier semble accueillir une équipe jeune et dynamique. La différence principale entre les deux distributions est la dispersion des données par rapport à la moyenne. Comment s'y prendre pour mesurer cette dispersion?

★ Pour chacune des données des pictogrammes, tracer un segment de droite représentant l'écart entre la donnée et la moyenne, et inscrire la valeur de cet écart sur le segment de droite.



a) Pour le voilier *Le Moussaillon*, calculer la moyenne des écarts d'âge par rapport à la moyenne.

$$\frac{(8)\times 2+(-6)\times 2+(-4)}{5}=\frac{0}{5}=0$$

b) Que remarques-tu?

La moyenne des écarts est nulle.

La moyenne des écarts donnera toujours 0. En effet, la moyenne étant le centre d'équilibre du pictogramme, la somme des écarts située à gauche de la moyenne sera toujours égale, mais de signe négatif, à la somme des écarts située à droite de la moyenne.

c) Pour le voilier *Le Moussaillon*, calculer la moyenne des écarts d'âge en valeur absolue par rapport à la moyenne.

$$\frac{|8|\times 2+|-6|\times 2+|-4|}{5}=\frac{32}{5}=6,4$$

Comme personne n'aime travailler avec des valeurs absolues, nous allons transformer cette formule pour qu'elle soit plus facile à exploiter. Un moyen approprié d'éliminer les valeurs absolues consiste à appliquer le principe suivant : la valeur absolue d'un nombre élevée au carré donne le même résultat que donnerait ce nombre élevé au carré.

Par exemple, $|-4|^2 = (-4)^2 = 16$. Ce principe se nomme « variance ».

d) Calculer la variance pour le voilier *Le Moussaillon*.

$$\sigma^2 = \frac{(8)^2 \times 2 + (-6)^2 \times 2 + (-4)^2}{5} = \frac{216}{5} = 43{,}2 \text{ ans}^2$$ *!!!*

Comme la variance nous donne la moyenne des carrés des écarts, il faut extraire la racine de la variance pour obtenir un écart typique de tous les écarts de la série. On donne le nom d'écart type à cet écart que l'on note σ, ou sigma (s dans l'alphabet grec); la variance quant à elle est notée σ^2.

e) Calculer l'écart type de l'âge de l'équipage du voilier *Le Moussaillon* et le représenter sur le pictogramme de la page précédente par un segment de droite.

$$\sigma = \sqrt{43{,}2 \text{ ans}^2} = 6{,}6 \text{ ans}$$

Généralement, on trouve **la plupart** des données d'une distribution entre la moyenne moins un écart type et la moyenne plus un écart type, soit entre $\mu - \sigma$ et $\mu + \sigma$. Lorsque la distribution prend la forme d'une cloche (modèle normal), environ les deux tiers des données sont comprises dans cet intervalle.

f) Donner l'interprétation de l'écart-type pour le voilier *Le Moussaillon*.

La plupart des matelots (3 sur 5) du voilier *Le Moussaillon* ont un âge se situant à ± 6,6 ans de la moyenne d'âge de l'équipage de ce navire, soit entre 13,4 ans et 26,6 ans.

(on peut revenir à p.23)

g) Calculer l'écart type de l'âge de l'équipage du voilier *Les Quatre Vents* et le représenter sur le pictogramme de la page précédente par un segment de droite. Donner l'interprétation de cet écart type.

Variance (σ^2) = $\boxed{\dfrac{(-2)^2+(-1)^2+(0)^2\times 2+(1)^2+(2)^2}{6}=\dfrac{10}{6}=1,7\ \text{an}^2}$

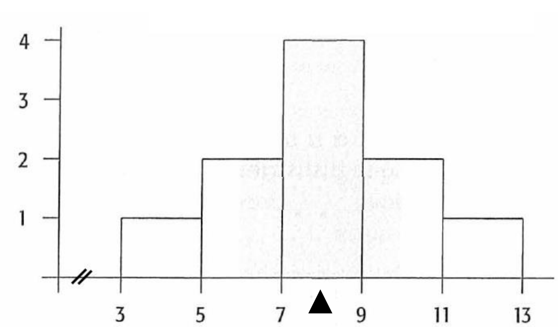
Écart type (σ) = $\boxed{\sqrt{1,7\ \text{an}^2}=1,3\ \text{an}}$

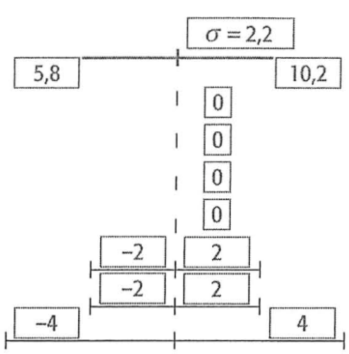
Interprétation

La plupart des matelots (4 sur 6) du voilier *Les Quatre Vents* ont un âge se situant à $\boxed{\pm 1,3}$ an(s) de la moyenne d'âge de l'équipage de ce navire, soit entre $\boxed{22,7}$ ans et $\boxed{25,3}$ ans.

Écart type et variance
$\sigma=\sqrt{\sigma^2}=\sqrt{\dfrac{\text{somme des carrés des écarts à la moyenne}}{\text{nombre total de données}}}=\sqrt{\dfrac{\Sigma(x_i-\mu)^2 n_i}{N}}=\sqrt{\Sigma(x_i-\mu)^2 f_i}$

Exemple 2: Calculer, représenter graphiquement et interpréter l'écart type de la distribution suivante.





$\mu=8$

$$\sigma^2=\frac{(-4)^2+(-2)^2\times 2+(0)^2\times 4+(2)^2\times 2+(4)^2}{10}$$

$$=\frac{16+8+0+8+16}{10}=\frac{48}{10}=4,8\ \text{unités}^2$$

$$\sigma=\sqrt{4,8}=2,2\ \text{unités}$$

La plupart des données se situent à ±2,2 unités de la moyenne, soit entre 5,8 et 10,2.