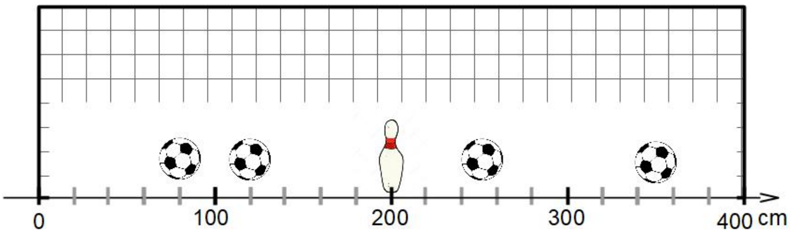
TD 6 : Statistiques descriptives : Paramètres de dispersion

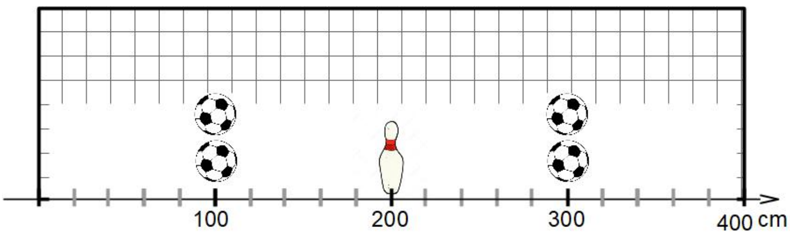
Ex 1. Julien s'entraîne au tir au but. Il vise une quille placée au milieu d'une cage de 4 mètres de largeur.



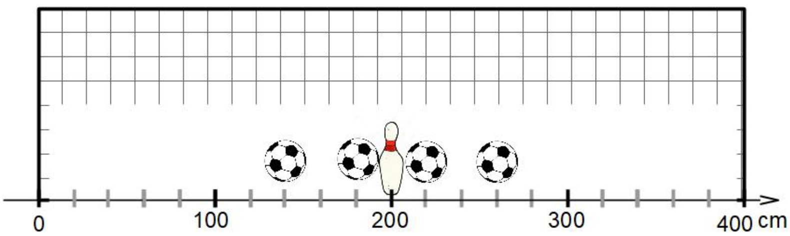
On repère la position de son i-ème tir par son absisse en centimètres x_i. Voici les résultats de ses quatre premiers tirs.

Tir	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	120		
2	250		
3	80		
4	350		
Moyenne			

a. Calculer la position moyenne $\overline{x}$ des tirs. Marquer $\overline{x}$ sur le graphique. Les tirs de Julien sont-ils bien centrés ?
b. Remplir le tableau et calculer la variance des x_i. En déduire l'écart-type σ.
Dessiner sur le graphique un intervalle de centre $\overline{x}$ et largeur 2σ.
c. Dans une autre série de 4 tirs, Julien a envoyé 2 fois le ballon en position 100 et 2 fois en position 300.



Après avoir vérifié que la position moyenne est toujours 200 cm, calculer l'écart-type de cette deuxième série. Aurait-on pu deviner ce résultat ?
d. Voici une dernière série de 4 tirs (placés à 140, 180, 220 et 260 cm respectivement).



L'écart-type σ est-il plus petit ou plus grand que pour sa deuxième série (question précédente) ? Essayer d'estimer sa valeur, puis calculer l'écart-type (pour des raisons de symétrie, la position moyenne est toujours 200 cm).

Solution. a. On calcule

$$\overline{x} = \frac{1}{4}(120 + 250 + 80 + 350) = 200.$$

La position moyenne des tirs est de 200 cm, ce qui correspond au milieu du but. Les tirs de Julien sont centrés.
b.

Tir	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	120	-80	6400
2	250	50	2500
3	80	-120	14400
4	350	150	22500
Moyenne	200	0	11450

La troisième colonne liste les écarts à la moyenne. Leur moyenne vaut 0. La dernière colonne liste les carrés des écarts à la moyenne. Leur moyenne est, par définition, la variance de la série de tir. Cette variance vaut 11450 cm^2 (attention aux unités : $x_i - \overline{x}$ est en cm, donc $(x_i - \overline{x})^2$ est en cm^2). L'écart-type est la racine carrée de la variance, soit $\sigma = \sqrt{11450} \simeq 107$ cm.
c. La moyenne vaut

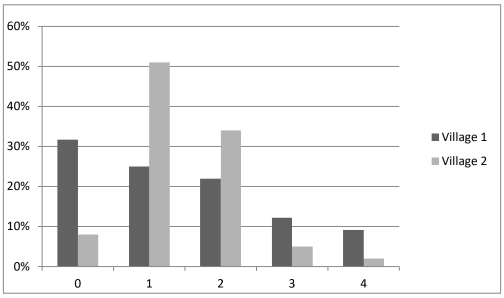
$$\frac{1}{4}(2 \times 100 + 2 \times 300) = 200.$$

Les tirs sont donc encore centrés. La variance vaut

$$\frac{1}{4}\left(2 \times (100 - 200)^2 + 2 \times (300 - 200)^2\right) = 10000.$$

et l'écart-type vaut $\sigma = \sqrt{10000} = 100$ cm. On aurait pu s'en douter : l'écart à la moyenne (ici, le milieu du but) est le même pour les quatre tirs, et vaut aussi 100 cm. **d.** Les tirs ont l'air plus dispersés, mais en fait, il sont tous plus rapprochés de la moyenne (Julien est plus *précis*. L'écart-type sera donc plus petit que pour la deuxième série.
Aucun tir n'est à plus de 60 cm de la moyenne, donc l'écart-type sera d'au plus 60 cm. De même, aucun tir n'est à moins de 20 cm de la moyenne, donc l'écart-type sera d'au moins 20 cm.
Le calcul exact donne un écart-type d'environ $44,72$ cm.

Ex 2. Le diagramme en barre ci-dessous donne la distribution du nombre d'enfants par foyer dans deux petits villages de province.



a. Pour quel village s'attend-on à avoir le plus grand écart-type ?
b. Voici le tableau des effectifs pour le village 1 :

Nb. d'enfants	0	1	2	3	4	Total
Effectif	52	41	36	20	15	164

Calculer le nombre moyen d'enfants par foyer, la variance et l'écart-type.
c. Voici le tableau des fréquences pour le village 2 :

Nb. d'enfants	0	1	2	3	4	Total
Fréquence	8%	51%	34%	5%	2%	100%

Calculer le nombre moyen d'enfants par foyer, la variance et l'écart-type. Conclure.
Solution. a. Dans le village 2, la majorité des foyers 1 ou 2 enfants. Dans le village 1, le nombre d'enfants est dispersé sur une plage plus large, allant de 0 à 4 enfants par foyers. On s'attend à ce que l'écart-type soit plus élevé dans le premier village.
b. On trouve une moyenne de

$$\frac{1}{164}(52 \times 0 + 41 \times 1 + 36 \times 2 + 20 \times 3 + 15 \times 4) = \frac{233}{164} \simeq 1,42 \text{ enfants}.$$

La variance est alors de :

$$\frac{1}{164}\left(52 \times (0 - 1,42)^2 + 41 \times (1 - 1,42)^2 + 36 \times (2 - 1,42)^2 + 20 \times (3 - 1,42)^2 + 15 \times (4 - 1,42)^2\right) \simeq 1,67,$$

et l'écart-type de $\sqrt{1,67} \simeq 1,29$ enfants.

c. On trouve une moyenne de

$$0,08 \times 0 + 0,51 \times 1 + 0,34 \times 2 + 0,05 \times 3 + 0,02 \times 4 = 1,42 \text{ enfants.}$$

La variance est alors de :

$0,08\times(0-1,42)^2+0,51\times(1-1,42)^2+0,34\times(2-1,42)^2+0,05\times(3-1,42)^2+0,02\times(4-1,42)^2 = 0,6236$,

et l'écart-type de $\sqrt{0,6236} \simeq 0,79$ enfants. Les deux villages ont le même nombre d'enfants par foyer en moyenne, mais, comme prévu, l'écart-type dans le second village est plus petit que dans le premier.

Ex 3. Pour chacune des variables suivantes, dire pour quelle population on doit s'attendre à l'écart-type le plus grand.

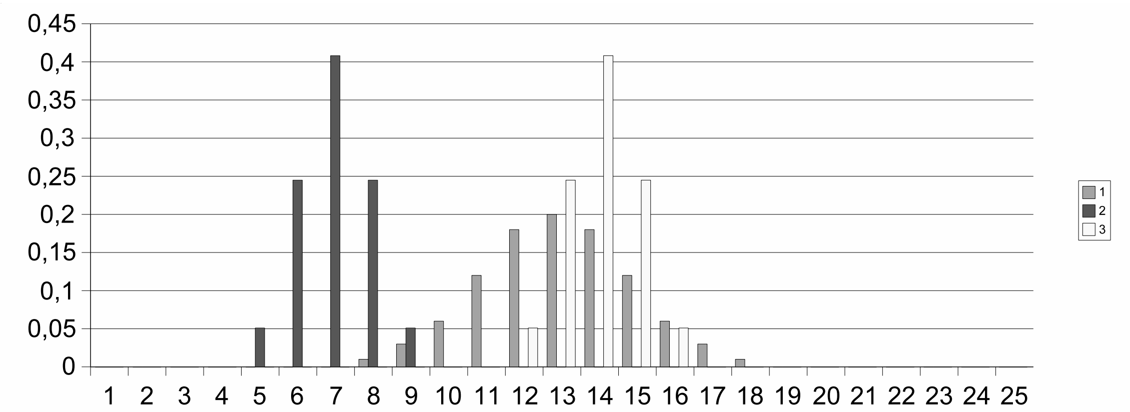
(1) X = taille des individus dans la population :
- des élèves d'un collège ;
- des élèves de 6ème ;
- des élèves de 4ème.

(2) X = revenus annuels dans la population :
- des maîtres d'école ;
- des habitants de Île-de-France.

(3) X = le prix au mètre carré dans la population :
- des appartements en vente à Orsay ;
- des appartements en vente en France.

Solution. a. Plus un caractère (variable statistique) est hétérogène au sein d'une population, plus l'on s'attend à ce que l'écart-type soit élevé. Par exemple, on s'attend à ce que :
Écart-type (élèves de collège) > Écart-type (élèves de 4ème) > Écart-type (élèves de 6ème),
la première inégalité parce que les élèves de collège vont de la 6ème à la 3ème, avec un taille variant beaucoup au cours de l'année, et la seconde inégalité parce que les élèves de 4ème, à cause de leur croissance, vont présenter des tailles plus variées qu'en 6ème.
De même, on s'attend à avoir :
Écart-type (revenus d'Île-de-France) > Écart-type (revenus de maîtres d'école),
et :
Écart-type (prix en France) > Écart-type (prix à Orsay).

Ex 4. La figure suivante représente les diagrammes en bâton des fréquences des notes à un examen dans trois classes 1, 2 et 3.



Répondre sans calcul aux questions suivantes :

a. Combien valent les moyennes des notes dans chacune des classes ?

b. L'écart-type des notes dans la classe 1 est-il plus petit, plus grand ou égal à l'écart-type dans la classe 2 ?

c. L'écart-type des notes dans la classe 1 est-il plus petit, plus grand ou égal à l'écart-type dans la classe 3 ?

d. L'écart-type des notes dans la classe 2 est-il plus petit, plus grand ou égal à 3 ?

Solution. a. Les notes sont distribuées de façon symétrique. La moyenne va donc être l'abscisse de leur axe de symétrie, soit :

— Classe 1 : moyenne de 13.
— Classe 2 : moyenne de 7.
— Classe 3 : moyenne de 14.

b. Les notes de la classe 1 s'étalent sur une plage plus grande que les notes de la classe 2. L'écart-type des notes dans la classe 1 est donc plus grand que l'écart-type dans la classe 2.

c. Les notes de la classe 1 s'étalent sur une plage plus grande que les notes de la classe 3. L'écart-type des notes dans la classe 1 est donc plus grand que l'écart-type dans la classe 3.

d. Dans la classe 2, toutes les notes sont entre 5 et 9, donc l'écart à la moyenne (qui est de 7) est d'au plus 2. L'écart-type des notes est donc aussi d'au plus 2.