

TD n°3 de Statistiques : Mesures de la dispersion

Les paramètres qui mesurent la dispersion *absolue* d'une série :

inconvenients

<i>Étendue</i>	<i>différence entre les 2 valeurs extrêmes (Max-min)</i>	les 2 valeurs extrêmes seulement élimination de 50% de l'effectif n'est pas donné par la calculatrice formule compliquée à comprendre
<i>Écart-inter-quartile</i>	<i>différence entre les quartiles 1 et 3 (= $Q_3 - Q_1$)</i>	
<i>Écart-moyen</i>	<i>moyenne des écarts à la moyenne</i>	
<i>Écart-type</i>	<i>racine carrée de la moyenne des carrés des écarts à la moyenne</i>	

I] Calculs de l'étendue, de l'écart-moyen et de l'écart-type

Tailles (en cm) de 15 élèves :

178	-	180	-	181	-	182	-	178	-	182	-	178	-	178	-	179	-	180	-	180	-	182
-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----

⊛ Déterminer l'étendue : $e = 182 - 178 = 4$.

L'écart-type est défini par la formule $\sigma = \sqrt{\frac{\sum n_i(x_i - \bar{x})^2}{\sum n_i}}$ mais on peut aussi le calculer avec cette variante de la formule

$\sigma = \sqrt{\frac{\sum n_i x_i^2}{\sum n_i} - \bar{x}^2}$ (racine carrée de la différence entre la moyenne des carrés et le carré de la moyenne) qui a l'avantage de n'introduire \bar{x} qu'à la fin des calculs : on calcule les $n_i x_i$, les $n_i x_i^2$, et leurs totaux nous permettent de trouver \bar{x} et σ .

Compléter le tableau ci-dessous.

x_i	178	179	180	181	182	183	Total	Moyenne	
n_i	5	2	3	1	4	0	15		
$n_i x_i$	890	358	540	181	728	0	2697	179,8	Écart-type 1,600
$n_i x_i^2$	158420	64082	97200	32761	132496	0	484959	32330,6	
$ x_i - m $	1,8	0,8	0,2	1,2	2,2	3,2			
$n_i x_i - m $	9	1,6	0,6	1,2	8,8	0	21,2	1,41	
$n_i (x_i - m)^2$	16,2	1,28	0,12	1,44	19,36	0	38,4	2,56	1,600

⊛ Calculer la moyenne $\bar{x} = \frac{T_1}{N} = \bar{x} = \frac{2697}{15} = 179,8$. Rappel : on a noté $N = \sum n_i$, $T_1 = \sum n_i x_i$ et $T_2 = \sum n_i x_i^2$.

Puis $\sigma^2 = \frac{T_2}{N} - \left(\frac{T_1}{N}\right)^2$. En déduire l'écart-type σ (σ^2 est appelé la variance de la série).

$\sigma = \sqrt{\frac{T_2}{N} - \left(\frac{T_1}{N}\right)^2} = \sqrt{32330,6 - (179,8)^2} = 1,6$. Nous avons utilisé les lignes 3 et 4 du tableau ci-dessus.

L'autre formule, la définition de l'écart-type, donne (7^{ème} ligne du tableau) le même résultat.

$\sigma = \sqrt{\frac{5(178-179,8)^2 + 2(179-179,8)^2 + \dots + 4(182-179,8)^2}{15}} = \sqrt{\frac{38,4}{15}} = 1,6$.

⊛ Calculs de l'écart-moyen : $e_m = \frac{21,2}{15} \approx 1,4$

L'écart-moyen nécessite le calcul préalable de \bar{x} et l'arrondi de cette valeur si on ne peut en garder la valeur exacte.

✎ Comparer les résultats de ces trois paramètres qui mesurent la dispersion absolue de cette série des tailles.

Les 3 paramètres de dispersion calculés indiquent de façon légèrement différente, la même chose. L'étendue est toujours le plus grand résultat. Il faudrait d'ailleurs prendre la moitié de l'étendue pour comparer aux autres valeurs (qui mesurent de 2 façons différentes l'écart des valeurs par rapport à la moyenne) : soit 2 fois cette demi-étendue. Vient ensuite l'écart-type : 1,6 qui est plus sensible aux valeurs extrêmes car leur carré pèse plus dans la moyenne. La plus petite mesure de la dispersion est donnée ici par l'écart-moyen : 1,4 environ. Est-ce toujours dans cet ordre que l'on trouve ordonnées ces valeurs ? Laissons la question ouverte. Chacun peut y réfléchir...

Maintenant, le commentaire que l'on peut faire : la moyenne des tailles est de 179,8 cm (ils sont plutôt grands) et la dispersion est faible, seulement 2 cm en plus ou en moins, et en moyenne environ 1,5 cm seulement (pour ne pas faire de jaloux on peut prendre la moyenne entre l'écart-type et l'écart-moyen) en plus ou en moins. Un Sherlock Holmes pourrait inférer qu'il s'agit d'une équipe de jeunes basketteurs vu la grande taille et la faible dispersion des valeurs (ils sont tous grands) mais sans doute manquons nous de certains éléments...

Utilisation du mode statistique de la calculatrice (voir pages 227-228 sur votre manuel) :

On entre les valeurs (x_i) et les effectifs (n_i) dans deux colonnes d'un tableau, puis on demande les statistiques à 1 variable (1Var) pour ces deux colonnes. On obtient tous les paramètres calculés ici plus quelques autres que vous reconnaîtrez comme Q_1 , M et Q_3 (on obtient $Q_1=178$, M=180 et $Q_3=182$ qui sont entiers car il ne s'agit pas de classe).

II] Comparaison de séries statistiques

Calculs d'indicateurs *relatifs* de la dispersion : en divisant un indicateur absolu par une valeur centrale (moyenne, médiane ou mode) on obtient un indicateur relatif (sans unité) qui peut être utilisé pour comparer des séries très différentes. L'écart-type relatif $\frac{\sigma}{\bar{x}}$ est appelé *coefficient de variation*.

Voici les notes moyennes de maths et physique/chimie pour le 1^{er} trimestre d'un groupe de douze élèves :

Maths	10	10,3	15	19	16,8	16,5	17,5	11	13,5	15,5	12,7	10,2
Phys./Ch.	9,2	10	10,8	14,6	12,6	14,1	15,2	8,9	12,1	12	12,7	7,4

★ Déterminer la moyenne, l'écart-type et les coefficients de variation de ces deux séries

voir le tableau ci-contre.

NB : On peut utiliser la calculatrice en mode statistique pour cela : il faut entrer les deux colonnes de valeurs et aussi une 3^{ème} colonne pour les effectifs (égaux à 1 pour chaque valeur).

✎ Comparer ces deux séries.

La moyenne de maths est de 2,32 points supérieure à celle de physique. L'écart-type en maths est aussi plus élevé qu'en physique (de 0,7 points dans la série des 12 notes mais de 0,98 dans la série complète). En valeurs relatives, les coefficients de variation sont de 0,22 (en maths) et 0,20 (en physique), et dans la série complète de 0,23 et 0,19 (les notes s'écartent en moyenne de 0,22 fois la moyenne en maths et de 0,20 fois la moyenne en physique), c'est quasiment la même dispersion relative.

Un commentaire sur ces notes :

On ne va pas supposer des différences d'aptitudes différentes chez les élèves (mais cela peut exister individuellement) mais plutôt ici, une différence de notation des professeurs (c'est un parti-pris, quelque chose dont il faut se méfier en général car il peut orienter le jugement sur une fausse piste...). En maths, la notation est plus élevée, plus généreuse (plus de 2 points en moyenne c'est beaucoup) qu'en physique. On peut remarquer aussi que c'est essentiellement pour les meilleures notes que cet effet se fait ressentir, les moins bonnes moyennes sont quasiment égales dans les deux matières. Si on compare les écart-types, on s'aperçoit que la dispersion des notes est quasiment identique, si on se raisonne en valeurs relatives. Les valeurs absolues des écart-types donnent un écart moyen par rapport à la moyenne, or celle-ci est plus élevée en maths, donc un écart-type plus élevé est normalement attendu, même si les dispersions relatives étaient rigoureusement égales.

Prolongement : reporter les points de coordonnées $(m;p)$ où m est la note de maths et p la note de physique/chimie dans un graphique. Vous obtenez un nuage de points. Ce nuage vous semble-t-il traduire une liaison entre les deux notes moyennes? Que signifie cette liaison éventuelle?

Voici les résultats pour la série complète des 38 notes, le tableau de calcul et le nuage de points. Nous remarquons que ce nuage est très fortement étiré en forme de droite. Nous avons tracé cette droite que l'on appelle droite de régression (pas au programme!). L'équation de cette droite, la meilleure possible qui s'ajuste au mieux à la série est donnée par le logiciel $\text{physique} = 0,61 \times \text{maths} + 3,12$.

Un élève qui aurait 0 en maths aurait 3,12 en physique.

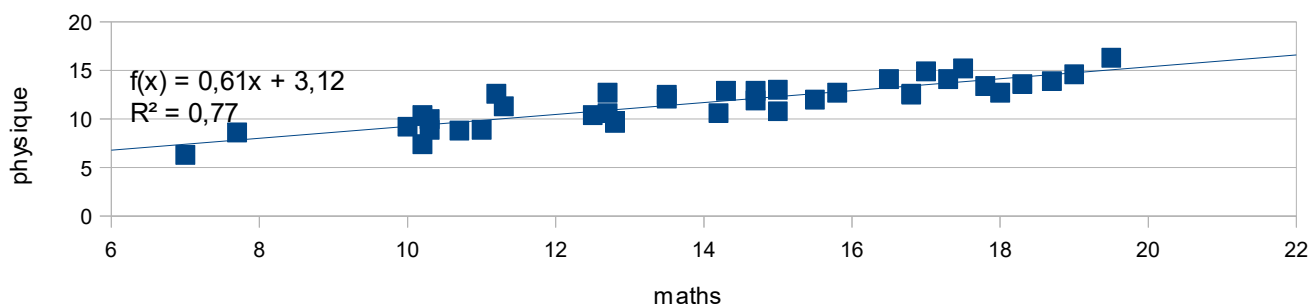
Un élève qui aurait 0 en physique aurait $\frac{-3,12}{0,61} \approx -5,11$ en maths...

Un élève qui aurait 20 en maths aurait $0,61 \times 20 + 3,12 = 15,32$ en physique.

Un élève qui aurait 20 en physique aurait $\frac{20-3,12}{0,61} \approx 27,67$ en maths...

Croisement des notes de maths et physique

1er trimestre 2013/2014

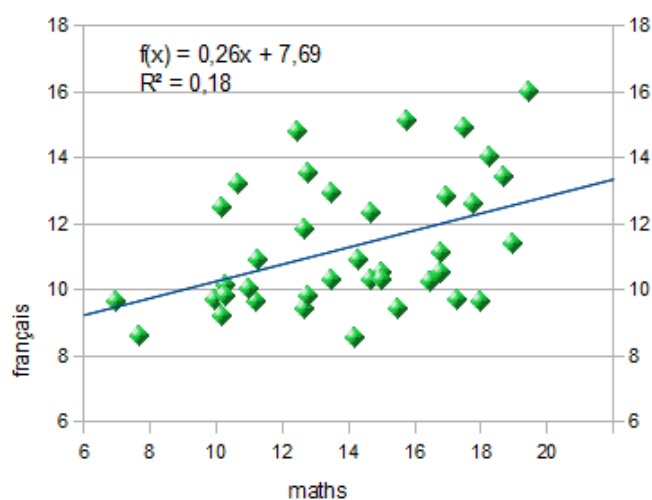


On mesure l'ajustement des données à cette droite par un coefficient appelé le coefficient de corrélation linéaire (pas, non plus, au programme!), noté R^2 . Ici ce coefficient est de 0,77 ce qui indique une bonne corrélation (une corrélation parfaite serait obtenue pour $R=1$, une absence de corrélation pour $R=0$).

Ce genre de comparaison peut se poursuivre à l'infini. On peut introduire d'autres matières, étudier des groupements de matières ou d'élèves... On peut aussi étudier les évolutions dans le temps... On pourrait écrire un livre entier (pas très intéressant sans doute) sur les statistiques relatives aux élèves d'une classe mais on s'arrêtera aujourd'hui à la même étude entre les notes de maths et de français. En français, la moyenne est un peu comme en physique (11,29) mais l'écart-type est plus faible (1,95), l'écart-type relatif vaut 0,17. Le croisement maths/français donne le nuage de points suivant : On voit un alignement des points mais celui-ci est beaucoup moins marqué. Cela ne se voit pas forcément très bien mais lorsqu'on cherche la droite de régression, on s'aperçoit que celle-ci est presque horizontale (français= $0,26$ maths+7,69) et surtout, le coefficient de corrélation est très faible (0,18) dénotant une quasi-indépendance des deux grandeurs. On aurait pu aussi bien tracer une droite complètement ou encore presque verticale, la corrélation n'aurait pas été beaucoup moins bonne. Le mieux à faire est sans doute de comparer les deux nuages avec des échelles identiques sur les axes.

Croisement des notes de maths et de français

1er trimestre 2013/2014



Croisement des notes de maths et physique

1er trimestre 2013/2014

