## Découvrir la notion de variance

### Introduction

L'objectif est de découvrir un nouveau paramètre de dispersion pour une série statistique.
A l'origine de ce paramètre, on trouve l'astronomie. Plaçons-nous au milieu du XVIII$^{e}$ siècle. A cette époque Copernic, Kepler et surtout Newton en ont fait une science étayée par une théorie mathématique. Mais le travail de l'astronome s'appuie sur des mesures et si celles-ci sont entachées d'erreurs, même si la théorie est bonne, le résultat ne sera pas bon. Or les enjeux de l'astronomie dépassent largement le fait de prévoir le passage de telle ou telle comète; ce sont des enjeux à la fois scientifiques, politiques et commerciaux : meilleure connaissance des étoiles et des planètes mais aussi meilleure connaissance de la Terre (dresser des cartes précises, permettre aux marins en mer de calculer correctement les longitudes, etc.).

D'où viennent ces erreurs? Il y a bien sûr des erreurs humaines mais aussi des erreurs provenant des instruments de mesure eux-mêmes. Or malgré l'amélioration de ceux-ci, et donc la précision accrue des mesures, des erreurs subsistent. Les scientifiques vont donc être amenés à s'intéresser aux erreurs. Il faut trouver des moyens de combiner de nombreuses mesures d'un même phénomène qui réduisent le plus possible l'erreur finale sur la vraie valeur du phénomène.

Le mathématicien, astronome et physicien allemand Carl Friedrich GAUSS (1777-1855) va mettre en place la méthode dite des moindres carrés et introduire un nouveau paramètre de dispersion : la variance.

### Etude d'un exemple

Un navire désire connaitre sa longitude. Plusieurs astronomes à bord réalisent les mesures nécessaires et obtiennent les résultats suivants (en degré) :

16,82 - 16,91 - 16,85 - 16,97 - 16,71 - 16,84 - 16,87 - 16,95 - 16,92 - 16,89
16,69 - 16,74 - 16,88 - 16,86 - 16,83 - 16,84 - 16,87 - 17,07

1) Quel nombre choisiriez-vous pour estimer la valeur réelle de la longitude du navire?

2) Proposer un (ou plusieurs) calcul(s) permettant d'estimer l'erreur commise sur les mesures. On pourra s'aider de l'extrait de l'ouvrage de Gauss :

*(extrait de la méthode des moindres carrés - 1821)*

*La question qui nous occupe a, dans sa nature même, quelque chose de vague et ne peut être bien précisée que par un principe jusqu'à un certain point arbitraire. La détermination d'une grandeur par l'observation peut se comparer, avec quelque justesse, à un jeu dans lequel il y aurait une perte à craindre et aucun gain à espérer: chaque erreur commise étant assimilée à une perte que l'on fait, la crainte relative à un pareil jeu doit s'exprimer par [...] la somme des diverses pertes possibles [...]. Mais quelle perte doit-on assimiler à une erreur déterminée? C'est ce qui n'est pas clair en soi; cette détermination dépend en partie de notre volonté. Il est évident, d'abord, que la perte ne doit pas être regardée comme proportionnelle à l'erreur commise; car, dans cette hypothèse, une erreur positive représentant une perte, l'erreur négative devrait être regardée comme un gain : la grandeur de la perte doit, au contraire, s'évaluer par une fonction de l'erreur dont la valeur soit toujours positive. Parmi le nombre infini de fonctions qui remplissent cette condition, il semble naturel de choisir la plus simple, qui est, sans contredit, le carré de l'erreur.*

3) Quelle est la nature de la fonction $V$ précédemment définie?

Visualiser l'allure de sa courbe représentative et déterminer graphiquement la valeur en laquelle le minimum de l'erreur est atteint. Comparer la réponse à celle de la question 1).

**Etude théorique**

1) Cas particulier d'une série de 3 valeurs

On considère une série statistique composée de 3 valeurs $x_1$, $x_2$ et $x_3$.

Soit $V$ la fonction définie sur $\mathbb{R}$ par $V(x) = \frac{1}{3}\left((x - x_1)^2 + (x - x_2)^2 + (x - x_3)^2\right)$

a) Mettre $V(x)$ sous forme développée.

b) Déterminer la valeur en laquelle le minimum de la fonction est atteint.

2) Cas général

On considère une série statistique $x_1$, $x_2$, …, $x_N$ de moyenne $\overline{x}$.

Soit $V$ la fonction définie sur $\mathbb{R}$ par $V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - x)^2$

a) Expliquer pourquoi $V$ est une fonction polynôme du second degré.

b) La forme développée de $V(x)$ peut s'écrire : $V(x) = ax^2 + bx + c$.

Justifier que l'on a : $a = 1$, $b = -2\overline{x}$ et $c = \frac{1}{N}\sum_{i=1}^{N} x_i^2$

En déduire que la moyenne est bien le nombre qui minimise la moyenne des carrés des écarts.