Devoir maison n°4

Statistiques

À rendre pour le jeudi 18 janvier.

Soit S une série statistique quantitative comportant N données : $S = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}$ et $d(x)$ la fonction qui à un réel x associe :

$$d(x) = \frac{1}{N}\left[(x-x_1)^2 + (x-x_2)^2 + \ldots + (x-x_N)^2\right]$$

On peut voir d comme la fonction qui, à un nombre x, associe la moyenne des distances de ce nombre à chacune des valeurs de la série, la distance entre deux nombres x et y étant définie ici comme $(x-y)^2$.
On cherche s'il existe une valeur centrale associée à cette distance, c'est-à-dire un nombre x_0 tel que $d(x_0)$ est le minimum de d, autrement dit un nombre qui serait le plus proche de toutes les valeurs de la série pour cette distance.

Cas particulier *Cette partie nécessite un tableur (Calc de la suite bureautique de LibreOffice, par exemple).*

Soit S la série constituée des moyennes des élèves d'une Première S en mathématiques au premier trimestre.
$S = \{7$; 9,8 ; 12,7 ; 11 ; 13,9 ; 6,3 ; 14,2 ; 12 ; 8,8 ; 5,9 ; 7,4 ; 16,9 ; 10,7 ; 11,2 ; 6,7 ; 11 ; 7,2 ; 12,9 ; 7,6 ; 10,3 ; 9,3 ; 7,3 ; 4,9 ; 10,5 ; 9,5 ; 6,1 ; 8,5 ; 12,2 ; 11,5 ; 17,1 ; 14,8 ; 9,1 ; 8,3 ; 5,1 $\}$

1. On commence avec $x = 10$. Entrer la valeur de x dans la première case du tableur (A1).
2. Entrer la série S dans la deuxième colonne qu'on nommera x_i (de B2 à B...).
3. Indiquer au tableur dans la troisième colonne, qu'on nommera $(x-x_i)^2$, le calcul à effectuer pour obtenir la distance du premier nombre de la série à x présent dans la case A1.
4. Copier coller ce calcul afin d'avoir dans la troisième colonne, les distances de chacun des nombres de la série à x.
5. En bas de cette colonne, indiquer au tableur le calcul à effectuer pour obtenir $d(x)$, la moyenne de ces distances.
   Vérifier que $d(10) \simeq 9{,}89$.
6. Recopier sur sa copie et compléter le tableau suivant (*les valeurs seront arrondies au centième*) :

x	-5	0	5	10	15	20	25
$d(x)$							

7. En tatonnant, déterminer la valeur de x, au dixième près, telle que $d(x)$ est minimun.
   Fournir le fichier avec sa copie.

Cas général

1. Montrer que $d(x) = x^2 - 2x\left(\frac{x_1+x_2+\ldots+x_N}{N}\right) + \frac{x_1^2+x_2^2\ldots+x_N^2}{N}$
2. En déduire que $d(x)$ admet un minimum.
3. Montrer que ce minimum est atteint en $x_0 = \overline{x}$, où $\overline{x}$ est la moyenne des valeurs de la série.
   Le vérifier dans le cas particulier du 6.3.4.
4. Montrer que $d(\overline{x}) = \frac{x_1^2+x_2^2\ldots+x_N^2}{N} - \overline{x}^2$.

Une application $d(\overline{x})$ est un indicateur statistique de dispersion d'une série quantitative. Plus il est élevé et plus les données sont dispersées. Il peut permettre de comparer la dispersion de deux séries.

Soit S' la série constituée des moyennes des élèves d'une autre Première S en mathématiques au premier trimestre :
$S' = \{14$; 14,3 ; 15,1 ; 12,9 ; 15,8 ; 10,8 ; 7,7 ; 15,4 ; 15,7 ; 10,9 ; 11 ; 13,7 ; 10,5 ; 14,5 ; 13,4 ; 10,8 ; 11,2 ; 13,7 ; 14 ; 11,5 ; 11,5 ; 15,8 ; 14,9 ; 12,7 ; 12,3 ; 14,7 ; 8,2 ; 12,4 ; 12,6 ; 12,3 ; 11 ; 14,5 ; 12,2 ; 13,4 ; 12,2$\}$

1. Calculer la moyenne $\overline{x}'$ de cette série et l'indicateur $d(\overline{x}')$ de cette série.
2. Comparer cette série à la précédente à l'aide de ces deux indicateurs statistiques.