

VALEURS CENTRALES
VALEURS DE DISPERSION

I / VALEURS CENTRALES

1- Définitions

Parmi les valeurs centrales qui permettent de décrire une série de chiffres, les diverses *moyennes* (géométrique ou arithmétique par exemple) sont les plus utilisées.

On rappelle que la formule qui permet de calculer la moyenne arithmétique d'une série $\{A_1, A_2, \dots, A_n\}$ est :

$$Moy(A) = \frac{\sum_{i=1}^n A_i}{n}$$

Une autre valeur centrale couramment utilisée est la *médiane*. Il s'agit du chiffre de la série tel qu'il y ait autant de valeurs de la série qui soient au dessous, et au dessus. Cette définition ne pose pas problème si le nombre de termes de la série est impair.

Par exemple, la médiane de $\{5, 7, 7, 9, 13, 18, 22\}$ est 9.

Si le nombre de termes de la série est pair, il n'y a plus de « médiane » au sens strict, mais un « intervalle-médian », dont les deux bornes sont les valeurs « du milieu » de la série. Par exemple, la série $\{5, 7, 7, 9, 13, 18, 22, 24\}$ a pour intervalle médian $[9, 13]$. On convient alors que la médiane de cette série est la moyenne des bornes de cet intervalle, soit ici, $(9+13)/2 = 11$.

2-Utilité

Pour ce qui est de la moyenne arithmétique, tout le monde en connaît l'usage et l'importance en tout domaine.

Pour ce qui est de la *médiane*, elle peut paraître assez peu intéressante du point de vue du mathématicien, mais a un intérêt réel dans la pratique du traitement de données historiques. En effet, contrairement à la moyenne, la médiane n'est pas sensible à des valeurs « aberrantes » qui se glisseraient dans la série, soit par caprice de la nature (?), soit surtout par erreur de saisie. La médiane est ainsi dite valeur centrale *robuste*. De plus, la médiane est une valeur concrète : elle fait effectivement partie de la série, et se repère plus qu'elle ne se calcule. Ainsi, il y bien des foyers de deux ou trois personnes, tandis qu'il est difficile d'imaginer un ménage composé de 3.05 ou de 2.87 personnes (qui sont des valeurs de moyennes) !

L'inconvénient de la médiane reste malgré tout les difficultés de manipulation, et en particulier la nécessité du tri de toutes les occurrences de la variables. Ainsi, pour notre tableau en feuille 2 faudrait-il écrire 3 935 100 fois 1, puis 4 936 840 fois 2 etc... Ce qui serait relativement fastidieux ! Il vaut bien mieux le faire « à la main ».

APPLICATION :

Calculer la moyenne de la première série de chiffres de la feuille 1.

Donner la valeur de la médiane de cette série. Remarque : on pourra utiliser l'outil « tri » présent dans la barre d'outils.

Calculer, à partir des données présentes sur la feuille 2, le nombre moyen de personnes dans un ménage en 1968, 1975, 1982, 1990 et 1999.

Même chose avec la médiane du nombre de personnes d'un ménage.

Recalculer les médianes en utilisant la fonction « MEDIANE » d'Excel : utiliser l' « Aide » pour des précisions.

II / VALEURS DE DISPERSION

1- Etendue, Ecart moyen

Les valeurs de dispersion se réfèrent bien souvent aux valeurs centrales d'une série, la plupart du temps à sa moyenne arithmétique. Elles expriment comment, en moyenne, les valeurs de la variable sont groupées autour de la valeur centrale. Elles permettent donc de juger si une série est homogène ou hétérogène.

L'indicateur le plus simple est *l'étendue* d'une série. Il s'agit de la différence entre sa valeur maximale et sa valeur minimale.

Un indicateur un peu plus complexe est *l'écart moyen* : il s'agit de la *moyenne des valeurs absolues des écarts des chiffres de la série à la moyenne*.

Pour le calculer, on effectue toutes les différences entre les chiffres de la série et la moyenne, on leur enlève leur signe, et on fait la moyenne de ces chiffres (tous positifs attention!).

Exemple : la série :

2, 5, 10, 15, 18

a pour moyenne : 10

pour écarts à la moyenne : -8, -5, 0, 5, 8

donc en valeur absolue : 8, 5, 0, 5, 8.

Enfin, la moyenne de ces chiffres est de $26/5 = 5,2$.

Cela signifie qu'en moyenne, les éléments de la série sont distants de 5.2 points par rapport à la moyenne arithmétique.

2- Variance et écart type

(source principale : Le pouvoir des données – statistiques Canada : http://www.statcan.ca/francais/edu/power/ch12/variance_f.htm)

a-Définitions

L'indicateur le plus utilisé en statistique est *l'écart type*, qui a des propriétés mathématiques plus intéressantes. Il est souvent noté σ , ou S et son carré, appelé *variance* est noté S^2 .

La *variance* se calcule sous la forme de *l'écart au carré moyen* de chaque nombre par rapport à la moyenne d'un ensemble de données. Pour les nombres 1, 2 et 3, par exemple, la moyenne est 2 et la variance, 0,667.

$$[(1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2] \div 3 = 0,667$$

Variance = [somme du carré des écarts à la moyenne] \div nombre d'observations

Variance, (S^2) = écart au carré moyen de valeurs par rapport à la moyenne

Comme le calcul de la variance se fait à partir des carrés des écarts, les unités de mesure ne sont pas les mêmes que celles des observations originales. Par exemple, les longueurs

mesurées en mètres (m) ont une variance mesurée en mètres carrés (m²). *La racine carrée de la variance nous donne les unités utilisées dans l'échelle originale.*

Écart-type (S) = Racine carrée de la variance

En résumé (et en formules...)

Variance d'une variable discrète x sur n observations:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n}$$

Ecart-type d'une variable discrète composée de n observations :

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

b-Utilité

L'écart-type est la mesure de dispersion la plus couramment utilisée en statistique lorsqu'on emploie la moyenne pour calculer une tendance centrale. Il mesure donc la dispersion autour de la moyenne. En raison de ses liens étroits avec la moyenne, l'écart-type peut être grandement influencé si cette dernière donne une mauvaise mesure de tendance centrale.

L'écart-type est aussi influencé par les valeurs aberrantes; une seule de ces valeurs pourrait avoir une grande influence sur les résultats de l'écart-type. Il s'agit donc d'un bon indicateur de l'existence de valeurs aberrantes, ce qui en fait une mesure de dispersion très utile pour les distributions symétriques ne comptant aucune valeur aberrante.

L'écart-type est aussi utile quand on compare la dispersion de deux ensembles de données séparés qui ont approximativement la même moyenne. La dispersion des mesures autour de la moyenne est plus étroite dans le cas d'un ensemble de données dont l'écart-type est plus petit. Habituellement, un tel ensemble renferme comparativement moins de valeurs élevées ou de valeurs faibles. Un élément sélectionné au hasard à partir d'un ensemble de données dont l'écart-type est faible peut se rapprocher davantage de la moyenne qu'un élément d'un ensemble de données dont l'écart-type est plus élevé.

Généralement, plus les valeurs sont largement distribuées, plus l'écart-type est élevé. Imaginez, par exemple, que nous devons séparer deux ensembles différents de résultats d'examens de 30 élèves; les notes du premier examen varient de 31 % à 98 % et celles du second, de 82 % à 93 %. Compte tenu de ces étendues, l'écart-type serait plus grand pour les résultats du premier examen.

Il n'est pas toujours facile d'évaluer l'importance que doit avoir l'écart-type pour que les données soient largement dispersées. L'importance de la valeur moyenne de l'ensemble des données dépend aussi de l'importance de l'écart-type. Lorsque vous mesurez quelque chose en millions, le fait d'avoir des mesures qui se rapprochent de la valeur moyenne n'a pas la même signification que si vous mesurez le poids de deux personnes. Par exemple, si après avoir

mesuré les recettes annuelles de deux grandes entreprises, vous constatez un écart de 10 000 \$, la différence est considérée comme étant peu significative, alors que si vous mesurez le poids de deux personnes, dont l'écart est de 30 kilogrammes, la différence est considérée comme étant très significative. Voilà pourquoi il est utile, dans la plupart des cas, d'évaluer quelle est l'importance de l'écart-type par rapport à la moyenne de l'ensemble de données.

c-Propriétés remarquables

Souvenez-vous des propriétés suivantes quand vous utilisez l'écart-type.

- On n'utilise l'écart-type que pour mesurer la dispersion autour de la moyenne d'un ensemble de données.
- L'écart-type n'est jamais négatif.
- L'écart-type est sensible aux valeurs aberrantes. Une seule valeur aberrante peut accroître l'écart-type et, par le fait même, déformer le portrait de la dispersion.
- Dans le cas des données ayant approximativement la même moyenne, plus la dispersion est grande, plus l'écart-type est grand.
- L'écart-type est zéro si toutes les valeurs d'un ensemble de données sont les mêmes (parce que chaque valeur est égale à la moyenne).

Quand on analyse des données normalement distribuées, on peut utiliser l'écart-type parallèlement à la moyenne pour calculer des intervalles de données.

Si \bar{x} = moyenne, S = écart-type et x = une valeur incluse dans l'ensemble de données, alors

- Environ 68 % des données x se situent à l'intérieur de l'intervalle :

$$\bar{x} - S < x < \bar{x} + S.$$

- environ 95 % des données se situent à l'intérieur de l'intervalle :

$$\bar{x} - 2S < x < \bar{x} + 2S.$$

- environ 99 % des données se situent à l'intérieur de l'intervalle :

$$\bar{x} - 3S < x < \bar{x} + 3S.$$

APPLICATION :

Calculer la moyenne de la deuxième série de chiffres de la feuille 1.

Faire apparaître les deux séries sur un même graphique. Qu'observe-t-on ?

Calculer l'étendue, l'écart moyen et l'écart type des deux séries de la feuille 1

Une fois qu'on a bien compris comment fonctionnent ces indicateurs, on peut utiliser les fonctions « ECART.MOYEN », « ECARTYPEP », et « VAR.P » pour la variance. On peut consulter l'aide d'Excel pour voir comment utiliser ces fonctions.

Calculer l'écart moyen et l'écart type pour les chiffres de la feuille 2 correspondant à l'année 1975.

Petit Tableau Récapitulatif : Les indicateurs statistiques de votre carnet de correspondance :

Voici les différentes questions que vous vous posez, ou que quelque parent attentif à vos efforts scolaires, pouvaient vous poser, pour juger (analyser) votre note brute, et brutale :

FAQ du petit parent illustré	Réponses du petit élève affolé - Analyse d'une mauvaise note...	Indicateurs statistiques correspondant
T'as eu combien ?	Note sur 20	Variable
Valeurs centrales : se situer par rapport à une « norme » pour la série		
Et c'est quoi la moyenne de la classe ?	Somme des notes sur nombre de notes	Moyenne
Combien d'élèves ont eu plus (moins) que toi ?	Tri de la série : évaluation de la position dans la série	Médiane (éventuellement quart etc.)
Valeurs de dispersion : connaître la situation moyenne de chacun par rapport à cette norme (judicieux donc la pertinence de ces « normes »)		
C'est quoi, la meilleure (plus mauvaise) note ?	Evaluation générale de l'extension de la série	Etendue
(pour les parents un peu critiques, et bien informés...)		
Il y en a beaucoup qui ont une très mauvaise ou très bonne note ? Ou bien tout le monde est-il autour de la moyenne ?	Moyenne des écarts à la moyenne Ou encore mieux : Moyenne des carrés des écarts à la moyenne !	Ecart moyen Ecart type (variance)