Aucun article ou résumé dans cette revue ne peut être reproduit sous forme d'imprimé, photocopie, microfilm ou par tout autre procédé sans l'autorisation expresse des auteurs et de l'éditeur. Editions Phlébologiques Françaises
No article or abstract in this journal may be reproduced in the form of print, photocopy, microfilm or any other means without the express permission of authors and the editor. Editions Phlébologiques Françaises

Phlébologie 2010, 63, 2, p. 87-88

Comment décrire statistiquement une population de patients ?

How to describe statistically a patients population?

Allaert F.A.

Lorsque l'on s'intéresse à une population ou à un échantillon de patients présentant une pathologie donnée, il est fondamental de déjà bien savoir la décrire. Deux types de paramètres sont disponibles en statistiques : les paramètres de position et les paramètres de dispersion.

Les paramètres de position

Les paramètres de positions sont ceux que nous utilisons les plus communément pour résumer les caractéristiques de la distribution d'une population.

La médiane correspond à la valeur qui divise la série des individus en deux groupes égaux. Ainsi, si l'on dit que la médiane du délai de reperméabilisation post-sclérothérapie est de 3 ans, cela signifie que la moitié des patients sont recanalisés avant 3 ans et la moitié après 3 ans.

Les quartiles sont les 3 valeurs qui divisent la distribution en 4 parties égales. Le premier quartile marque la limite supérieure des 25 % des données les plus faibles, le deuxième quartile correspond à la médiane décrite ci-dessus et le troisième quartile désigne la limite inférieure des 25 % des données les plus élevées.

Les déciles correspondent au même principe mais divisent la distribution de la population de 10 % en 10 %.

Le mode est la valeur de la distribution qui est la plus fréquente et on l'appelle souvent aussi le pic de la distribution.

La moyenne qui est le plus connu de ces paramètres de position est la somme algébrique de toutes les données de la distribution divisée par le nombre de sujets de la population.

Mais à quoi servent tous ces éléments et pourquoi ne peut-on pas se contenter de la moyenne seulement par exemple ? La réponse réside dans le fait qu'une moyenne sans autre information peut induire en erreur sur la répartition générale des caractéristiques d'une population.

Certes lorsque les distributions sont symétriques et tendent vers la loi Normale, la médiane, la moyenne et le mode sont confondus en une seule et même valeur mais très souvent ce n'est pas le cas comme le montrent les **Figure 1**, **Figure 2** et **Figure 3** : ils montrent des situations très différentes.

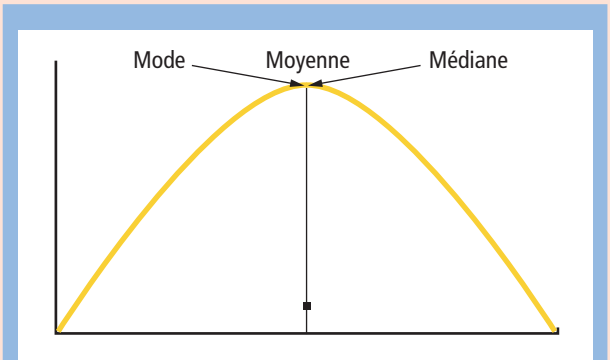

FIGURE 1 : Distribution symétrique.

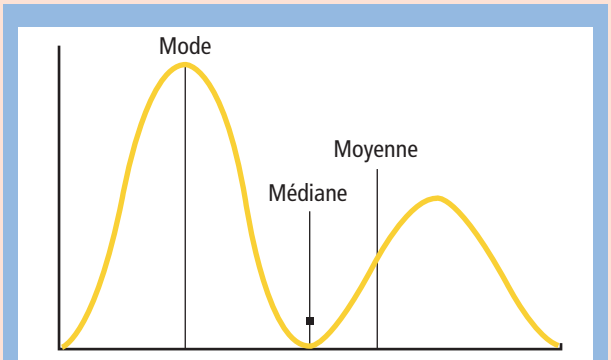

FIGURE 2 : Distribution polymodale.

Professeur titulaire de la Chaire d'Évaluation des Allégations de Santé, Ceren ESC & DIM, CHRU, 21000 Dijon, France.

Aucun article ou résumé dans cette revue ne peut être reproduit sous forme d'imprimé, photocopie, microfilm ou par tout autre procédé sans l'autorisation expresse des auteurs et de l'éditeur. Editions Phlébologiques Françaises
No article or abstract in this journal may be reproduced in the form of print, photocopy, microfilm or any other means without the express permission of authors and the editor. Editions Phlébologiques Françaises

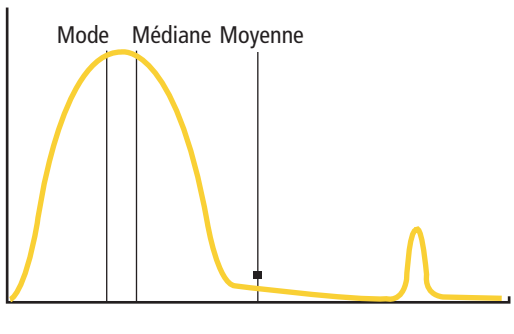

FIGURE 3 : Distribution décentrée.

En particulier, comme l'illustre la **Figure 3**, la moyenne est sensible au poids des valeurs extrêmes et quelques valeurs très excentrées sur la droite vont la tirer vers celles-ci du fait qu'elles sont très élevées, alors que la médiane ne va guère être influencée par elles.

Par exemple pour l'ancienneté de la maladie veineuse, un sujet chez qui elle a débuté il y a 36 mois va influencer la moyenne 36 fois plus qu'un sujet chez qui elle a débuté il y a un mois. Mais par contre, les deux sujets pèsent le même poids pour la médiane : qu'un sujet soit juste au dessus d'elle ou beaucoup au-dessus d'elle, il ne compte toujours que pour un ! Quand il existe des extrêmes importants, la médiane décrit souvent mieux la distribution que la moyenne.

Les paramètres de dispersion

La moyenne à elle seule ne permet pas d'appréhender la dispersion des valeurs et que des distributions très différentes peuvent avoir des moyennes identiques comme l'illustre la **Figure 4**.

Il en est d'ailleurs de même pour la médiane. Pour résumer la dispersion des informations, les statisticiens ont inventé le concept de variance qui correspond à la moyenne des carrés des écarts de chaque valeur à la moyenne.

L'inconvénient de ce paramètre est qu'il est un carré et donc qu'il correspond à une unité au carré, ce qui n'est pas d'un usage aisé.

En prenant sa racine carrée, on a ce que l'on appelle l'écart-type qui a l'avantage d'être dans la même unité de mesure que la moyenne. Cet écart-type mesure la dispersion de la distribution autour de la moyenne.

Plus il est élevé, plus la dispersion est grande et à l'inverse, plus il est petit, plus la distribution est resserrée autour de la moyenne.

Lorsque les populations sont assez grandes pour que l'on puisse considérer que les variables tendant vers la loi normale ou gaussienne (la fameuse courbe en « cloche »), 95 % des valeurs qui peuvent être prises par la variable étudiée se situent entre la moyenne moins 2 fois l'écart-type et la moyenne plus deux fois l'écart-type **(Figure 5)**.

On comprend ainsi pourquoi une moyenne doit toujours être présentée avec son écart type pour pouvoir être interprétée correctement.

Enfin, on appelle **étendue** d'une distribution la différence entre les deux extremums, c'est-à-dire les valeurs minimales et maximales de la distribution de la population.

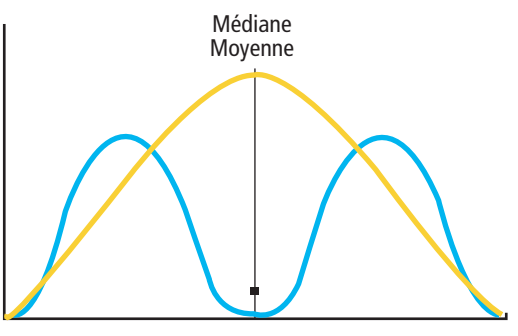

FIGURE 4 : Deux populations très différentes, mais même moyenne et médiane.

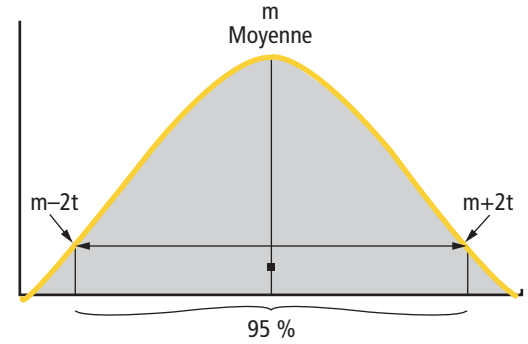

FIGURE 5 : Distribution de la population en fonction de la moyenne m et de l'écart-type t.