

Séance 3 : Les paramètres statistiques élémentaires

C'est le caractère quantitatif qui est le plus général. Il s'applique sur une donnée statistique qui prend en compte un nombre de valeurs précises. En variable quantitative, cela peut correspondre à la taille des membres d'un club de volley-ball par exemple. C'est un caractère plus général puisqu'il n'indique pas précisément ce qu'il représente mais seulement une quantité, donc on ne connaît pas la valeur en tant que telle derrière. Les caractères quantitatifs discrets servent à compter. Ils s'utilisent dans des décomptes et ne fonctionnent que pour certaines valeurs. Par exemple pour compter le nombre de touches défaillantes sur piano : la valeur sera forcément entre 0 et 88 puisqu'un piano compte 88 touches. En revanche les caractères quantitatifs continus servent à mesurer, donc il est obligatoire de mesurer la valeur au préalable. On se sert alors d'outils, comme la moyenne, la somme, l'écart-type, etc. Il existe plusieurs types de moyenne pour pouvoir calculer des populations statistiques de différentes nature. Par exemple, on ne peut pas utiliser une moyenne « classique » (c'est-à-dire la moyenne arithmétique) pour faire une moyenne sur une population statistique représentant des vitesses. Il faut alors faire une moyenne harmonique, moyenne qui ne prend pas en compte le zéro. Calculer une médiane sert à diviser la série statistique en deux en l'ayant rangée par ordre croissant ou décroissant. On peut savoir grâce à la médiane que 50% des valeurs de la série sont supérieures à la médiane et 50% inférieures à la médiane. Il est possible de calculer un mode quand une valeur revient fréquemment dans la série statistique. On connaît alors la densité maximale. La médiane est calculée relativement aux valeurs globales. De plus les valeurs de la médiane combinées aux valeurs de la médiane servent à calculer la concentration. L'intérêt de l'indice de Gini ou coefficient de Gini permet de rendre compte de la concentration d'une population statistique. Il s'applique sur un repère orthonormé. Il est souvent utilisé pour décrire la répartition des richesses au sein d'un pays. La variance permet de calculer au mieux la dispersion dans une série statistique. En revanche l'écart-type est plus efficace puisqu'il caractérise la dispersion en pouvant comprendre si la série statistique est plus ou moins homogène. L'étendue permet de connaître la différence entre les valeurs extrêmes traitées, donc pour connaître les valeurs extrêmes de la série statistique étudiée. Les quantiles servent à diviser une série statistique ordonnée en parties égales. En divisant la série statistique en quatre, on crée des quartiles, en cinq des quintiles, en dix les déciles... Les quartiles sont les quantiles les plus utilisés puisque le deuxième quartile correspond à la médiane. Les quartiles sont schématiquement visibles lors de la construction d'une boîte de dispersion ou boîte à moustache. La boîte de dispersion est avant tout une représentation graphique d'un caractère quantitatif. On peut donc interpréter plus visuellement la médiane, ainsi que la répartition en quartiles et déciles. Un moment est une sorte de fonction qui permet de bien comprendre comment fonctionne la variable. Le moment centré est une variable aléatoire qui a une espérance nulle. Le moment absolu concerne la densité de la probabilité de la variable. Les moments servent à s'intéresser à la dispersion de la variable. Vérifier la symétrie d'une distribution statistique permet de savoir si elle n'est pas dissymétrique (c'est-à-dire positive ou négative). On utilise alors les coefficients de Pearson et de Fisher.