

### Séance 3

Le caractère le plus général est le caractère quantitatif.

Les paramètres statistiques se rapportent essentiellement aux variables quantitatives, tout en abordant occasionnellement les variables qualitatives. Cela illustre que la dimension quantitative est la plus universelle, puisqu'elle permet de définir l'ensemble des paramètres statistiques tels que la position, la dispersion et la forme. En revanche, la dimension qualitative ne permet que d'établir certains paramètres de façon restreinte.

Un caractère quantitatif discret se caractérise par des valeurs distinctes et dénombrables. Les paramètres statistiques tels que la moyenne, la variance ou la médiane sont déterminés à partir de sommes.

En revanche, un caractère quantitatif continu peut prendre n'importe quelle valeur dans un intervalle réel. Les paramètres sont alors calculés à l'aide d'intégrales, par exemple pour la moyenne, la médiane ou les moments.

La distinction entre ces deux types de caractères est essentielle, car les méthodes de calcul des paramètres statistiques diffèrent selon que la variable est discrète ou continue.

Il existe différents types de moyenne : la moyenne arithmétique, la moyenne quadratique, la moyenne harmonique, la moyenne géométrique, et la moyenne mobile.

La médiane est la valeur, observée ou possible, dans la série des données classées par ordre croissant (ou décroissant) qui partage cette série en deux parties comprenant exactement le même nombre de données de part et d'autre de cette valeur. Une médiane permet de Remarque 1 La médiane n'est pas influencée par les valeurs extrêmes (ou aberrantes), à la différence de la moyenne arithmétique. Remarque 2 Par contre, la médiane est influencée par le nombre des données. Elle ne peut ainsi être utilisée en théorie de l'estimation. La médiane est déterminée par le classement des valeurs, et non par les valeurs extrêmes, donc elle résume bien des distributions fortement dissymétriques.

Il est possible de calculer une mode quand il y a une modalité ou valeurs qui ressort le plus (la plus fréquente) dans une série statistique. Il est aussi possible de la calculer quand il y a deux ou plusieurs modalités.

La médiale correspond à la valeur qui divise en deux parties égales la masse totale des valeurs globales (et non les effectifs). Elle est utile pour analyser la répartition effective de la variable étudiée, comme par exemple dans le cas de la répartition des salaires.

L'indice de concentration, quant à lui, se base sur la comparaison entre la médiale et la médiane. Plus l'écart entre ces deux valeurs est grand par rapport à l'étendue, plus la concentration est élevée.

"La variance tient compte de toutes les données; il s'agit de la meilleure caractéristique de dispersion"

"La variance est l'indicateur de dispersion par excellence, mais, exprimé dans la même unité que la moyenne, il est souvent plus pratique d'utiliser l'écart type."

L'étendue est intéressante à calculer car elle permet de voir si la série statistique est concentrée (étendue faible) ou au contraire étendue.

Plus l'étendue d'une série est grande, plus la série est hétérogène. - Plus l'étendue est petite, plus la série est homogène.

Elle permet ainsi de voir la variation entre deux valeurs extrêmes (valeur minimum et valeur maximum). Et ainsi de voir l'écart entre ces deux valeurs.

Les quantiles sont des caractéristiques de position partageant la série statistique ordonnée en k parties égales"

Ainsi créer un quantile permet de voir les probabilité qu'un groupe de données apparaisse dans le jeu de données. Ils servent à diviser une série statistique ordonnée en plusieurs parties ayant le même effectif. et à identifier des repères représentatifs de la distribution. Par exemple, en partageant la série en quatre parties égales ( $k = 4$ ), on obtient les quartiles  $Q_1$ ,  $Q_2$  et  $Q_3$ , le deuxième quartile  $Q_2$  correspondant à la médiane. L'écart interquartile, défini par  $Q_3 - Q_1$ , contient 50 % des valeurs et permet de mesurer la dispersion de la moitié centrale des observations.

Les quantiles ont différents noms selon le nombre de parts dans la population. Ainsi les quantiles les plus utilisées sont la médiane ( $k = 2$ ), les quartiles, les décile et les centiles.

La différence entre les moments centrés et les moments absolus réside dans les valeurs utilisées. En effet

Les moments centrés sont calculés à partir des écarts à la moyenne (par exemple  $(x - \bar{x})$ ). Ils servent à caractériser la dispersion, L'asymétrie et l'aplatissement d'une distribution. Les moments absolus utilisent la valeur absolue des écarts par rapport à un point donné et mesurent • l'importance moyenne des écarts sans tenir compte du signe

Les moments permettent de caractériser complètement une distribution statistique. Par exemple :

- le moment centré d'ordre 2 correspond à la variance,
- les moments d'ordre 3 et 4 servent à mesurer l'asymétrie et l'aplatissement

Une distribution symétrique se caractérise par l'égalité entre la moyenne, la médiane et le mode. La symétrie est vérifiée à l'aide du coefficient d'asymétrie  $B_1$  :