

Question de cours de la séance 3 :

1. Quel caractère est le plus général : le caractère quantitatif ou le caractère qualitatif ?

Justifier pourquoi.

2. Que sont les caractères quantitatifs discrets et caractères quantitatifs continus ? Pourquoi les distinguer ?

3. Paramètres de position

— Pourquoi existe-t-il plusieurs types de moyenne ?

— Pourquoi calculer une médiane ?

— Quand est-il possible de calculer un mode ?

4. Paramètres de concentration

— Quel est l'intérêt de la médiane et de l'indice de C. Gini ?

5. Paramètres de dispersion

— Pourquoi calculer une variance à la place de l'écart à la moyenne ? Pourquoi la remplacer par l'écart type ?

— Pourquoi calculer l'étendue ?

— À quoi sert-il de créer un quantile ? Quel(s) est (sont) le(s) quantile(s) le(s) plus utilisé(s) ?

— Pourquoi construire une boîte de dispersion ? Comment l'interpréter ?

6. Paramètres de forme

— Quelle différence faites-vous entre les moments centrés et les moments absolus ?

Pourquoi les utiliser ?

— Pourquoi vérifier la symétrie d'une distribution et comment faire ?

1. Le caractère qualitatif est le plus général. Le cours montre que tout caractère quantitatif peut être ramené à une forme qualitative par discrétisation (classes), alors que l'inverse n'est pas toujours possible. Les caractères qualitatifs permettent de décrire des catégories sans supposer de mesure numérique préalable. Ainsi, le caractère quantitatif constitue un cas particulier du caractère qualitatif, auquel s'ajoute une structure numérique permettant le calcul de paramètres statistiques.

2. Les caractères quantitatifs discrets prennent un nombre fini ou dénombrable de valeurs distinctes (par exemple des notes, des effectifs). Les caractères quantitatifs continus peuvent prendre toutes les valeurs d'un intervalle réel (par exemple une distance, une durée). Le cours insiste sur la nécessité de les distinguer car les outils statistiques diffèrent : une variable discrète se traite par des sommes ; une variable continue se traite par des intégrales. Cette distinction conditionne le calcul des moyennes, médianes, variances et représentations graphiques.

3. Il existe plusieurs types de moyenne car toutes les variables ne répondent pas aux mêmes contraintes mathématiques ni aux mêmes objectifs d'analyse. Le cours montre que la moyenne arithmétique, bien que centrale, est sensible aux valeurs extrêmes. D'autres moyennes (harmonique, géométrique, quadratique, mobile) permettent de s'adapter à la nature des données et à la question posée. La pluralité des moyennes reflète donc la diversité des situations statistiques.

La médiane permet de partager une population en deux parties de même effectif. Le cours souligne qu'elle est robuste aux valeurs extrêmes, contrairement à la moyenne. Elle est particulièrement adaptée aux distributions dissymétriques et rend compte d'une position centrale indépendante des valeurs aberrantes.

Le mode est calculable lorsqu'une modalité présente un effectif ou une densité maximale. Le cours précise qu'il peut ne pas exister ou ne pas être unique (distributions bimodales ou plurimodales). Il est surtout pertinent pour les variables discrètes ou les distributions présentant des pics nets.

4. La médiane partage la masse totale de la variable en deux parties égales, et non les effectifs. Elle permet donc d'analyser la répartition réelle d'un caractère (revenus, salaires, surfaces, etc.). L'indice de concentration de C. Gini, fondé sur la courbe de Lorenz, permet de mesurer le degré d'inégalité dans la distribution. Plus la courbe s'éloigne de la diagonale, plus la concentration est forte. Ces outils sont essentiels pour analyser les inégalités spatiales.

5. La variance est calculée à partir des carrés des écarts à la moyenne, car les écarts simples s'annulent algébriquement. Le carré possède en outre des propriétés mathématiques utiles.

Cependant, la variance est exprimée dans une unité au carré ; l'écart type, qui en est la racine carrée, permet de revenir à l'unité de la variable et facilite l'interprétation.

L'étendue mesure la différence entre la valeur maximale et la valeur minimale. Le cours précise qu'elle est facile à calculer mais peu robuste, car elle ne repose que sur les valeurs extrêmes. Elle donne néanmoins une première idée de l'amplitude globale de la distribution.

Les quantiles servent à partager une série ordonnée en parts égales d'effectifs. Les plus utilisés sont : les quartiles (Q1, Q2, Q3), les déciles, les centiles. Ils permettent de décrire la structure interne d'une distribution et sont à la base de nombreux indicateurs de dispersion.

La boîte de dispersion (ou boîte à moustaches) permet de représenter graphiquement les principales caractéristiques d'une distribution : médiane, quartiles, valeurs extrêmes. Elle facilite la ****comparaison visuelle de plusieurs séries**** et permet d'identifier la dispersion, la dissymétrie et la présence éventuelle de valeurs atypiques.

6. Les moments non centrés sont calculés par rapport à l'origine, tandis que les moments centrés sont calculés par rapport à la moyenne. Les moments centrés permettent de caractériser la dispersion, la dissymétrie et l'aplatissement d'une distribution. Ils sont essentiels pour décrire la forme globale d'une loi statistique. Vérifier la symétrie permet de mieux comprendre la structure de la distribution et d'évaluer la pertinence de certains indicateurs. Le cours indique que l'on peut utiliser le coefficient d'asymétrie β_1 : $\beta_1 = 0$: distribution symétrique ; $\beta_1 > 0$: dissymétrie positive ; $\beta_1 < 0$: dissymétrie

négative. Dans une distribution parfaitement symétrique, la moyenne, la médiane et le mode sont confondus.