

Questions cours 2

1. Positionnement de la Géographie par rapport aux Statistiques

La relation entre la géographie et la statistique est souvent tendue et complexe, voire paradoxale. Bien que la géographie soit une discipline qui génère des données massives (comme la démographie, l'occupation des sols, etc.) qui nécessitent l'outil statistique pour être étudiées, elle a traditionnellement eu tendance à mépriser ou négliger les définitions et les méthodes mathématiques élémentaires de la statistique. Cette situation est due au fait que l'analyse mathématique n'entrait pas historiquement dans son champ disciplinaire. Pourtant, l'étude des phénomènes géographiques actuels est indissociable de l'analyse statistique des données qu'ils produisent.

2. Le concept de Hasard en Géographie

D'un point de vue philosophique, l'idée de hasard pur est souvent rejetée, car on considère qu'il existe une cause à tout phénomène.

Toutefois, dans les modélisations mathématiques et statistiques appliquées à la géographie, le hasard est un concept opérationnel qui se décline en deux catégories :

- Le hasard bénin : Il suit généralement une distribution de probabilité normale (ou gaussienne).
- Le hasard sauvage : Il correspond à des distributions de probabilité moins fréquentes, souvent caractérisées par des événements extrêmes.

Historiquement, dès le début du XXe siècle, les géographes ont intégré deux grandes lois de probabilité pour décrire la répartition des phénomènes : la loi normale et la loi de Vilfredo Pareto.

3. Types d'Information Géographique

L'information géographique pouvant faire l'objet d'une analyse statistique se divise en deux séries principales :

1. L'étude des caractéristiques : Pour une entité territoriale donnée (un pays, une région, etc.), il s'agit d'étudier toutes les données pouvant la caractériser, qu'elles relèvent de la géographie humaine (population, activités...) ou de la géographie physique (relief, climat...).
2. L'étude de la morphologie : Il s'agit d'analyser la géométrie même des ensembles délimités (taille, forme, densité des réseaux...). La statistique s'applique également à l'étude de la forme spatiale des phénomènes.

4. Besoins de la Géographie en Analyse de Données

L'analyse de données, qui repose sur les probabilités et les statistiques, permet d'étudier la structure interne des données produites par la géographie. Elle répond à un besoin fondamental :

- Confronter les résultats obtenus avec la méthodologie utilisée pour la production des données.
- Valider et enrichir la connaissance que l'on a déjà du phénomène géographique étudié.

5. Types de Visualisation de Données Géographiques

Le choix de la méthode de visualisation dépend de la nature des données à analyser :

- Visualisation Quantitative : Utilisée pour l'analyse de données numériques, souvent appliquée dans le cadre d'une Analyse Factorielle en Composantes Principales (AFC).
- Visualisation Qualitative : Utilisée pour l'analyse de données catégorielles, elle est typiquement appliquée dans le cadre d'une Analyse Factorielle des Correspondances (AFC) ou d'une Analyse Factorielle des Correspondances Multiples (AFCM).
- Visualisation Mixte : De nombreux logiciels statistiques offrent des options permettant de visualiser simultanément des variables qualitatives et quantitatives.

6. Statistiques Descriptives vs. Statistiques Explicatives

Il est essentiel de distinguer les deux grandes branches de l'analyse statistique :

Caractéristique	Statistique Descriptive	Statistique Explicative
Objectif Principal	Résumer, visualiser, classer et comprendre les grandes dimensions d'un phénomène.	Relier une variable "à expliquer" (ou réponse) à des variables "explicatives".
Rôle des Variables	Toutes les variables (les k variables du tableau individus) jouent le même rôle. Pas de variable à expliquer.	Vise à ajuster un modèle (régression, logistique, etc.) dont la forme dépend de la nature de la variable réponse (numérique, catégorielle...).
Application Typique	Analyse factorielle, étude des distributions, calcul des moyennes et des écarts-types.	Modélisation, inférence, prévision (ex. : prévoir la croissance urbaine en fonction de facteurs socio-économiques).

7. Méthodes d'Analyse de Données

Les méthodes d'analyse de données se classent en trois grandes catégories :

- Méthodes Descriptives : Viser à synthétiser et à représenter l'information contenue dans les données (cf. point 6).
- Méthodes Explicatives : Viser à modéliser les relations de cause à effet entre variables (cf. point 6).
- Méthodes de Prévision : Viser à anticiper des valeurs futures ou l'occurrence de phénomènes (ex. : séries temporelles).

8. Définitions des Concepts de Base

- Population Statistique : C'est l'ensemble (au sens mathématique) que l'on étudie. Elle peut être spatiale (ex. : les habitants d'un territoire) ou non spatiale (ex. : le personnel d'une entreprise).

- Individu Statistique (ou Unité Statistique) : C'est un élément de la population statistique.
 - Particularités en Géographie :
 - Les individus statistiques sont souvent localisables et cartographiables (unités spatiales).
 - Ils sont fréquemment composés eux-mêmes d'un ensemble d'éléments de niveau inférieur (ex. : une commune est composée d'habitants, de points de vente, etc.).
- Caractère Statistique (ou Variable Statistique) : C'est la propriété que l'on étudie sur les individus.
 - Types de Caractères :
 - Qualitatif : Décrit une qualité, une catégorie (ex. : couleur, type d'habitat).
 - Quantitatif : Représente une quantité mesurable (ex. : âge, revenu, superficie).
 - Note : Il n'existe pas de hiérarchie intrinsèque entre ces deux types de caractères.
- Modalités Statistiques : Ce sont les valeurs possibles prises par un caractère. Elles doivent être incompatibles (disjointes) et exhaustives (couvrir toutes les possibilités) pour former une partition du caractère et permettre de caractériser l'appartenance de chaque individu.

9. Mesure de l'Amplitude et de la Densité

Ces mesures sont utilisées après la discrétisation des caractères quantitatifs (c'est-à-dire le regroupement des valeurs en classes) :

- Amplitude (A) : C'est la longueur d'une classe (intervalle). Elle se calcule comme la différence entre sa valeur maximale (b) et sa valeur minimale (a) : $A=b-a$.
- Densité (d) : Elle correspond au rapport entre l'effectif de la classe et son amplitude (A). La densité permet de comparer des classes qui n'ont pas la même longueur.

10. Formules de Sturges et de Yule

Ces formules servent de guides pour organiser les données en classes (regroupement de valeurs) :

- Formule de Sturges : Elle donne une valeur approximative du nombre de classes à créer.
- Formule de Yule : Elle permet de calculer l'amplitude des classes en fonction de l'étendue totale des observations (différence entre la valeur maximale et minimale de la série) et du nombre de classes choisi.

11. Effectif, Fréquence et Distribution Statistique

- Effectif : C'est le nombre d'apparitions d'une variable ou d'une modalité donnée au sein de la population étudiée.

- Fréquence Relative : C'est le rapport entre l'effectif d'une modalité et l'effectif total de la population. Elle s'exprime souvent en pourcentage et se calcule à partir d'une fonction.
- Fréquence Cumulée : Pour une modalité k , c'est la somme des effectifs (ou des fréquences) associés aux valeurs du caractère qui sont inférieures ou égales à k . Elle s'applique principalement aux variables quantitatives.
- Distribution Statistique Empirique : C'est la manière dont se répartissent les effectifs ou les fréquences entre les différentes modalités du caractère étudié. Son analyse permet de se prononcer sur le type de loi de probabilité qui pourrait modéliser le phénomène.