

Questions de cours de la séance 2 :

1. Quel est le positionnement de la géographie par rapport aux statistiques ?
2. Le hasard existe-t-il en géographie ?
3. Quels sont les types d'information géographique .
4. Quels sont les besoins de la géographie au niveau de l'analyse de données ?
5. Quelles sont les différences entre la statistique descriptive et la statistique explicative ?
6. Quelles sont les types de visualisation de données en géographie ? Comment choisir celles-ci ?
7. Quelles sont les méthodes d'analyse de données possibles ?
8. Comment définiriez-vous : (a) population statistique ? (b) individu statistique ? (c) caractères statistiques ? (d) modalités statistiques ? Quels sont les types de caractères ? Existe-t-il une hiérarchie entre eux ?
9. Comment mesurer une amplitude et une densité ?
10. À quoi servent les formules de Sturges et de Yule ?
11. Comment définir un effectif ? Comment calculer une fréquence et une fréquence cumulée ? Qu'est-ce qu'une distribution statistique ?

1. La géographie entretient une relation ancienne, souvent ambiguë, avec les statistiques. Longtemps, la discipline s'en est méfiée, considérant que la rigueur mathématique s'accordait mal avec la complexité et la diversité des phénomènes territoriaux. Pourtant, cette position est paradoxale : la géographie produit et mobilise des données massives, qu'il est impossible de comprendre sans outils statistiques. Le cours insiste sur cette tension historique entre une géographie parfois rétive aux mathématiques et la nécessité, aujourd'hui incontournable, de l'analyse statistique. Les statistiques ne sont pas un corps étranger à la géographie : elles constituent au contraire un outil essentiel pour transformer l'information géographique en connaissance scientifique, notamment dans une discipline qui se définit comme une science des échelles.
2. Le hasard en géographie n'est pas nié, mais il est pensé. Le cours montre que deux visions s'opposent : une vision déterministe, pour laquelle le hasard n'est qu'un bruit négligeable, et une vision contingente, dans laquelle les phénomènes peuvent se produire ou non. En géographie, il est impossible de prévoir précisément chaque événement ou chaque comportement individuel. En revanche, il est possible de dégager des tendances globales, des régularités statistiques. Le hasard existe donc à l'échelle des réalisations individuelles, mais il devient structurable à une échelle plus large. Cette approche rejoint le raisonnement multiscalaire : ce qui est imprévisible localement peut devenir intelligible globalement.

3. Le cours distingue deux grands types d'information géographique. D'une part, les informations attributaires, qui décrivent les caractéristiques humaines ou physiques d'un territoire (population, variables sociales, économiques, climatiques, etc.). D'autre part, les informations géométriques, qui concernent la forme, la morphologie et la géométrie des objets spatiaux. Dans un système d'information géographique (SIG), ces deux types sont complémentaires : les premières constituent la base attributaire, tandis que les secondes relèvent de la structure spatiale elle-même.
4. La géographie a besoin de l'analyse de données pour structurer, résumer et interpréter l'information géographique. Face à la massification des données, le géographe ne peut rester passif : il doit comprendre les conditions de production des données, leur nomenclature, leurs méta-données, puis analyser leur structure interne. L'analyse de données permet ainsi de dégager des tendances, d'identifier des relations entre phénomènes et d'évaluer la fiabilité des résultats. Elle ne remplace pas le raisonnement géographique, mais elle le complète, en offrant une vision synthétique de phénomènes complexes.
5. La statistique descriptive vise à décrire et résumer les données. Elle cherche à mettre de l'ordre dans l'information à l'aide d'indicateurs numériques et de représentations graphiques. Elle fournit une image simplifiée de la réalité et constitue une étape préalable indispensable. La statistique explicative, quant à elle, cherche à établir une relation entre une variable à expliquer (Y) et une ou plusieurs variables explicatives (X). Elle repose sur des modèles statistiques et vise à comprendre, voire prédire, un phénomène. Ainsi, la première décrit, tandis que la seconde explique.
6. Les visualisations dépendent étroitement du type de variables étudiées. Le cours mentionne notamment : l'histogramme pour les variables quantitatives continues ; les diagrammes en secteurs pour les variables qualitatives ; les diagrammes en bâtons pour les variables quantitatives discrètes ; les représentations issues des analyses factorielles (ACP, AFC, ACM). Le choix d'une visualisation repose donc sur la nature des données, leur distribution et l'objectif de l'analyse. Une mauvaise visualisation peut produire une interprétation erronée : le graphique n'est jamais neutre.
7. Le cours distingue trois grandes familles de méthodes : les méthodes descriptives, notamment les analyses factorielles (ACP, AFC, ACM) et les classifications ; les méthodes explicatives, comme les régressions, l'analyse de la variance ou la régression logistique ; les méthodes de prévision, appliquées aux séries chronologiques. Ces méthodes permettent respectivement de visualiser, expliquer et anticiper les phénomènes étudiés.
8.
 - a) La population statistique est l'ensemble des unités étudiées.
 - b) L'individu statistique est un élément de cette population ; en géographie, il est souvent spatialisé.
 - c) Les caractères statistiques sont les caractéristiques observées sur les individus.
 - d) Les modalités statistiques sont les valeurs prises par ces caractères.

Les caractères peuvent être qualitatifs (nominal ou ordinal) ou quantitatifs (discret ou continu). Il n'existe pas de hiérarchie absolue entre eux, mais leur nature conditionne les traitements statistiques possibles.

9. L'amplitude correspond à la longueur d'une classe statistique, calculée comme la différence entre la borne supérieure et la borne inférieure. La densité est le rapport entre l'effectif d'une classe et son amplitude. Elle permet de comparer des classes de tailles différentes et joue un rôle essentiel dans la construction des histogrammes.
10. Les formules de Sturges et de Yule servent à estimer le nombre de classes lors de la discrétisation d'une variable quantitative. Elles permettent d'éviter un découpage trop fin ou trop grossier, qui entraînerait une perte d'information. Ce sont des outils d'aide à la décision, non des règles absolues.
11. Un effectif est le nombre d'occurrences d'une valeur ou d'une modalité. La fréquence est le rapport entre cet effectif et l'effectif total. La fréquence cumulée s'obtient en additionnant les fréquences jusqu'à une valeur donnée. Une distribution statistique est l'organisation des effectifs ou des fréquences d'une variable, permettant d'en analyser la structure et, éventuellement, d'identifier une loi de probabilité sous-jacente.