

## Séance 2

### Questions de cours :

La géographie occupe une position complexe vis-à-vis des sciences statistiques : se considérant comme une science-humaine depuis plusieurs décennies, elle refuse de traiter les données statistiques comme une partie à part entière. Pour autant, la géographie produit un nombre conséquent de données, qu'il est alors nécessaire d'analyser. La géographie doit donc travailler avec la statistique. Le hasard n'existe pas en tant que tel en géographie, ce sont davantage les notions de contingence, c'est-à-dire qui peut être autrement, et de nécessité, c'est-à-dire qui ne peut pas ne pas être autrement qui rentrent dans le domaine d'analyse. On ne peut pas établir de certitude mais on peut dégager une tendance générale contingente. Il y a deux types d'information géographique déclinés en séries statistiques possibles : celle qui s'appuie sur l'ensemble délimité par des éléments de géographie humaine ou physique et celle qui étudie directement les éléments en tant que tel donc qui s'intéresse à la morphologie de ces éléments. L'analyse de données en géographie repose sur les statistiques et sur les probabilités. Une fois l'étude de la structure interne de ces données faites, il faut confronter les résultats obtenus avec la méthodologie de production des données ainsi qu'avec le phénomène étudié. La statistique descriptive sert à étudier et analyser les données collectées. Le but est de dégager des tendances générales ou remarquables par rapport à une distribution théorique connue. On met ainsi de l'ordre dans les données. On rend alors compte de distribution graphique et on peut élaborer des comparaisons et des prédictions. La statistique explicative permet de créer une information systématique, de traiter l'information créée ou encore de connaître la fiabilité de l'information mais elle ne remplacera jamais un raisonnement explicatif, la culture ou les connaissances. La statistique reste complémentaire, elle n'est en aucun cas substituable. Les différents types de visualisation de données en géographie rentrent dans les méthodes descriptives de l'analyse de données. Le type de visualisation varie en fonction de la nature des données qu'on souhaite étudier. L'analyse factorielle en composantes principales permet de visualiser des données quantitatives. L'analyse factorielle des correspondances est utilisée pour des données qualitatives. Elle permet de visualiser un tableau de contingence. L'analyse des proximités permet de construire une carte des objets étudiés. On retrouve trois types de méthode en analyse de données : les méthodes descriptives, les méthodes explicatives et les méthodes de prévision. Les méthodes descriptives s'appliquent à des tableaux individuels avec des variables et permettent de résumer le tableau de variables et de comprendre les grandes dimensions du phénomène étudié. Les méthodes explicatives cherchent à relier une variable à expliquer à des variables explicatives. Les méthodes de prévisions concernent l'analyse et la prévision d'une série chronologique. La population statistique correspond à un ensemble au sens mathématique. En géographie cela désigne souvent l'ensemble des données qu'on doit traiter lors d'une étude comme le nombre d'habitants d'un territoire par exemple. Un individu statistique (ou unité statistique) correspond à un élément de la population statistique comme un arbre au sein d'une forêt par exemple. Les caractères statistiques sont les particularités de l'individu statistique pris parmi la population statistique. Les modalités statistiques sont les valeurs prises par un caractère. Elles sont incompatibles et exhaustives. Les caractères peuvent être de deux natures différentes, elles-mêmes subdivisées en deux sous-catégories : les variables qualitatives (nominales ou ordinales) et les variables quantitatives (discrètes ou continues). Ces caractères s'appliquent à des types de données différentes, il n'y a pas véritablement de hiérarchie entre eux. L'amplitude correspond à la différence entre la valeur maximale de la classe et la valeur minimale de la classe. La densité correspond au rapport entre l'effectif et l'amplitude de la classe ce qui décrit une modalité. Les formules de Sturges et de Yule permettent de calculer approximativement le nombre de classe selon le nombre d'observations effectuées et l'étalement des données. L'effectif quand il est associé à une valeur de variable aléatoire correspond au nombre d'apparitions de cette variable.

dans la population statistique. La fréquence se calcule par le rapport entre l'effectif et l'effectif total (la somme de tous). La fréquence cumulée jusqu'à  $k$  modalités est la somme des effectifs associés aux valeurs du caractère qui sont inférieures ou égales à  $k$ . Une distribution statistique sert à établir une « probabilité réelle » où l'on pourra conclure sur le type de loi de probabilité utilisée.