

Raphaël JODER (étudiant étranger)

M1 GAED

GéoSuds

Rapport analyse de données en géographie

Parcours : débutants

Séance 2 - Les principes généraux de la statistique

1. Quel est le positionnement de la géographie par rapport aux statistiques ?

La géographie utilise les statistiques comme outil d'analyse pour décrire, comparer et expliquer les phénomènes spatiaux (démographie, économie, mobilités, environnement).

Elle ne se limite pas au calcul : elle interprète les résultats dans l'espace, en tenant compte des contextes territoriaux, sociaux, politiques et culturels. La statistique est donc un moyen, pas une fin.

2. Le hasard existe-t-il en géographie ?

Oui, mais de manière relative. Les phénomènes géographiques sont déterminés par des logiques structurelles (économie, politiques publiques, environnement), mais ils comportent aussi une part d'aléa (catastrophes naturelles, comportements individuels, événements imprévus). La statistique permet de modéliser cette incertitude.

3. Quels sont les types d'information géographique ?

- Information qualitative (type de sol, catégorie socio-professionnelle)
- Information quantitative (population, revenu, densité)
- Information spatiale (localisation, distance, surface)
- Information temporelle (évolution, dynamique)
- Information relationnelle (flux, réseaux)

4. Quels sont les besoins de la géographie au niveau de l'analyse de données ?

- Décrire des territoires
- Comparer des espaces
- Expliquer des dynamiques spatiales
- Mettre en évidence des inégalités
- Modéliser des processus territoriaux
- Aider à la décision publique (aménagement, politiques)

5. Quelles sont les différences entre la statistique descriptive et la statistique explicative ?

Statistique descriptive : résume et décrit les données (moyennes, graphiques, cartes)

Statistique explicative (ou inférentielle) : cherche à expliquer, généraliser, tester des hypothèses et établir des relations entre variables

6. Quels sont les types de visualisation de données en géographie ? Comment choisir ceux-ci ?

- Cartes (choroplèthes, cartes de flux, anamorphoses)
- Tableaux statistiques
- Boîtes à moustaches

Le choix dépend :

- du type de variable (qualitative/quantitative)
- de l'objectif (comparaison, distribution, relation)
- de l'échelle spatiale

7. Quelles sont les méthodes d'analyse de données possibles ?

Analyse descriptive

Analyse multivariée (ACP, CAH)

Analyse spatiale (autocorrélation, clustering spatial)

Statistiques inférentielles

Modélisation

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 ?

- Population statistique : ensemble étudié (ex. : communes du Costa Rica)
- Individu statistique : unité d'observation (ex. : une commune)
- Caractère statistique : variable observée (ex. : population)
- Modalités : valeurs prises par le caractère (ex. : urbain/rural)

Types de caractères :

- Qualitatifs (nominal / ordinal)
- Quantitatifs (discret / continu)

9. Comment mesurer une amplitude et une densité ?

Amplitude : différence entre la valeur maximale et minimale d'une classe

Densité : rapport entre effectif et amplitude (utile pour histogrammes)

10. A quoi servent les formules de Sturges et de Yule ?

Elles servent à déterminer le nombre optimal de classes dans une distribution statistique.

11. Comment définir un effectif ? Comment calculer une fréquence et une fréquence cumulée ? Qu'est-ce qu'une distribution statistique ?

Effectif : nombre d'individus par modalité

Fréquence : effectif / effectif total

Fréquence cumulée : somme progressive des fréquences

Distribution statistique : organisation des données selon leurs modalités

Séance 3 - Les paramètres statistiques élémentaires

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

Le caractère quantitatif, car il permet des calculs mathématiques et statistiques plus poussés.

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

Pourquoi les distinguer ?

Discret : valeurs entières (nombre d'enfants)

Continu : valeurs infinies possibles (revenu, superficie)

3. Paramètres de position

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

Car chaque moyenne (arithmétique, médiane, géométrique) répond à des contextes différents et à des distributions parfois asymétriques.

❖ Pourquoi calculer une médiane ?

Elle est robuste aux valeurs extrêmes et représente mieux la tendance centrale en cas d'inégalités.

❖ Quand est-il possible de calculer un mode ?

Pour les variables qualitatives ou discrètes.

4. Paramètres de concentration

❖ Quel est l'intérêt de la médiale et de l'indice de C. Gini ?

Ils mesurent la concentration et les inégalités spatiales ou sociales (revenus, population).

5. Paramètres de dispersion

❖ Pourquoi calculer une variance à la place de l'écart à la moyenne ?

Pourquoi la remplacer par l'écart type ?

Variance : mesure théorique

Écart type : plus lisible car exprimé dans l'unité de la variable

❖ **Pourquoi calculer l'étendue ?**

Pour une première mesure simple de la dispersion.

❖ **A quoi sert-il de créer un quantile ? Quel(s) est (sont) le(s) quantile(s) le(s) plus utilisé(s) ?**

Ils découpent la population en parts égales. Les plus utilisés : quartiles, déciles, percentiles.

❖ **Pourquoi construire une boîte de dispersion ? Comment l'interpréter ?**

Elle permet de visualiser :

- dispersion
- asymétrie
- valeurs aberrantes

6. Paramètres de forme

❖ **Quelle différence faites-vous entre les moments centrés et les moments absolus ? Pourquoi les utiliser ?**

Absolus : liés aux valeurs brutes

Centrées : liées à la moyenne

Utiliser pour analyser la forme d'une distribution

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

Elle permet d'identifier les inégalités.

Méthodes : comparaison moyenne/médiane, coefficient d'asymétrie.

Séance 4 - Les distributions statistiques

1. Quels critères mettriez-vous en avant pour choisir entre une distribution statistique avec des variables discrètes et une distribution statistique avec des variables continues ?

Critères :

- nature des données
- précision
- taille de l'échantillon
- objectif de l'analyse

2. Expliquez selon vous quelles sont les lois les plus utilisées en géographie ?

Lois les plus utilisées en géographie

Loi normale (phénomènes sociaux)

Loi log-normale (revenus)

Loi de Pareto (concentration)

Loi binomiale / Poisson (événements rares)

Séance 5 - Les statistiques inférentielles

1. Comment définir l'échantillonnage. Pourquoi ne pas utiliser la population en entier ? Quelles sont les méthodes d'échantillonnage ? Comment les choisir ? \$

Sélection d'une partie de la population pour en tirer des conclusions générales.

Pourquoi pas toute la population :

- coût
- temps
- accès limité

Méthodes :

- aléatoire simple
- stratifié
- systématique
- par grappes

2. Comment définir un estimateur et une estimation ?

Estimateur : formule

Estimation : valeur obtenue

3. Comment distingueriez-vous l'intervalle de fluctuation et l'intervalle de confiance ?

Fluctuation : variabilité attendue

Confiance : encadrement du paramètre inconnu

4. Qu'est-ce qu'un biais dans la théorie de l'estimation ?

Erreur systématique faussant l'estimation.

5. Comment appelle-t-on une statistique travaillant sur la population totale ? Faites le lien avec la notion de données massives (*big data*) ?

Lien avec le big data, mais attention aux biais et à la qualité des données.

6. Quels sont les enjeux autour du choix d'un estimateur ?

Les enjeux sont les biais, la précision et la robustesse.

7. Quelles sont les méthodes d'estimation d'un paramètre ? Comment en sélectionner une ?

Méthode des moments, et le maximum de vraisemblance.

8. Quels sont les tests statistiques existants ? À quoi servent-ils ? Comment créer un test ?

Ils servent à valider ou rejeter une hypothèse.

Étapes : hypothèses → statistique de test → seuil → décision.

9. Que pensez-vous des critiques de la statistique inférentielle ?

- simplification excessive
- dépendance aux hypothèses

- déconnexion du terrain

Séance 6 - La statistique d'ordre des variables qualitatives

- 1. Qu'est-ce qu'une statistique ordinale ? À quel autre statistique catégorielle s'oppose-t-elle ? Quel type de variables utilise-t-elle ? En quoi cela peut matérialiser une hiérarchie spatiale ?**

Statistique basée sur un ordre, opposée au nominal.

Variables : rangs, niveaux.

Elle permet de matérialiser des hiérarchies spatiales.

- 2. Quel ordre est à privilégier dans les classifications ?**

Ordre logique, cohérent avec le phénomène étudié.

- 3. Quelle est la différence entre une corrélation des rangs et une concordance de classements ?**

Corrélation : lien entre deux variables ordonnées

Concordance : accord entre classements

- 4. Quelle est la différence entre les tests de Spearman et de Kendal ?**

Spearman : plus sensible aux valeurs extrêmes

Kendall : plus robuste

5. À quoi servent les coefficients de Goodman-Kruskal et de Yule ?

Ils mesurent l'association entre variables qualitatives.