SORBONNE UNIVERSITE

M1 GAED- Parcours ENVITERR

Analyse de données

M.FORRIEZ

CHIBANI Sihem

N°étudiant : 21206828

Décembre 2025

Séance 2

I- Question de cours

La géographie est très proche et paradoxale avec les statistiques. Bien qu'elle produise une quantité importante de données, la discipline a longtemps sous-estimé l'apport et l'intérêt des méthodes statistiques, notamment en raison d'une formation mathématique limitée chez de nombreux géographes. Cette situation a parfois conduit à des usages imprécis des outils statistiques. Toutefois, les apports récents de données et le développement de l'analyse spatiale ont rendu ces méthodes incontournables. La géographie s'appuie désormais largement sur les statistiques pour analyser, représenter et modéliser les phénomènes spatiaux, et pour dégager des régularités à différentes échelles.

La place du hasard en géographie s'inscrit au centre de la réflexion philosophique et de la pratique scientifique. Si certaines approches préfèrent une vision déterministe des phénomènes, la pratique statistique reconnaît l'existence de l'aléa. Il est impossible de prévoir précisément les comportements individuels, mais des tendances globales peuvent être mises en évidence. On distingue ainsi un hasard dit bénin, souvent modélisé par la loi normale, et un hasard dit sauvage, associé à des événements extrêmes décrits par des lois à queue lourde comme celle de Pareto. Cette prise en compte de l'aléa permet à la géographie d'expliquer des régularités générales malgré des variations locales imprévisibles.

L'information géographique se décline principalement en deux formes complémentaires : les informations attributaires, qui décrivent les caractéristiques sociales, économiques ou physiques des objets spatiaux, et les informations géométriques, qui concernent leur localisation, leur forme et leur étendue. Dans un système d'information géographique, ces deux types de données sont liés et permettent une analyse spatiale cohérente.

L'analyse des données géographiques repose sur plusieurs étapes essentielles. Elle nécessite d'abord la définition de nomenclatures claires et la disponibilité de métadonnées détaillées afin de garantir la fiabilité des données. Les données sont ensuite résumées à l'aide de statistiques descriptives, puis analysées à l'aide de modèles explicatifs ou prédictifs fondés sur des lois de probabilité adaptées. Des méthodes de réduction de dimension ou de classification peuvent également être mobilisées. L'interprétation des résultats doit toujours être confrontée aux conditions de collecte des données et aux connaissances du terrain.

Dans ce cadre, la statistique descriptive joue un rôle exploratoire en synthétisant les données à travers des indicateurs et des représentations graphiques, tandis que la statistique explicative ou inférentielle vise à comprendre et à modéliser les relations entre les variables, notamment à l'aide de régressions ou de tests statistiques. Ces deux approches sont complémentaires.

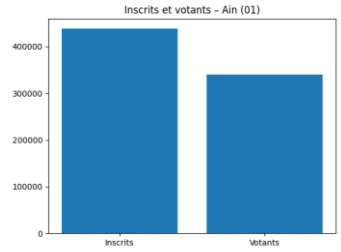
La visualisation des données constitue un outil central en géographie. Le choix des représentations dépend à la fois de la nature des variables (qualitatives, quantitatives, discrètes ou continues) et des objectifs de l'analyse. Histogrammes, diagrammes en bâtons,

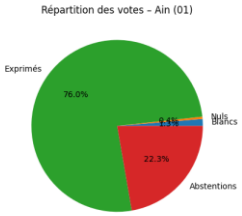
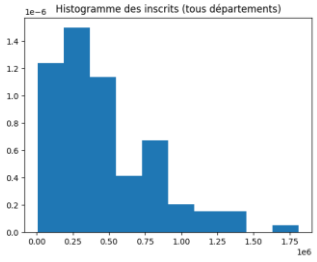
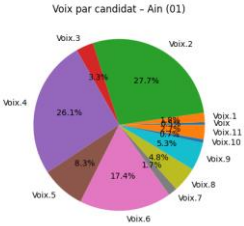
camemberts, boîtes à moustaches ou graphiques issus d'analyses multivariées permettent de décrire les distributions, comparer des groupes ou mettre en évidence des structures spatiales.

Les méthodes d'analyse de données peuvent être regroupées en trois grandes catégories : les méthodes descriptives et multidimensionnelles, les méthodes explicatives et les méthodes de prévision, notamment à travers l'analyse de séries temporelles. Elles sont généralement utilisées de manière complémentaire au cours d'une même étude.

Enfin, l'analyse statistique repose sur des notions fondamentales telles que la population statistique, les individus, les caractères et leurs modalités. Les caractères peuvent être qualitatifs ou quantitatifs, discrets ou continus, sans hiérarchie de valeur entre eux. La structuration des données passe aussi par la construction de distributions statistiques, l'usage d'indicateurs comme les effectifs, fréquences et fréquences cumulées, et par le choix de classes adaptées, guidé notamment par les formules de Sturges ou de Yule. Ces outils constituent la base de toute analyse statistique rigoureuse en géographie.

II- Mise en œuvre avec Python

Numéro de la question	Application	Illustration
Question 4 à 10	Affichage lignes et colonnes	
Question 11	Diagrammes en barre par départements	

Question 12	Diagrammes circulaires par département	
Question 13	Histogramme de la distribution des inscrits	
Bonus	Voix par candidat	

Découverte à tâtons, j'ai rapidement compris qu'à la moindre erreur/faute de frappe le code ne marcherait tout simplement pas.

Le principal problème de la séance venait de moi : j'ai passé environ 2h sur tous les forums, contacté tous mes amis pour comprendre pourquoi le code ne voulait pas se lancer avec "docker-compose run python", j'avais simplement oublié d'enregistrer mon main.py...

Séance 3

I- Question de cours

En statistique, le caractère qualitatif est généralement considéré comme le plus général. Il permet de classer les individus selon des catégories, sans nécessiter de mesure numérique, comme le type de logement, la couleur des yeux ou le genre. À l'inverse, le caractère quantitatif correspond à une valeur mesurable, telle que l'âge, la taille ou le revenu. On peut considérer que le quantitatif constitue un cas particulier du qualitatif, dans la mesure où une variable quantitative peut toujours être transformée en variable qualitative par un regroupement en classes, par exemple en tranches d'âge. L'opération inverse n'est en revanche pas possible.

Parmi les caractères quantitatifs, on distingue les variables discrètes et continues. Les variables discrètes ne prennent que des valeurs isolées, souvent entières, comme le nombre d'enfants ou le nombre de pièces dans un logement. Les variables continues, quant à elles, peuvent prendre n'importe quelle valeur dans un intervalle réel, comme la température, le poids ou la durée. Cette distinction est importante car ces deux types de variables ne se traitent pas de la même manière : les variables discrètes se comptent, tandis que les variables continues se mesurent et nécessitent souvent un regroupement en classes pour être analysées ou représentées graphiquement.

Pour résumer une distribution, on utilise des paramètres de position, parmi lesquels figurent différentes formes de moyennes. Il existe plusieurs types de moyennes car aucune ne permet, à elle seule, de représenter correctement toutes les situations. La moyenne arithmétique est adaptée à des valeurs homogènes, la moyenne pondérée est utilisée lorsque certaines observations ont plus de poids que d'autres, et les moyennes géométriques ou harmoniques sont plus pertinentes dans des contextes spécifiques, comme l'étude des taux de croissance ou des vitesses. La médiane constitue une mesure complémentaire essentielle : elle partage la population en deux groupes égaux et présente l'avantage d'être peu sensible aux valeurs extrêmes, ce qui en fait une mesure de tendance centrale plus robuste. Le mode, enfin, correspond à la valeur ou à la modalité la plus fréquente ; il peut être calculé pour tous les types de variables, mais il est particulièrement utile pour les données qualitatives et les variables discrètes.

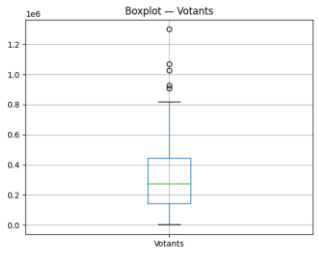
L'analyse statistique s'intéresse également à la concentration et à la dispersion des données. La médiale permet d'apprécier la concentration en divisant une distribution en deux parties égales, tandis que l'indice de Gini mesure le degré d'inégalité au sein d'une distribution. Un indice proche de zéro indique une répartition très égalitaire, alors qu'un indice proche de 1 traduit une forte concentration, comme c'est souvent le cas pour la répartition des revenus ou des patrimoines.

Concernant la dispersion, la variance est préférée à l'écart moyen à la moyenne car ce dernier tend à s'annuler lorsque les écarts positifs et négatifs se compensent. En élevant les écarts au carré, la variance évite ce problème. L'écart type, qui correspond à la racine carrée de la variance, est souvent privilégié car il s'exprime dans la même unité que la variable étudiée et se révèle donc plus facile à interpréter. L'étendue, définie comme la différence entre la valeur maximale et la valeur minimale, donne une première idée de la dispersion globale, mais elle reste très sensible aux valeurs extrêmes et doit être interprétée avec prudence.

Les quantiles constituent un autre outil fondamental pour analyser la répartition des données. Ils permettent de découper une distribution en parts égales afin d'étudier la position relative des valeurs. Les plus couramment utilisés sont la médiane, les quartiles, mais aussi les déciles et les centiles pour des analyses plus détaillées. La boîte à moustaches, (boxplot), synthétise ces informations en un seul graphique : elle met en évidence la médiane, les quartiles, l'étendue interquartile et les éventuelles valeurs aberrantes. Elle facilite ainsi la comparaison entre plusieurs distributions et permet de repérer rapidement les asymétries.

Enfin, l'étude de la forme des distributions repose sur les moments statistiques. Les moments centrés prennent en compte les écarts à la moyenne et permettent de caractériser la variance, l'asymétrie et l'aplatissement de la distribution. Les moments absolus, qui ne tiennent pas compte du signe des écarts, complètent cette analyse. Vérifier la symétrie d'une distribution est essentiel, car elle conditionne la relation entre la moyenne, la médiane et le mode. Une distribution symétrique présente des valeurs proches pour ces trois indicateurs, tandis qu'une distribution asymétrique montre des écarts caractéristiques selon qu'elle est orientée vers la droite ou vers la gauche. Cette symétrie peut être évaluée à l'aide de graphiques ou d'indicateurs numériques dédiés.

II- Mise en œuvre avec Python

Numéro de la question	Réponse	Illustration
Question 1 à 7	Affichage et calcul	
Question 8	Boite à moustache	
Question 9 et 10	Ouverture du ficher	

Bonus :	Voir les fichiers en .csv et .xlsx dans le dossier /images de la séance	

Pas de réel problème pour cet exercice. On m'a conseillé d'automatiser la création du fichier "images", j'ai donc importé "import os" à chaque séance pour faciliter le tout.

Séance 4

I- Question de cours

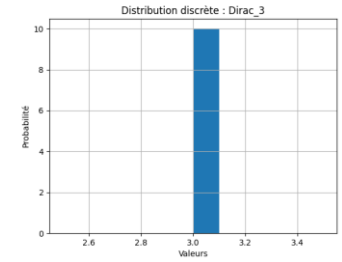
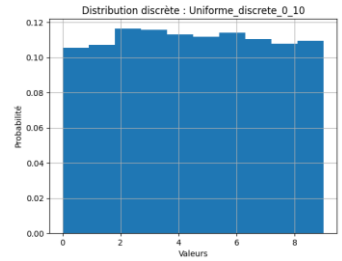
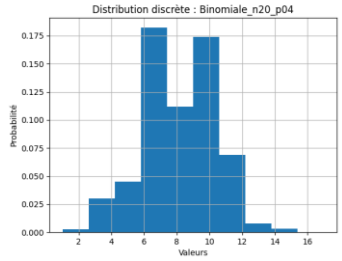
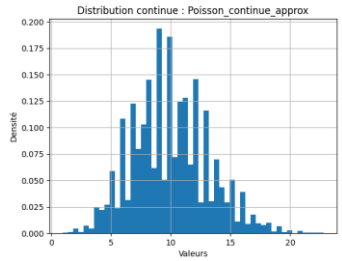
Le choix entre une distribution statistique à variables discrètes ou continues repose sur la nature de la variable étudiée. Lorsqu'une variable ne peut prendre que des valeurs entières et dénombrables, comme un nombre d'habitants ou un nombre d'événements, elle est dite discrète. Ce type de variable se prête bien à des lois de comptage telles que les lois de Bernoulli, binomiale ou de Poisson. À l'inverse, lorsque la variable peut prendre n'importe quelle valeur dans un intervalle continu, par exemple l'altitude, la température, le revenu ou la distance, elle est considérée comme continue et se modélise plus volontiers à l'aide de lois comme la loi normale, la loi log-normale, l'exponentielle ou la Weibull.

L'aspect pratique de la distribution constitue également un critère important. Une distribution qui apparaît sous forme de valeurs isolées ou en "marches" renvoie généralement à une variable discrète, tandis qu'une distribution lisse, décrite par une densité continue, correspond à une variable continue. Le choix de la loi dépend aussi du processus à l'origine du phénomène observé. Les phénomènes résultant de l'addition de nombreux effets indépendants suivent souvent une loi normale, les processus de croissance multiplicative tendent vers une loi log-normale, et les comptages d'événements rares dans l'espace sont fréquemment modélisés par une loi de Poisson. Enfin, l'échelle d'observation et la précision des données doivent être prises en compte, car une variable continue arrondie ou mesurée grossièrement peut donner l'illusion d'un comportement discret.

En géographie, certaines lois statistiques sont particulièrement mobilisées car elles permettent de rendre compte des structures spatiales, des hiérarchies et des dynamiques territoriales. La loi de Zipf, ou loi rang-taille, est largement utilisée pour analyser la hiérarchie urbaine et la distribution des tailles de villes, en mettant en évidence les fortes inégalités entre grandes et petites agglomérations. La loi log-normale, également appelée loi de Gibrat, est fréquemment employée pour modéliser des phénomènes de croissance proportionnelle, comme l'évolution des populations, des revenus ou des entreprises. La loi de Pareto, caractéristique des distributions à queue lourde, décrit des situations très inégalitaires dans lesquelles une minorité d'unités concentre une part importante des ressources. La loi normale, enfin, intervient dans de nombreux contextes dès lors que les phénomènes étudiés résultent d'un grand nombre de petites influences indépendantes, comme certaines variations naturelles ou les erreurs de mesure. La loi de Poisson complète cet ensemble en permettant de modéliser la répartition spatiale d'événements rares et indépendants, tels que des accidents ou des occurrences environnementales.

Ces lois jouent un rôle central en géographie car elles permettent de relier les formes observées sur le terrain (comme la répartition, concentration ou hiérarchie) aux mécanismes statistiques et spatiaux qui les produisent.

II- Mise en œuvre avec Python

Numéro de la question	Réponse	Illustration
Question 1 (L'entièreté des graphiques sont visibles dans le dossier "images" de la séance")	La loi Dirac	
	Loi Uniforme Discrète	
	Loi Binomiale	
	Distribution continue de la Loi Poisson	
Question 2	Calcul de la moyenne et de l'écart-type des variables discrètes et des variables continues	

Pour cette séance, j'ai eu du mal à importer "scipy.stats", j'ai dû forcer avec :

from scipy.stats import (

poisson,

binom,

zipf,

randint,

norm,

lognorm,

uniform,

chi2,

pareto

)

Pour que cela fonctionne correctement.

Séance 5

I- Question de cours

L'échantillonnage consiste à étudier une partie d'une population afin d'en tirer des conclusions sur l'ensemble. On y a recours parce qu'une étude exhaustive est souvent irréaliste, soit pour des raisons de coût, de temps, soit parce que la population est trop vaste ou difficile à observer dans sa totalité. L'enjeu principal est donc de construire un échantillon suffisamment représentatif pour limiter l'erreur liée à l'échantillonnage.

Il existe deux grandes familles de méthodes. Les méthodes aléatoires, comme le tirage au sort simple ou stratifié, avec ou sans remise, reposent sur le hasard et offrent de bonnes garanties théoriques de représentativité. Les méthodes non aléatoires, telles que l'échantillonnage systématique, par quotas ou certaines approches de type Monte Carlo, sont davantage utilisées lorsque la base de sondage est incomplète ou lorsque des contraintes pratiques s'imposent. Le choix de la méthode dépend avant tout des objectifs de l'étude, du niveau de précision recherché et des moyens disponibles.

Dans ce cadre, un estimateur désigne une fonction mathématique construite à partir des données de l'échantillon et destinée à approcher un paramètre inconnu de la population, comme une moyenne ou une proportion. L'estimation correspond à la valeur numérique obtenue à partir de cet estimateur. Par exemple, la moyenne de l'échantillon est un estimateur de la moyenne de la population, et le résultat calculé constitue l'estimation.

Il est important de distinguer l'intervalle de fluctuation de l'intervalle de confiance. L'intervalle de fluctuation décrit les variations attendues d'une fréquence observée lorsque la proportion théorique est connue, et permet d'apprécier les effets du hasard d'échantillonnage. L'intervalle de confiance, en revanche, sert à estimer un paramètre inconnu en fournissant une plage de valeurs dans laquelle ce paramètre a une forte probabilité de se situer.

En théorie de l'estimation, le biais correspond à l'écart entre la valeur moyenne fournie par un estimateur et la vraie valeur du paramètre. Un estimateur est dit sans biais lorsque cet écart est nul ; dans le cas contraire, il est biaisé. Un bon estimateur doit idéalement être sans biais, précis, convergent et présenter une variance faible.

Lorsqu'une statistique porte sur l'ensemble de la population, on parle de statistique exhaustive. Cette situation limite le recours à l'inférence, puisque les paramètres sont directement observés. Avec le développement des données massives, certaines analyses se rapprochent de cette logique, même si l'inférence reste nécessaire pour comprendre les mécanismes sous-jacents et corriger les imperfections des données.

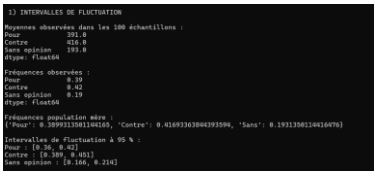
Le choix d'un estimateur soulève plusieurs enjeux, notamment le compromis entre absence de biais, précision, robustesse face aux valeurs extrêmes et efficacité globale. Les théorèmes de Rao-Blackwell et de Lehmann-Scheffé montrent qu'un estimateur sans biais et de variance minimale peut être considéré comme optimal, même si, en pratique, le choix dépend souvent du contexte et des données disponibles.

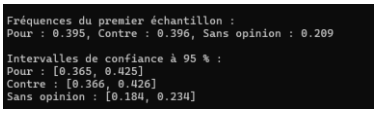
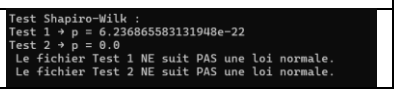
Les méthodes d'estimation se divisent principalement entre l'estimation ponctuelle (qui fournit une valeur unique du paramètre) et l'estimation par intervalle (qui propose une plage de valeurs). Ces estimations peuvent être obtenues par différentes approches, comme la méthode des moments, le maximum de vraisemblance ou l'approche bayésienne, qui intègre une information a priori.

Enfin, les tests statistiques permettent de vérifier la plausibilité d'une hypothèse sur la population en contrôlant le risque d'erreur. Ils reposent sur la formulation d'une hypothèse nulle et d'une hypothèse alternative, le choix d'une statistique de test, la fixation d'un seuil de signification et une règle de décision. On distingue notamment les tests paramétriques, qui supposent certaines conditions sur la distribution, et les tests non paramétriques, utilisés lorsque ces conditions ne sont pas vérifiées.

Bien que la statistique inférentielle fasse l'objet de critiques (notamment en raison de sa dépendance à des hypothèses parfois irréalistes ou de certains usages abusifs) elle reste un outil central pour produire des connaissances généralisables à partir de données partielles.

II- Mise en œuvre avec Python

Numéro de la question	Réponse	Illustration
Question : lien entre l'intervalle de fluctuation et les valeurs réelles de la population mère	L'intervalle de fluctuation ne **décrit pas les valeurs** de la population mère, mais rend compte de la **variabilité attendue** des résultats d'échantillons tirés au hasard. Il permet néanmoins d'évaluer la cohérence d'un échantillon avec une population supposée. Lorsque la valeur observée se situe à l'intérieur de l'intervalle, l'échantillon est considéré comme **compatible avec l'hypothèse formulée**. En revanche, une valeur située en	

	dehors de cet intervalle permet **de remettre en question cette hypothèse.** Ici, les fréquences réelles de la population mère sont d'environ 0,39 pour les "Pour", 0,42 pour les "Contre" et 0,19 pour les "Sans opinion". Les fréquences observées dans les 100 échantillons (0,39 ; 0,42 ; 0,19) se situent **toutes à l'intérieur des intervalles de fluctuation à 95 % correspondants.** Cela signifie que les écarts observés entre les échantillons et la population mère **sont compatibles avec le simple effet du hasard d'échantillonnage.**	
Théorie de l'estimation	L'intervalle de **confiance est plus resserré que l'intervalle de fluctuation**. Les fréquences observées sont de 0,395 pour les "Pour", 0,396 pour les "Contre" et 0,209 pour les "Sans opinion". L'estimation est assez précise des proportions dans la population. Les fréquences estimées restent proches de celles de la population mère et ne présentent pas de décalage significatif.	
Théorie de la décision	Aucune n'est une distribution normale	

Séance 6

I- Question de cours

La statistique ordinale regroupe l'ensemble des méthodes qui reposent sur le classement des individus ou des objets étudiés, en mettant l'accent sur leur ordre relatif plutôt que sur leurs valeurs exactes. Elle s'appuie sur des rangs obtenus après avoir ordonné les observations, ce qui la distingue des statistiques nominales, lesquelles se limitent à des catégories sans relation d'ordre entre elles. Les variables mobilisées sont donc des variables qualitatives ordinales, pour lesquelles un ordre naturel peut être défini, le plus souvent croissant. Ce type d'approche est utile en géographie, car de nombreux phénomènes spatiaux s'organisent spontanément sous forme de hiérarchies, qu'il s'agisse de la taille des villes, de l'intensité d'événements naturels ou encore du dynamisme socio-économique des territoires. Le classement permet alors de rendre visibles des structures hiérarchiques entre espaces.

Dans les classifications, l'ordre croissant est généralement privilégié. Il correspond à l'ordre naturel des observations et facilite l'interprétation des rangs, l'identification des valeurs extrêmes et l'analyse globale des distributions. Cet ordre est aussi celui qui est le plus intuitif dans la majorité des applications statistiques.

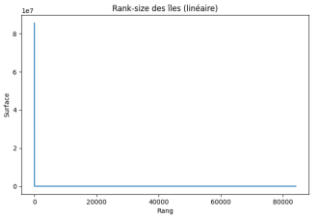
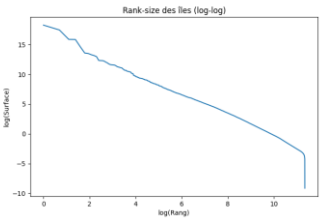
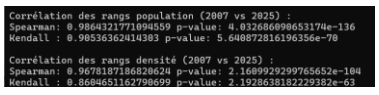
La corrélation des rangs et la concordance de classements répondent à des logiques proches mais distinctes. La corrélation des rangs vise à mesurer la proximité globale entre deux classements en comparant les rangs attribués aux mêmes individus. Elle est généralement évaluée à l'aide des coefficients de Spearman ou de Kendall. La concordance de classements, en revanche, s'intéresse plus finement à la cohérence de l'ordre en examinant, paire par paire, si les individus sont classés dans le même sens ou non. Une concordance parfaite signifie que toutes les paires respectent le même ordre, tandis qu'une concordance nulle correspond à un équilibre entre paires concordantes et discordantes. Ainsi, la corrélation donne une mesure synthétique de proximité, alors que la concordance analyse la cohérence interne des classements.

Les tests de Spearman et de Kendall, bien qu'ils poursuivent le même objectif de comparaison de classements, reposent sur des principes différents. Le test de Spearman calcule une corrélation à partir des différences entre les rangs et s'inscrit dans une logique plus quantitative. Il est sensible à la présence d'ex æquo et, pour des effectifs suffisants, peut être rapproché d'une distribution normale. Le test de Kendall, quant à lui, repose sur le comptage des paires concordantes et discordantes. Conceptuellement plus simple, il compare directement l'ordre de chaque paire d'individus et se généralise plus facilement lorsque plusieurs classements sont analysés simultanément. De manière générale, Spearman mesure la proximité des rangs, tandis que Kendall évalue la cohérence de l'ordre.

Les coefficients de Goodman-Kruskal et de Yule permettent de mesurer l'association entre variables ordinales. Le coefficient de Goodman-Kruskal repose sur la comparaison du nombre

de paires concordantes et discordantes et fournit une mesure comprise entre −1 et +1, allant de l'inversion totale à la concordance parfaite. Il est proche, dans son esprit, du coefficient de Kendall. Le coefficient de Yule, quant à lui, constitue un cas particulier de cette approche, limité aux tableaux de contingence en 2x2. Il est utilisé pour analyser l'association entre variables dichotomiques et s'interprète de manière similaire. Ensemble, ces outils offrent des moyens rigoureux pour quantifier les relations d'ordre et analyser les hiérarchies ou dépendances observées dans les données géographiques.

II- Mise en œuvre avec Python

Numéro de la question	Réponse	Illustration
Question 5 : visualiser la loi rang-taille	Comme énoncé, celle-ci est illisible	
Question 6 : convertir les axes en logarithme	Graphique visible. La courbe **décroit de manière proportionnelle.**	
Question 7 :	Il est impossible car les rangs ne sont pas des variables continues donc **non normalisables.**	
	PARTIE II	
Question 8 à 13 : Préparation du fichier à l'analyse.		
Question 14 : utilisation des méthodes de la corrélation de Spearmanr et la concordance de Kendalltau	Très forte **stabilité des hiérarchies spatiales**, Pour la population, les valeurs très élevées de Spearman (0,99) et de Kendall (0,91), avec des p-values quasi nulles, indiquent une correspondance quasi parfaite entre les rangs aux deux dates. La hiérarchie démographique n'est donc pas changée sur la période. Pour la densité, les c**oefficients restent également très élevés** (Spearman : 0,97 et Kendall : 0,86), ce qui indique une forte	

	concordance des classements (moins marquée que pour la population totale). Les structures hiérarchiques territoriales sont stables dans le temps, et que les évolutions observées restent rares.	
Bonus	Voir le fichier “correlations_population_2007_2025.csv” dans le dossier “images”	

```
def analyse_concordance(popA, popB):
    """Retourne Spearman + Kendall"""
    s = scipy.stats.spearmanr(popA, popB)
    k = scipy.stats.kendalltau(popA, popB)
    return s.correlation, s.pvalue, k.correlation, k.pvalue

# Analyse 2007 à 2025 pour population
annees = list(range(2007, 2026))
correlations = []

for year in annees:
    colA = f"Pop 2007"
    colB = f"Pop {year}"

    popA = list(monde[colA])
    popB = list(monde[colB])

    # classements
    ordA = ordrePopulation(popA, etat)
    ordB = ordrePopulation(popB, etat)
    fusion = classementPays(ordA, ordB)

    rA = [x[1] for x in fusion]
    rB = [x[2] for x in fusion]

    corrS, pS, corrK, pK = analyse_concordance(rA, rB)
    correlations.append([year, corrS, pS, corrK, pK])

pd.DataFrame(correlations, columns=["year", "spearman", "pS", "kendall", "pK"]).to_csv(
    "images/correlations_population_2007_2025.csv", index=False
)
```

Je remarque que j’ai de plus en plus de réflexe quand j’écris le code, il n’est pas parfait mais il fonctionne.

Retour sur le semestre - Réflexion

Je vous avoue que j'ai eu énormément de mal avec la matière.

Tout d'abord d'un point de vue technique : j'ai commencé le semestre sans ordinateur, et les machines prêtées sont, à mes yeux, pas très performantes (pour en avoir déjà emprunter durant toutes ma licence.) J'ai donc commencé par les questions, mais je trouve cela assez compliqué sans pratique.

La pratique quant à elle était moins laborieuse que ce que je pensais, mais ce n'était quand même pas de tout repos. Je me suis entourée de certains de mes amis de vidéos YouTube pour le codage, ainsi que de ChatGPT pour les erreurs (cela m'a vraiment été d'une grande aide, notamment pour forcer l'ouverture d'un fichier etc.). Certaines formules ont fini par entrer dans ma tête un peu machinalement, j'ai créé un fichier "bloc-note" pour le reste, qui m'était très utile.

Les séances 4 et 5 étaient particulièrement difficiles à faire, que ce soit la réalisation en elle-même, mais il s'agit des séances où j'ai fait le plus d'erreurs de code.

Je suis malgré tout contente d'avoir eu cette introduction à Python, étant habituée à R.

Les questions étaient relativement accessibles et faciles à répondre, mais cela prend un temps monstre.

Je retiens que cette matière est une bonne introduction au codage, et que cela pourrait très probablement m'aider pour la suite de mes études et plus.

Les principales difficultés que j'ai rencontrées était des problèmes de matériel.

Fin du rapport
