



## MASTER 1 : GAED

### GéoSuds : Sociétés, Territoires et Développement

# Rapport final d'Analyse de données

```
0     response = requests.get(url) # code from the website
1     # checking response.status_code (if you get 502, try rerunning the code)
2     if response.status_code != 200:
3         print(f"Status: {response.status_code} - Try rerunning the code")
4     else:
5         print(f"Status: {response.status_code}\n")
6
7     # using BeautifulSoup to parse the response object
8     soup = BeautifulSoup(response.content, "html.parser")
9
10    # finding Post images in the soup
11    images = soup.find_all("img", attrs={"alt": "Post image"})
12
13    # downloading images
14    for i in range(0, len(images)):
15        images[i].get_attribute("src").split("/")[-1]
```

## **Retour d'expérience et remarques personnelles :**

Dans le cadre de mon travail, j'ai été confrontée à plusieurs défis techniques qui m'ont permis d'approfondir ma compréhension de la manipulation de données en Python. La principale difficulté a été logistique : elle était liée à la gestion des chemins de fichiers sur mon environnement macOS. Utiliser iCloud a rendu l'accès aux données assez instable, ce qui m'a obligée à automatiser la localisation du répertoire de travail via la bibliothèque os (que j'ai utilisé plusieurs fois durant mon travail et que je continuerai d'utiliser). La deuxième difficulté était liée à l'extraction des colonnes car les caractères spéciaux comme « ² » dans les unités de surface en km : elle provoquait des erreurs de lecture assez redondantes. J'ai dû nettoyer les données. Mis à part ces inconvénients, je n'ai pas d'autres problèmes qui me viennent en tête.



## **Séance 2 :**

### **Question 1 : Quel est le positionnement de la géographie par rapport aux statistiques ?**

La géographie mobilise les statistiques comme un instrument d'analyse spatiale. Les statistiques permettent de quantifier les phénomènes observés, de valider des modèles de distribution et d'apporter une rigueur scientifique.

### **Question 2 : Le hasard existe-t-il en géographie ?**

En géographie, on considère que le hasard n'existe pas vraiment. Même si un événement semble arriver n'importe où (comme l'installation d'un commerce ou un vote particulier), il y a presque toujours une explication.

Au lieu de parler de hasard, on observe des logiques d'organisation : les choses se regroupent souvent parce qu'elles sont proches les unes des autres (effet de cluster) ou parce que le territoire l'impose (présence d'une route, d'une ville, d'une montagne). Ce qui ressemble à du hasard à petite échelle finit par révéler des modèles très organisés dès que l'on regarde la carte de plus haut.

### **Question 3 : Quels sont les types d'information géographique ?**

On distingue les données géométriques, relatives à la localisation et à la forme des objets (points, lignes, polygones), et les données attributaires, qui décrivent les propriétés de ces objets (données qualitatives ou quantitatives).

#### **Question 4 : Quels sont les besoins de la géographie au niveau de l'analyse de données ?**

La discipline nécessite des outils pour synthétiser des volumes de données importants, comparer des unités spatiales, mesurer des évolutions temporelles et détecter des phénomènes d'autocorrélation spatiale. Aussi le fait d'automatiser des tâches répétitives (par exemple, Python intervient dans la réalisation de plusieurs diagrammes pour chaque département) réalisables en cartographie.

#### **Question 5 : Quelles sont les différences entre la statistique descriptive et la statistique explicative ?**

La statistique descriptive résume et présente l'information de manière synthétique (pour la moyenne ou les graphiques par exemple). La statistique explicative cherche à établir des liens de causalité ou de dépendance entre différentes variables pour comprendre les facteurs explicatifs d'un phénomène.

#### **Question 6 : Quelles sont les types de visualisation de données en géographie ? Comment choisir celles-ci ? Quel est le nombre de lignes et de colonnes du jeu de données ?**

Le choix d'une visualisation dépend de la nature de la variable : les diagrammes en barres ou circulaires conviennent aux données discrètes, tandis que l'histogramme est plutôt utilisé pour les distributions continues. La cartographie permet de représenter la dimension spatiale. L'exécution du script main.py sur le fichier source indique que la population étudiée comprend 107 individus (lignes) pour 56 caractères (colonnes).

#### **Question 7 : Quelles sont les méthodes d'analyse de données possibles et quels sont les types de variables présents ?**

Les méthodes varient de l'analyse univariée à l'analyse multivariée (comme l'ACP). L'analyse automatisée via le script main.py montre que le jeu de données combine des variables qualitatives (type object pour les noms de départements) et des variables quantitatives (types int64 ou float64 pour les effectifs de voix).

**Question 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 ?**

La population est l'ensemble étudié (l'électorat français par département), l'individu est l'unité de base (un département), le caractère est la variable mesurée (ex : les inscrits) et la modalité est la valeur spécifique prise par ce caractère. Les caractères sont qualitatifs ou quantitatifs. Une hiérarchie existe puisque les données quantitatives permettent des calculs mathématiques complexes inaccessibles aux données qualitatives.

**Question 9 : Comment mesurer une amplitude et une densité ?**

L'amplitude est l'écart entre la valeur maximale et la valeur minimale d'une distribution. La densité est le rapport entre un effectif (nombre d'habitants, de voix par exemple) et la surface de l'unité spatiale considérée.

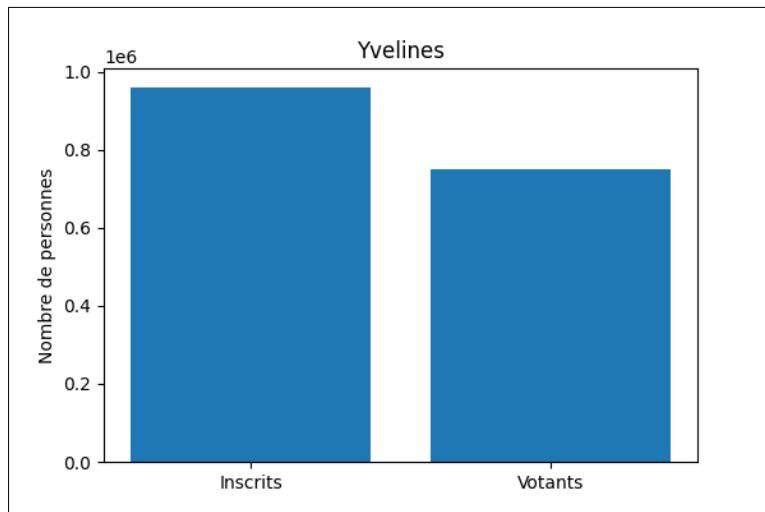
**Question 10 : À quoi servent les formules de Sturges et de Yule et comment calculer la somme des colonnes quantitatives ?**

Ces formules permettent de déterminer statistiquement le nombre « optimal » de classes pour construire un histogramme sans grande perte d'information majeures. D'un point de vue pratique, le script main.py utilise une structure conditionnelle pour identifier les colonnes numériques et en calculer la somme. Cela permet d'obtenir des totaux nationaux cohérents.

## **Question 11 : Comment définir un effectif, une fréquence et une distribution statistique ? Réalisation des diagrammes en barres.**

L'effectif est le nombre d'individus pour une modalité. La fréquence est la proportion de cet effectif par rapport au total. Une distribution statistique décrit la répartition des individus selon les valeurs du caractère. Le script génère pour chaque département un diagramme en barres comparant le nombre d'inscrits et de votants, que vous retrouverez dans le dossier images\_barres/.

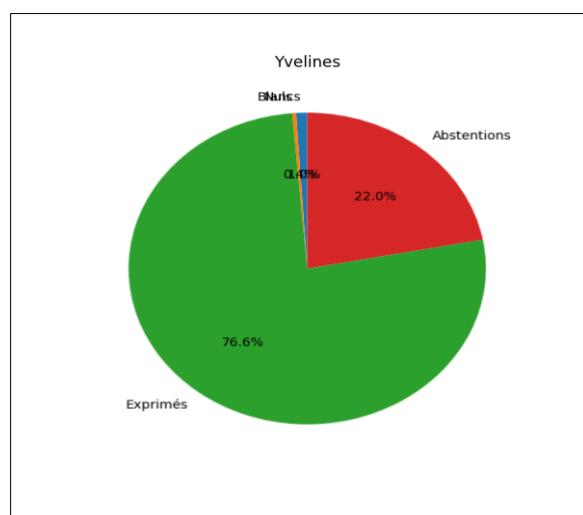
**Exemple :** Diagramme en barre du nombre d'inscrits et de votants dans le département des Yvelines (78) :



## **BONUS : Réalisation des diagrammes circulaires (Blancs, Nuls, Exprimés, Abstentions).**

À l'aide d'une boucle dans le fichier main.py, des diagrammes circulaires sont produits pour chaque ligne du jeu de données. La méthode permet d'analyser visuellement la structure du vote et le poids de l'abstention par territoire. J'ai placé les fichiers dans le dossier images\_pie/.

**Exemple :** Diagramme circulaire de la quantité de votes blancs, nuls, exprimés ainsi que les abstentions dans le département des Yvelines (78) :



## **BONUS : Réalisation de l'histogramme de la distribution des inscrits.**

Contrairement au diagramme en barres, l'histogramme représente ici la distribution statistique de la variable "Inscrits" sur l'ensemble de la population. Le script génère ce graphique pour visualiser la concentration des volumes électoraux à travers les départements français.

## **BONUS : Analyse des voix par candidat et résultat national.**

Le script a été complété par une routine générant la répartition des voix par candidat pour chaque département. Enfin, une agrégation des données a permis de produire le diagramme circulaire des résultats du premier tour à l'échelle nationale.

## **Séance 3 :**

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

Le caractère qualitatif est considéré comme le plus général car il peut s'appliquer à n'importe quel objet par simple catégorisation (par exemple le nom, la couleur ou encore le type). À l'inverse, le caractère quantitatif est plus « restrictif » car il nécessite une mesure numérique précise. Cependant, en termes de richesse d'information, le caractère quantitatif est supérieur car il permet des opérations mathématiques (ce que nous avons réalisé avec la moyenne ou encore l'écart-type) impossibles sur des données qualitatives.

### **Question 2 : Que sont les caractères quantitatifs discrets et caractères quantitatifs continus ? Pourquoi les distinguer ?**

Un caractère discret ne peut prendre que des valeurs entières isolées (exemple : un nombre de voix), alors qu'un caractère continu peut prendre n'importe quelle valeur dans un intervalle donné (exemple : la surface d'une île en km<sup>2</sup>). Les outils de traitement diffèrent : on utilise souvent des diagrammes en bâtons pour le discret et des histogrammes ou des densités pour le continu (généralement).

### **Question 3 : Pourquoi existe-t-il plusieurs types de moyenne ? Pourquoi calculer une médiane ? Quand est-il possible de calculer un mode ?**

Il existe plusieurs moyennes (arithmétique, géométrique etc..) pour s'adapter à la nature des données (linéaire, taux de croissance, etc..). La médiane est calculée car, contrairement à la moyenne, elle n'est pas influencée par les valeurs extrêmes ; elle sépare la population en deux groupes égaux. Le mode, qui est la valeur la plus fréquente, peut être calculé pour tous les types de caractères, y compris qualitatifs.

En pratique, le script main (3).py calcule ces trois indicateurs pour chaque colonne quantitative du fichier électoral. Les résultats sont arrondis à deux décimales avec la méthode .round(2).

#### **Question 4 : Quel est l'intérêt de la médiale et de l'indice de C. Gini ?**

Ces paramètres mesurent la concentration. La médiale est la valeur qui divise la masse totale du caractère en deux parts égales. L'indice de Gini, compris entre 0 et 1, mesure le degré d'inégalité de la répartition (par exemple, si quelques départements concentrent la majorité des inscrits).

#### **Question 5 : Pourquoi calculer une variance à la place de l'écart à la moyenne ? Pourquoi la remplacer par l'écart type ? Pourquoi calculer l'étendue ? À quoi sert-il de créer un quantile ? Pourquoi construire une boîte de dispersion ?**

La variance est utilisée car la somme simple des écarts à la moyenne s'annule toujours ; le passage au carré permet de ne manipuler que des valeurs positives. On utilise l'écart-type pour revenir à l'unité de mesure d'origine des données. L'étendue (différence entre maximum et minimum) donne une première idée de la dispersion totale. Les quantiles (quartiles, déciles) permettent de diviser la série pour observer la répartition par tranches (exemple : les 10 % les plus riches).

Dans le script, la distance interquartile et interdécile est obtenue via la méthode .quantile(). Enfin, la boîte de dispersion (ou boîte à moustaches) est construite pour visualiser simultanément la médiane, les quartiles et les valeurs aberrantes.

Référence : script main (3).py et dossier img/ pour les graphiques.

#### **Question 6 : Quelle différence faites-vous entre les moments centrés et les moments absous ? Pourquoi vérifier la symétrie d'une distribution et comment faire ?**

Les moments centrés sont calculés par rapport à la moyenne, ce qui permet de neutraliser l'effet de position. Les moments absous utilisent les valeurs brutes. Le fait de vérifier la symétrie permet de savoir si les individus se répartissent de manière équilibrée autour de la moyenne. Par ailleurs, si la moyenne est égale à la médiane, la distribution est symétrique.

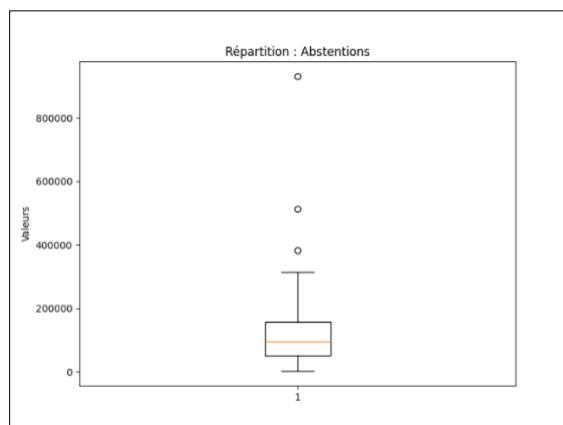
### **BONUS : Calcul de la distance interquartile et interdécile dans Pandas.**

Pour chaque colonne quantitative du jeu de données électORALES, le script utilise df.quantile([0.25, 0.75]) pour l'écart interquartile et df.quantile([0.1, 0.9]) pour l'interdécile. Ces mesures permettent d'évaluer la dispersion centrale et les queues de distribution des voix par département. J'ai effectué cela dans le dossier intitulé question Bonus/.

### **BONUS : Réalisation des boîtes à moustaches.**

Le script génère, via une boucle, une boîte à moustache pour chaque variable numérique. Ces graphiques permettent d'identifier visuellement les départements dont le comportement électoral s'écarte significativement de la norme (valeurs atypiques).

**Exemple :** Boîte à moustache de dispersion du taux d'abstention par département :



### **BONUS : Algorithme de catégorisation et dénombrement des surfaces.**

Pour le fichier island-index.csv, un algorithme de discréétisation a été mis en place en utilisant la fonction pd.cut(). Les surfaces ont été découpées selon des bornes spécifiques (0, 10, 25, 50, 100, 2500, 5000, 10000, + ∞).

L'application de value\_counts().sort\_index() permet de dénombrer les îles par classe. On observe une immense majorité d'îles dans la tranche ]0, 10], soit 78 423 individus, ce qui souligne une distribution très asymétrique vers les petites surfaces.

Référence : Fichier intitulé resultat\_categories\_iles.csv.

## Séance 4 :

### **Question 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 ?**

Le choix repose principalement sur la nature du phénomène observé et la précision de la mesure. On privilégie une distribution discrète lorsque les données correspondent à des comptes entiers et finis (ex : nombre d'habitants, nombre de villes). À l'inverse, on utilise une distribution continue pour des phénomènes mesurés sur une échelle infinie ou fractionnable, comme des distances, des surfaces ou des densités de population. Un autre critère est la taille de l'échantillon : pour une population très large, une variable discrète est souvent approximée par une loi continue (comme la loi normale) pour simplifier les calculs.

### **Question 2 : Expliquez selon vous quelles sont les lois les plus utilisées en géographie ?**

En géographie, trois lois se distinguent par leur pertinence spatiale :

**La loi Normale** : utilisée pour modéliser des phénomènes qui se répartissent de manière équilibrée autour d'une moyenne (ex : tailles, températures).

**La loi de Pareto** : fondamentale pour l'étude de la hiérarchie urbaine (loi rang-taille) et des inégalités de richesse, où un petit nombre d'individus concentre la majorité de la valeur.

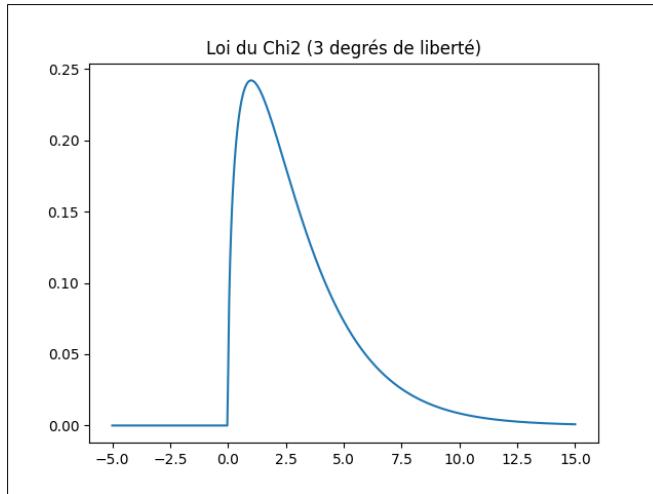
**La loi de Poisson** : essentielle pour analyser la distribution spatiale de points et déterminer si des événements (ex : implantations d'entreprises) se produisent de manière aléatoire ou agrégée.

**La loi de Zipf** : très proche de Pareto, elle est spécifiquement utilisée pour modéliser la distribution de la taille des villes au sein d'un système urbain.

## PRATIQUE : Visualisation des distributions de variables discrètes.

À l'aide du script main (4).py, plusieurs lois discrètes ont été modélisées et sauvegardées dans le dossier img/.

Exemple : Distribution théorique de la loi Chi<sup>2</sup> (3 degrés de liberté) :



- **Loi de Dirac** : modélise un événement certain en un point unique.
- **Loi Uniforme discrète** : chaque modalité a la même probabilité d'apparaître (ex : lancer de dé).
- **Loi Binomiale** : modélise le nombre de succès dans une répétition d'expériences.
- **Loi de Poisson** : appliquée aux événements rares dans un intervalle de temps ou d'espace.
- **Loi de Zipf-Mandelbrot** : mise en évidence pour illustrer la hiérarchie des fréquences (utilisée pour les populations des villes).

Référence : script main (4).py et fonctions plt.stem() pour le rendu des probabilités ponctuelles.

## PRATIQUE : Visualisation des distributions de variables continues.

Le script permet également de visualiser les lois continues en utilisant des fonctions de densité de probabilité.

- **Loi Normale** : représentée par la célèbre courbe en cloche (centrée sur 0 avec un écart-type de 1).
- **Loi Log-normale** : modélise des variables dont le logarithme suit une loi normale (fréquent en géographie économique).
- **Loi du Chi<sup>2</sup>** : utilisée pour les tests d'indépendance entre deux variables géographiques.
- **Loi de Pareto** : visualisée pour montrer la décroissance rapide de la courbe, caractéristique des systèmes très hiérarchisés.

### **PRATIQUE : Calcul de la moyenne et de l'écart-type théoriques.**

Une fonction informatique nommée `calculer_stats` a été créée pour automatiser l'extraction des paramètres de chaque loi via la méthode `.stats(moments='mv')` de Scipy.

L'exécution du script affiche dans le terminal les valeurs théoriques. Par exemple, pour la loi normale standard, le script confirme une moyenne de 0.00 et un écart-type de 1.00. Pour la loi de Pareto (avec un paramètre  $b = 2.6$ ), la moyenne théorique calculée est de 1.62.

---

## **Séance 5 :**

### **Question 1 : Comment définir l'échantillonnage ? Pourquoi ne pas utiliser la population en entier ? Quelles sont les méthodes d'échantillonnage et comment les choisir ?**

L'échantillonnage est le processus de sélection d'un sous-ensemble d'individus à partir d'une population parente. On ne travaille pas sur la population entière pour des raisons de coût ou de temps. Les méthodes incluent l'échantillonnage aléatoire simple (chaque individu a la même chance d'être choisi), stratifié (groupes homogènes) ou en grappes. Le choix dépend de la structure des données et de la précision souhaitée.

### **Question 2 : Comment définir un estimateur et une estimation ?**

Un estimateur est une règle mathématique ou une fonction appliquée aux données de l'échantillon pour évaluer un paramètre inconnu de la population. L'estimation est la valeur numérique précise obtenue en appliquant cet estimateur à un échantillon donné.

### **Question 3 : Comment distingueriez-vous l'intervalle de fluctuation et l'intervalle de confiance ?**

L'intervalle de fluctuation est déterminé à partir de la population mère pour prévoir où se situera la statistique d'un échantillon. L'intervalle de confiance est calculé à partir d'un échantillon pour estimer où se situe le paramètre réel de la population mère, que l'on ne connaît pas.

#### **Question 4 : Qu'est-ce qu'un biais dans la théorie de l'estimation ?**

Un biais est une erreur systématique. Un estimateur est biaisé s'il surestime ou sous-estime systématiquement la valeur réelle de la population. Un bon estimateur doit être "sans biais", c'est-à-dire que sa moyenne sur un grand nombre d'échantillonnages est égale au paramètre réel.

#### **Question 5 : Comment appelle-t-on une statistique travaillant sur la population totale ? Faites le lien avec la notion de données massives (Big Data).**

On parle de statistique exhaustive ou de recensement. Avec le Big Data, on traite souvent la quasi-totalité des données (population complète), ce qui réduit le besoin d'inférence statistique traditionnelle, même si l'analyse de la qualité des données reste nécessaire.

#### **Question 6 : Quels sont les enjeux autour du choix d'un estimateur ?**

Les enjeux majeurs sont la convergence (précision accrue avec la taille de l'échantillon), l'absence de biais et l'efficacité (variance minimale pour garantir une estimation stable).

#### **Question 7 : Quelles sont les méthodes d'estimation d'un paramètre ? Comment en sélectionner une ?**

On distingue l'estimation ponctuelle (une valeur unique) et l'estimation par intervalle (une plage de valeurs). On sélectionne la méthode selon le niveau de confiance requis (souvent 95 %).

#### **Question 8 : Quels sont les tests statistiques existants ? À quoi servent-ils ? Comment créer un test ?**

Il existe des tests de conformité, d'ajustement (comme le Chi<sup>2</sup>) ou de comparaison (test de Student). Ils servent à rejeter ou non une hypothèse nulle (H<sub>0</sub>). On crée un test en définissant H<sub>0</sub> et H<sub>1</sub>, en choisissant un seuil de risque (Alpha), et en calculant une p-value.

### **Question 9 : Que pensez-vous des critiques de la statistique inférentielle ?**

Les critiques pointent souvent la mauvaise utilisation de la p-value ou le manque de représentativité des échantillons. Même si elle est puissante, il faut une rigueur méthodologique pour ne pas produire de conclusions qui comprennent des erreurs.



### **Analyse de l'échantillonnage et intervalles de fluctuation**

À partir de la population mère (\$N=2185\$), le script main (5).py a permis de calculer les intervalles de fluctuation au seuil de 95 %. Ces intervalles indiquent les zones où les fréquences des 100 échantillons devraient théoriquement se situer :

**Pour** : \$[0,37 ; 0,41]\$

**Contre** : \$[0,40 ; 0,44]\$

**Sans opinion** : \$[0,18 ; 0,21]\$

## **Estimation par intervalle de confiance**

En isolant le premier échantillon (taille  $n=1000$ ), nous avons calculé les intervalles de confiance :

**Pour** :  $[0,36 ; 0,43]$

**Contre** :  $[0,37 ; 0,42]$

On constate que l'intervalle de confiance est plus large que l'intervalle de fluctuation, ce qui traduit l'incertitude liée à l'utilisation d'un échantillon unique pour estimer la population globale.

## **Théorie de la décision et test de normalité**

Le test de Shapiro-Wilk a été appliqué aux deux séries de données pour vérifier leur normalité :

**Fichier Test 1** : p-value  $\approx 0,24$ . Comme  $p > 0,05$ , on ne rejette pas l'hypothèse nulle. La distribution est normale.

**Fichier Test 2** : p-value  $\approx 0,00$ . Comme  $p < 0,05$ , on rejette  $H_0$ . La distribution est non normale.

L'analyse visuelle du Test 2 montre une forte concentration sur les petites valeurs (1 et 2) avec une traîne vers la droite (valeurs jusqu'à 14), suggérant une Loi de Poisson plutôt qu'une distribution gaussienne.

---

## Séance 6 :

### **Question 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 ?**

Une statistique ordinale porte sur des données que l'on peut classer selon un ordre logique, mais sans que l'écart numérique entre les rangs soit forcément régulier. Elle s'oppose à la statistique nominale (où les catégories n'ont pas d'ordre, comme les noms de pays). Elle utilise des variables qualitatives ordinaires ou des variables quantitatives transformées en rangs. En géographie, cela matérialise une hiérarchie spatiale en classant les territoires selon leur importance (exemple : classement des villes par population, classement des pays par IDH), permettant d'identifier des centres et des périphéries.

### **Question 2 : Quel ordre est à privilégier dans les classifications ?**

L'ordre à privilégier dépend du phénomène, mais en géographie et en économie, on utilise majoritairement l'ordre décroissant (du plus grand au plus petit). Cela permet de mettre en évidence les entités dominantes (le "haut du classement") qui structurent souvent l'espace étudié.

### **Question 3 : Quelle est la différence entre une corrélation des rangs et une concordance de classements ?**

La corrélation des rangs mesure la force et la direction de la relation entre deux classements (si un pays est premier dans l'un, l'est-il aussi dans l'autre ?). La concordance évalue plus largement la similarité globale ou l'accord entre plusieurs juges ou plusieurs séries de classements pour déterminer s'il existe un consensus dans l'organisation de la hiérarchie.

## **Question 4 : Quelle est la différence entre les tests de Spearman et de Kendall ?**

Le coefficient de Spearman ( $\rho$ ) est basé sur la différence des rangs au carré ; il est très sensible aux écarts de position. Le coefficient de Kendall ( $\tau$ ) est basé sur le nombre de paires concordantes et discordantes ; il est souvent jugé plus robuste pour les petits échantillons et plus facile à interpréter en termes de probabilité (risque de désaccord entre deux paires)

## **Question 5 : À quoi servent les coefficients de Goodman-Kruskal et de Yule ?**

Ces coefficients servent à mesurer l'association entre des variables qualitatives ordinaires, notamment dans des tableaux de contingence. Le gamma de Goodman-Kruskal est particulièrement utile lorsqu'il y a beaucoup d'exæquos ("ties") dans les classements, tandis que le Q de Yule est un cas particulier pour les tableaux 2x2.

## **PRATIQUE : Analyse des classements mondiaux**

### Traitements des données de population et densité

À l'aide du script main (6).py, nous avons automatisé le classement des États du monde. La fonction ordrePopulation a permis de transformer les données brutes de population et de densité en rangs (de 1 à N).

Résultats de la comparaison (Année 2007)

En comparant le rang des pays selon leur population totale et leur rang selon leur densité en 2007, le script a calculé les coefficients suivants :

**Coefficient de Spearman** : 0,4839

**Coefficient de Kendall** : 0,3373

*Partie interprétative* : Il existe une corrélation positive modérée. Cela signifie que les pays les plus peuplés ont une tendance certaine à être également les plus denses,

mais cette relation est loin d'être systématique (exemple : de très grands pays comme la Russie ou le Canada sont très peuplés mais très peu denses).

### **BONUS : Automatisation et Factorisation**

Le code a été optimisé par la création d'une fonction `calculerCoefficientsRangs` qui encapsule tout le processus (tri, alignement des pays, calcul de Spearman et Kendall). Cette approche permet d'analyser rapidement n'importe quel couple de variables.

Pour les îles, l'application de cet algorithme montre une corrélation très forte entre la surface et le trait de côte, ce qui est géométriquement logique, bien que l'indice de sinuosité puisse faire varier certains rangs.

---

### **Avis personnel sur le cours :**

Honnêtement, mon avis est assez positif parce que j'ai pu faire le pont entre la géographie et les outils numériques actuels. Ce format de cours nous a permis de tester nous-mêmes des vrais jeux de données (résultat des élections présidentielles), lorsque les lignes de codes fonctionnent, il y a un vrai aspect satisfaisant. Cependant, cela représente une marche assez haute au début. Par exemple, une parenthèse oubliée ou un fichier mal renseigné (plusieurs fois dans mon cas) peut bloquer tout le travail. Je tiens également à remercier Zara Huston et Stelline Théverin pour leur aide dans la réalisation de mon TP ainsi que monsieur Maxime Forriez pour le temps qu'il a accordé à mon cas.