

Cours d'analyse de données en géographie

Niveau Master 1 - GEANDO

Séance 6. La statistique d'ordre des variables qualitatives

Maxime Forriez^{1,a}

¹ Institut de géographie, 191, rue Saint-Jacques, Bureau 105, 75 005 Paris,
^amaxime.forriez@sorbonne-universite.fr

21 septembre 2025

1 Questions de cours

Les réponses comptent pour 20 % de la note finale du parcours « débutants ».

Les réponses comptent pour 10 % de la note finale du parcours « intermédiaires ».

Les réponses comptent pour 10 % de la note finale du parcours « confirmés ».

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 ?
2. Quel ordre est à privilégier dans les classifications ?
3. Quelle est la différence entre une corrélation des rangs et une concordance de classements ?
4. Quelle est la différence entre les tests de Spearman et de Kendal ?
5. À quoi servent les coefficients de Goodman-Kruskal et de Yule ?

2 Mise en œuvre avec Python

La sous-partie « Bonus » vous permet d'obtenir des points supplémentaires.

2.1 Objectifs

- Manipuler des fonctions locales et comprendre la nécessité de factoriser son code en une liste de fonctions ou de procédures exécutant une tâche unique
- Créer des fonctions locales spécifiques au traitement d'un problème
- Comprendre l'analyse de variables qualitatives ordinales

2.2 Manipulations

Le fichier obtenu compte pour 20 % de la note finale du parcours « débutants ».

Le fichier obtenu compte pour 15 % de la note finale du parcours « intermédiaires ».

Le fichier obtenu compte pour 10 % de la note finale du parcours « confirmés ».

1. Dans le dossier `src`, introduire le dossier `data` le fichier `island-index.csv` disponible dans la Seance-06 du GitHub
2. Ouvrir le fichier en utilisant la fonction locale `ouvrirUnFichier()`. Elle prend en paramètre une chaîne de caractères matérialisant l'adresse et le nom du fichier.
3. Isoler la colonne « Surface (km²) » et lui ajouter dans cette liste :
 - Asie / Afrique / Europe : 85 545 323 km²
 - Amérique : 37 856 841 km²
 - Antarctique : 7 768 030 km²
 - Australie : 7 605 049 km²

Attention ! Il faudra forcer le typage en castant les valeurs en `float()`.

Attention ! Ne mettez pas l'unité de mesure dans la liste informatique

4. Ordonner la liste obtenue avec la fonction locale `ordreDecroissant()` proposée. Elle prend en paramètre une liste.
5. Visualiser la loi rang-taille en créant une image de sortie.
6. L'image obtenue est illisible. Il vous faut convertir les axes en logarithme. Pour ce, utiliser la fonction locale `conversionLog()` proposée. Elle prend en paramètre une liste.
7. Est-il possible de faire un test sur les rangs ? (mettre votre réponse sous la forme d'un commentaire dans le fichier)
8. Dans le dossier `src`, introduire le dossier `data` le fichier `Le-Monde-HS-Etats-du-monde-2007-2025.csv` disponible dans la Seance-06 du GitHub
9. Ouvrir le fichier en utilisant la fonction locale `ouvrirUnFichier()`. Elle prend en paramètre une chaîne de caractères matérialisant l'adresse et le nom du fichier.
10. Isoler les colonnes « État », « Pop 2007 », « Pop 2025 », « Densité 2007 » et « Densité 2025 »

11. En utilisant la fonction locale `ordrePopulation()`, ordonner de manière décroissante les listes « Pop 2007 », « Pop 2025 », « Densité 2007 » et « Densité 2025 ». Elle prend en paramètres la liste à ordonner et la liste des États.
12. En utilisant la fonction locale `classementPays()`, préparer la comparaison des listes sur la population et la densité. Elle prend en paramètres les deux classements (pour le nombre d'habitants et pour la densité) obtenus avec `ordrePopulation()`. En classant le résultat avec la méthode `sort()`, vous obtenez une liste avec deux colonnes ordonnées par rapport au classement de 2007.
13. Isoler les deux colonnes sous la forme de liste différents en utilisant une boucle
14. Dans la bibliothèque `scipy.stat`, utiliser les méthodes `spearmanr()` et `kendalltau()` pour calculer le coefficient de corrélation des rangs et la concordance des rangs. Les deux méthodes prennent en paramètres les deux classements que vous avez respectivement calculés pour le nombre d'habitants et pour la densité. Vous commenterez ce résultat dans votre rapport d'activité.

2.3 Bonus

Afin de vous faciliter la vie, car les codes n'étaient pas simples à trouver, vous avez utilisé des fonctions que j'ai créées. C'est l'un des bonus qui peut vous rapporter le plus de points. Vous devrez créer des fonctions locales permettant de calculer rapidement une analyse de rangs.

Pour les îles, proposer un algorithme et un code pour comparer le classement obtenu avec les surfaces et un autre classement obtenu avec les traits de côte en utilisant le coefficient de corrélation et la concordance des rangs.

Pour la population mondiale, 1. factoriser le code d'analyse des classements sous la forme d'une fonction renvoyant les coefficients de corrélation des rangs et la concordance des rangs ; 2. faire un algorithme et un code permettant d'analyser la concordance des rangs de l'ensemble des classements annuels de 2007 à 2025.