Cours d'analyse de données en géographie
Niveau Master 1 - GEANDO
Séance 3. Les paramètres statistiques élémentaires

Maxime Forriez[1,a]

[1] Institut de géographie, 191, rue Saint-Jacques, Bureau 105, 75 005 Paris,
[a]maxime.forriez@sorbonne-universite.fr

20 septembre 2025

1 Questions de cours

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

1. Quel caractère est le plus général : le caractère quantitatif ou le caractère qualitatif ? Justifier pourquoi.
2. Que sont les caractères quantitatifs discrets et caractères quantitatifs continus ? Pourquoi les distinguer ?
3. **Paramètres de position**
   - Pourquoi existe-t-il plusieurs types de moyenne ?
   - Pourquoi calculer une médiane ?
   - Quand est-il possible de calculer un mode ?
4. **Paramètres de concentration**
   - Quel est l'intérêt de la médiale et de l'indice de C. Gini ?
5. **Paramètres de dispersion**
   - 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 ? Quel(s) est (sont) le(s) quantile(s) le(s) plus utilisé(s) ?
   - Pourquoi construire une boîte de dispersion ? Comment l'interpréter ?
6. **Paramètres de forme**
   - Quelle différence faites-vous entre les moments centrés et les moments absolus ? Pourquoi les utiliser ?
   - Pourquoi vérifier la symétrie d'une distribution et comment faire ?

2 Mise en œuvre avec Python

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

2.1 Objectifs

— Découvrir les méthodes de `Pandas` permettant de calculer les paramètres d'une série statistique

— Tracer une boîte de dispersion

2.2 Manipulations

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

1. Dans le dossier `src`, créer un dossier `data` et y introduire le fichier `resultats-elections-presidentielles-2022-1er-tour.csv` disponible dans la `Seance-03` du `GitHub`
2. Dans le dossier `src`, introduire le fichier `main.py` de la séance disponible dans la `Seance-03` du `GitHub`
3. Ouvrir le fichier `main.py` dans votre éditeur de code (`Notepad++` ou `VS Code`)
4. Utiliser l'instruction `with` pour ouvrir le fichier C.S.V. avec la méthode `read_csv(...)` de la bibliothèque `Pandas`
5. En reprenant le code précédemment créé dans la séance 2, sélectionner les colonnes contenant des caractères quantitatifs ? Calculer sous la forme d'une liste :
   — les moyennes de chaque colonne avec la bonne méthode de `Pandas` ;
   — les médianes de chaque colonne avec la bonne méthode de `Pandas` ;
   — les modes de chaque colonne avec la bonne méthode de `Pandas` ;
   — l'écart type de chaque colonne avec la bonne méthode de `Pandas` ;
   — l'écart absolu à la moyenne de chaque colonne avec la bonne méthode de `Pandas` ;
   — l'étendue de chaque colonne.

   N.B. 1. Utiliser la méthode de la valeur absolue `abs()` disponible dans `Numpy`

   N.B. 2. Utiliser les méthodes `min()` et `max()` disponibles dans `Pandas`

   En utilisant la méthode `round()` de `Pandas`, arrondir tous les paramètres à deux décimaux.
6. Afficher la liste des paramètres sur le terminal
7. Calculer la distance interquartile et interdécile de chaque colonne quantitative avec la méthode `quantile()` dans `Pandas` ?
8. À l'aide de `Matplolib` et d'une boucle, faire des boîtes à moustache de chaque colonne quantitative. Stocker les résultats dans un dossier `img`.

9. Dans le dossier `src`, introduire le dossier `data` le fichier `island-index.csv` disponible dans la `Seance-03` du `GitHub`
10. Sélectionner la colonne « Surface (km²) » et écrire un algorithme pour catégoriser et dénombrer le nombre d'îles ayant une surface comprise :
    - entre 0 et 10 km² ou $]0, 10]$;
    - entre 10 et 25 km² ou $]10, 25]$;
    - entre 25 et 50 km² ou $]25, 50]$;
    - entre 50 et 100 km² ou $]50, 100]$;
    - entre 100 et 2 500 km² ou $]100, 2500]$;
    - entre 2 500 et 5 000 km² ou $]2500, 5000]$;
    - entre 5 000 et 10 000 km² ou $]5000, 10000]$;
    - supérieur ou égal 10 000 km² ou $]10000, +\infty[$.

    Vous concevrez un organigramme pour expliquer votre solution. L'objectif de cette dernière question est d'apprendre à catégoriser des variables quantitatives.

2.3 Bonus

Sans remarque pour vous aider (conditions réelles), sortir les listes calculées avec la bonne méthode `Pandas` au format C.S.V. et `Excel`. N'oubliez pas de titrer les colonnes et les lignes de vos sorties si nécessaire.