

Projet de traitement de données en SHS - TP 1

L2 MIASHS - S3

1

1. Calculs et variables

On s'intéresse ici à la gestion d'une classe de 7 élèves : 3 garçons et 4 filles. Ce premier exercice a pour but de vous familiariser avec le langage python. N'oubliez pas d'écrire votre code dans un fichier .py et d'y mettre des commentaires. Vous pouvez utiliser ou l'éditeur de texte de votre choix.

- 1 Vérifier (avec python) que 3 garçons et 4 filles font bien 7 élèves en tout.
- 2 Créer une liste *noms* contenant les noms des enfants : Simone, Jacqueline, Jeanne, Philippe, Bart, Thomas et Tiffany.
- 3 Créer une seconde liste *notes* contenant leurs notes au dernier devoir, respectivement : 15, 10, 3, 17, 18, 7 et 13.
- 4 Transformer ces deux listes en tableaux. Les concaténer en une matrice.
- 5 À partir des deux listes, créer un DataFrame.
- 6 **Sans** utiliser les fonctions python associées, calculer la moyenne de la classe à ce devoir..
- 7 Un deuxième devoir a été fait, les notes sont respectivement de 11, 16, 19, 8, 4, 18 et 14. Ajoutez ce nouveau devoir au data-frame dans une colonne *notes2*. Donnez ensuite moyenne et variance pour ce deuxième devoir en utilisant les fonctions python.
- 8 Calculez le coefficient de corrélation entre les deux devoirs. Notez vous quelque chose d'étrange ?

Checkpoint ! Appelez l'encadrant pour qu'il vienne vérifier votre travail.

2. Résidences étudiantes

On s'intéresse à un ensemble de résidences étudiantes dans une ville américaine. Le fichier *immeublesUSA.data* contient le revenus globaux net des résidences ainsi que leurs nombre d'appartements. Le fichier est disponible à l'adresse <http://lietard.fr/teaching/immeublesUSA.data>.

- 1 Chargez le fichier de données dans un data-frame.
- 2 Calculez le revenu moyen par appartement pour chaque résidence.
- 3 Représentez la variable "revenu moyen par appartement" avec un diagramme à moustache. Quel problème remarquez vous ? Comment suggérez vous de le résoudre ?

Checkpoint ! Appelez l'encadrant pour qu'il vienne vérifier votre travail.

3. Nos députés

Nous allons étudier le fichier *deputes.data* qui contient les informations sur les députés de la législature actuelle disponibles sur nosdeputes.fr; le fichier se trouve à l'adresse <http://lietard.fr/teaching/deputes.data>.

- 1 Chargez le fichier de données dans un data-frame.
- 2 Combien y a t-il d'hommes et de femmes ?
- 3 Combien y a t-il de députés sans profession (codé '0') ?

Checkpoint ! Appelez l'encadrant pour qu'il vienne vérifier votre travail.

4. Les employés municipaux

Une mairie vous demande d'effectuer quelques statistiques sur l'ensemble de leurs employés municipaux. Vous trouverez le fichier de données correspondant à l'adresse <http://lietard.fr/teaching/employees.data>.

Pour chaque employé, le fichier de données contient les informations suivantes : prénom, nom, sexe, taille (en cm), poids (en kg), et année de naissance.

- 1 Charger le fichier de données dans un data-frame.
- 2 La parité est-elle respectée dans cette mairie ?
- 3 Donnez une description statistique complète des variables *taille* et *annee*.
- 4 Y a t-il une corrélation entre la taille et le poids ?
- 5 Une fonctionnalité intéressante de python est que les vecteur peuvent être considérés comme des termes dans un calcul arithmétique ; Autrement dit, pour ajouter 2 à tous les éléments d'un vecteur x , il suffit d'écrire $x = x + 2$. En vous servant de cela, calculez l'IMC de tous les employés (IMC = poids en kg / (taille en m)²).
- 6 Un IMC normal est compris entre 18.5 et 25. Créez une fonction qui prend en paramètre le data-frame et qui affiche le nom complet (prénom+nom) des personnes qui ont un IMC en dehors de ces bornes.
- 7 Donnez la taille moyenne des femmes et celle des hommes. Tracer un diagramme à moustache représentant la distribution de la variable *taille* pour les 2 catégories.

Bonus Est-ce que la taille des hommes a tendance à augmenter ? Et celle des femmes ? Et globalement ?

Checkpoint ! Appelez l'encadrant pour qu'il vienne vérifier votre travail.