

Analyse des données

TP 1

Remarque

L'utilité d'utiliser quelques *bibliothèques* et *modules indispensables* :

- Module **math** : Donne accès aux fonctions sur les réels.
- Module **statistics** : Permet de calculer les valeurs statistiques basiques
- Module **geometry** : variables, fonctions qui calculent l'aire, classes (carré, triangle)
- Bibliothèques **numpy**, **Scipy**, **Pandas** : Donnent des packages avec beaucoup de modules pour le calcul numérique scientifique.
- Bibliothèque **matplotlib** : Pour tracer des courbes et visualiser les données.
- Pour pouvoir utiliser ces modules et bibliothèques, nous utilisons le suffixe **import** afin de les importer et les exploiter.

Exemple

```
>>> import math  
>>> math.cos (math.pi/3)
```

0.5000000000000001

Vous pouvez toujours réduire le résultat obtenu !!!!!!!!!!!!!

- Les principales structures de données liées à notre domaine :

Les vecteurs, les listes, les matrices de données, les fonctions statistiques, graphisme,

La bibliothèque numpy est spécifique pour les vecteurs et les matrices.

Objectif

Le but de ce TP est de découvrir et d'appliquer quelques fonctions de base de python utiles pour l'analyse factorielle en s'appuyant sur les bibliothèques de base citées ci-dessus : Savoir explorer et manipuler les données.

Dans ce premier exercice proposé, vous allez essayer de voir de près et d'appliquer les différentes notions vues en cours sur les matrices des données. Aussi, savoir lire et interpréter les résultats obtenus.

Vous pouvez faire appel à quelques fonctions prédéfinies en exploitant les bibliothèques de base pour l'analyse factorielle ou bien faire des changements sur les fonctions prédéfinies pour les adapter à nos concepts.

Un autre point qui est très important en analyse des données est de visualiser les données et savoir interpréter les graphes obtenus.

Analyse des données

Exercice 1

Considérons la matrice des données X suivante, résultat de 4 observations sur 10 individus.

$I \backslash J$	<i>Moyenne</i>	<i>Age</i>	<i>Taille</i>	<i>Poids</i>
<i>Enfant 1</i>	14	13	1.50	45
<i>Enfant 2</i>	16	13	1.60	50
<i>Enfant 3</i>	15	13	1.65	50
<i>Enfant 4</i>	9	15	1.75	60
<i>Enfant 5</i>	10	14	1.70	60
<i>Enfant 6</i>	7	14	1.70	60
<i>Enfant 7</i>	8	14	1.60	70
<i>Enfant 8</i>	13	13	1.60	65
<i>Enfant 9</i>	17	15	1.55	60
<i>Enfant 10</i>	11	14	1.70	65

- Déclarer (Créer) la matrice X ainsi que sa transposée X^t . Afficher les deux matrices.
- Donner la liste des individus.
- Extraire les vecteurs colonnes (les variables).
- Accéder (ou Extraire) aux individus 4, 5, 6.
- Créer un tableau noté $X(J)$ qui nous permet d'avoir pour chaque variable les informations suivantes : la moyenne arithmétique, la variance, l'écart type. Dans vos résultats (calculs) fixez le nombre de chiffres décimaux à trois.
- Calculez le centre de gravité du nuage des données. Afficher le résultat.
- Calculer toutes les covariances possibles entre les variables données. Afficher les résultats obtenus dans une matrice notée V .
- Calculer les coefficients de corrélations des couples : (Moyenne, Taille), (Age, Poids). Que remarquez-vous ? Commenter les résultats obtenus.
- Déterminer les valeurs propres de la matrice V ainsi que les vecteurs propres associés aux valeurs propres déterminées.
- Proposer une instruction qui vous permettra de vérifier si un vecteur donné Y est le vecteur propre associé à une valeur propre λ d'une matrice A . Appliquer à ce qui précède (aux résultats de la question 9).

Pour le graphisme

- Représenter graphiquement le nuage des individus dans le plan des couples :

(Moyenne, Taille), et (Age, Poids).

Présenter les deux graphes dans une même fenêtre. C'est-à-dire **partitionner** la fenêtre de visualisation en **deux sous fenêtres** (horizontales ou verticales à vous de choisir). Dans la première sous fenêtre, présenter le graphe des individus dans le premier plan et le deuxième graphe dans la deuxième sous fenêtre. N'oublier pas de **nommer** (légender) les axes pour les deux graphes et de donner un **titre** pour chaque graphe.

- Interprétez les graphes obtenus.