

## Traitement statistique des données

### TP $n^01$

## Analyse en composantes principales

Les fichiers Matlab mis à votre disposition se trouvent à l'adresse :  
<http://www.esiee.fr/~decambro>.

**Pollco2.mat** est un fichier de données

**etoile.m** est une fonction matlab

Vous pouvez récupérer les données en transférant le fichier sur votre compte et en tapant load Pollco2 ou directement à partir du Workspace Matlab.

Vous chargerez 3 variables :

**X** : données numériques

**Pays** : nom des pays

**Source** : variables retenues pour décrire les données

Le but est d'étudier la source d'émission de CO2 pour chacun des pays.

Données OCDE (2005) en Tonnes de CO2 par habitants et par an

- $X_1$  : Transport : émissions dues aux transports à l'intérieur d'un pays
- $X_2$  : Soutages : émissions dues aux carburants utilisés par les transports maritimes et aériens internationaux
- $X_3$  : Energie : émissions dues à la transformation de l'énergie
- $X_4$  : Industrie : émissions dues à l'activité industrielle
- $X_5$  : Résidentiel : émissions dues aux particuliers
- $X_6$  : Autres : émissions dues à d'autres sources (agriculture, pêche, forêts...)

### - A - Statistiques descriptives

1. Déterminer les statistiques descriptives principales des variables : moyenne, écart-type, médiane, minimum, maximum, distance interquartile. Les fonctions Matlab utiles sont : **mean**, **std**, **median**, **min**, **max**, **prctile**.
2. Réaliser un profil en étoile des individus (utiliser la fonction **etoile** fournie).
3. Réaliser la régression linéaire de  $X_2$ , par rapport à  $X_6$ . Peut-on dire que la température de fusion d'un métal dépend de sa dureté ?

## - B - Analyse en composantes principales

1. *Normaliser les données.*
2. *Représenter les variables à l'aide de la fonction **boxplot**.*
3. *Calculer les axes principaux du nuage. Les fonctions Matlab utiles sont : **corrcoef** et **eig**.  
En déduire la fidélité de la représentation dans le plan principal. Faut-il tenir compte du troisième axe ?*
4. *Déterminer les représentations graphiques : cercle des corrélations et projection sur les plans utiles à la compréhension.*
5. *Utiliser ces représentations pour analyser ces données .*

Chaque binôme rendra un rapport comprenant le programme, les résultats et les commentaires avant le 13/10/11.

La qualité de l'interprétation des résultats sera un des éléments principaux pris en compte dans la note.