# Démarche en analyse de données

1. Y a-t-il des groupes de variables ? Oui ? -> AFM (Analyse factorielle multiple)
   * Est-ce que le jeu de donnée est structuré avec différentes sources d’informations, chaque source d’informations contenant plusieurs variables ?
2. Quel est le type d’information ?
   * Est-ce que l’on a un ou plusieurs tableaux de contingence (tableaux avec des lignes et des colonnes et dans une cellule du tableau un effectif) ? -> AFC (analyse des correspondances) ou AFMTC (analyse factorielle multiple sur tableau de contingence) si on veut comparer plusieurs tableaux de contingence
   * Autre type : Tableau avec individus statistique en ligne et des variables en colonne -> Les individus sont décrits par plusieurs variables
     1. Selon la nature des données : ACP si variable quantitative
     2. ACM si variable qualitative
     3. AFDM si données sont mixte
     4. AFM si variable structuré en groupe

# Analyse en composantes principales (ACP)

1. Données – Exemple
2. Etudes des individus
3. Etudes des variables
4. Aides à l’interprétation

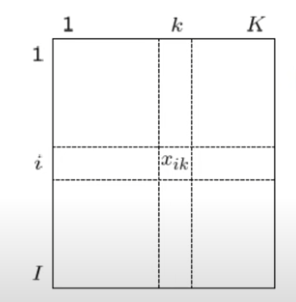
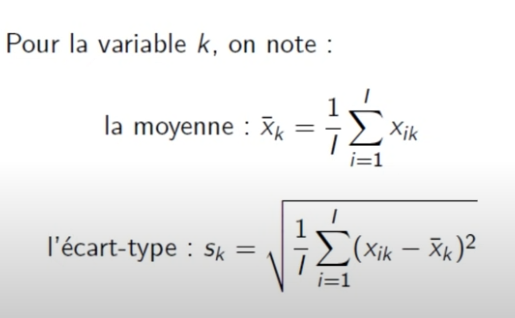
## 1 Données – Exemple

### Quel type de données ?

L’ACP s’intéresse à des tableaux de données rectangulaires avec des individus en lignes et des variables quantitatives en colonnes.

I = individus

K = variables quantitatives

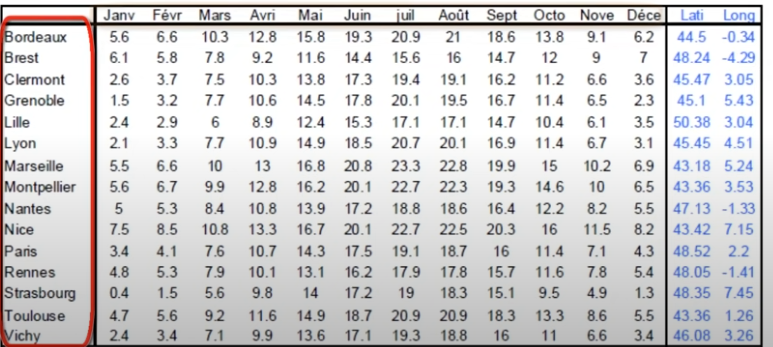


Exemples :

* Analyse sensorielle : note du descripteur k pour le produit i
* Ecologie : concentration du polluant k dans la rivière i
* Economie : valeur de l’indicateur k pour l’année i
* Génétique : expression du gène k pour le patient i
* Biologie : mesure k pour l’animal i
* Marketing : valeur d’indice de satisfaction k pour la marque i
* Sociologie : temps passé à l’activité k par les individus de la CSP i
* Etc .

## Exemple Données température

* 15 individus (lignes) : villes de France
* 14 variables (colonnes) :
  + 12 températures mensuelles moyennes (sur 30 ans)
  + 2 variables géographiques (latitude, longitude)



## Problèmes – objectif

Deux façons de voir le tableau :

* Comme un ensemble de lignes
* Comme un ensemble de colonnes

Etudes des individus

* Quand dit-on que 2 individus se ressemblent du point de vue de l’ensemble des variables ?
* Si beaucoup d’individus, peut-on faire un bilan des ressemblances ?

=>Construction de groupes d’individus, partition des individus

### Etude des variables

* Recherche des ressemblances entre variables
* Entre variables, on parle plutôt de liaisons
* Liaisons linéaires sont simples, très fréquentes et résument de nombreuse liaison = > coefficient de corrélation
* Visualisation de la matrice de corrélations
* Recherche d’un petit nombre d’indicateurs synthétiques pour résumer beaucoup de variables ( ex. d’indicateur synthétique a priori : la moyenne, mais ici on recherche des indicateurs synthétiques a posteriori, à partir des données)