

## TP3 M2 MIAGE FST

Les données qui seront utilisées en TP sont les données `DonnéesGrossistes.txt`.

### Etude du fichier grossiste (`DonnéesGrossistes.txt`)

Vous disposez d'un échantillon extrait d'une base de clients d'une entreprise (BtoB : business to business, vente d'entreprise à entreprises).

Les données sont maquillées, par multiplication par un certain nombre, suivi des arrondis adaptés, mais la structure est entièrement conservée.

Il y a deux familles d'articles : A et B. Les données concernent deux périodes d'activité : 1999, et avant 1999. Pour chaque famille et chaque période on a outre le numéro du client Client 1) le nombre d'articles (contrats) vendus, 2) le chiffre d'affaire total. Les variables sont :

- NA99 nombre de contrats A vendus en 99,
- CA99 chiffre d'affaires contrats A en 99,
- CB99 chiffre d'affaires contrats B en 99,
- NB99 nombre de contrats B vendus en 99,
- CBA99 chiffre d'affaires contrats B avant 99,
- NBA99 nombre de contrats B vendus avant 99,
- CAA99 chiffre d'affaires contrats A avant 99,
- NAA99 nombre de contrats A vendus avant 99,

L'objectif final est de créer une segmentation stratégique de cette clientèle, en 8 classes, (avec une contrainte : la classe des plus gros clients doit comporter environ 160 clients dans votre échantillon). Pour cela, il faudra faire de la descriptive, utiliser l'ACP et clustering, puis transformer les données et recommencer les méthodes factorielles et le clustering (tout cela sur deux Tps).

Ces données vont aussi être utilisées dans le TP4. Vous avez donc intérêt à conserver vos scripts et tous les résultats du TP3. A l'issue du TP4, vous ferez un compte-rendu des TP3 et TP4. Mais attention, il ne s'agit pas de répondre uniquement aux questions des TPs mais de faire des remarques appropriées en tenant compte de l'objectif de l'étude.

1. Lisez ce fichier de données : les séparateurs sont des tabulations et les décimales des virgules.
2. Appelez l'objet résultant "grossiste" et affichez le contenu de grossiste.
3. Faites un tri à plat des variables.

4. Illustrez le problème de ces données en construisant les diagrammes de dispersion des variables deux par deux et en calculant le coefficient de corrélation ( $\text{cor}(\text{grossiste})$ ). Faites le pour toutes les variables. Qu'en pensez-vous.
5. Chargez la librairie FactoMineR. Faites une ACP de ces données et interprétez les résultats en vous inspirant ce que ce qu'on a fait dans le TP sur les iris : combien d'axes conservez-vous ? Pourquoi ? Comment interprétez-vous les axes conservés ? Quels sont les individus qui contribuent le plus à l'inertie du premier, puis du second etc... axe ? Vous pouvez aussi représenter les points dans le plan 3-4 avec la commande `plot.PCA(res.pca,axes=c(3,4))`. Déterminez les points extrêmes. Que concluez-vous ?
6. A cette étape, si vous le jugez pertinent, vous pouvez faire un clustering (charger la librairie cluster) en vous inspirant du TP2. N'oubliez pas de caractériser les classes obtenues.
7. Commentez cette première partie de l'étude.