

# (Re-)prise en main de

## Analyse statistique de données d'enquêtes

---

Cette séance comprend deux parties distinctes. Dans une première partie, nous travaillerons sur un jeu de données d'enquêtes type et nous verrons comment manipuler classiquement un tel jeu de données. Dans une seconde partie, nous travaillerons sur un tableau de contingence et nous réaliserons une analyse des correspondances à l'aide du package *FactoMineR*.

### 1 Manipulation d'un jeu de données

Les données sont issues d'une enquête réalisée lors d'un projet INSFA3 sur la perception des OGM par les français à laquelle ont répondu 135 individus.

Cette enquête comporte 17 questions concernant les OGM :

- Vous sentez-vous concerné par la polémique sur les OGM (beaucoup, moyen, un peu, pas du tout) ?
- Quelle est votre position quant à la culture d'OGM en France (très favorable, favorable, plutôt défavorable, pas favorable du tout) ?
- Quelle est votre position quant à l'incorporation de matière première OGM dans les produits alimentaires destinés à l'alimentation humaine (très favorable, favorable, plutôt défavorable, pas favorable du tout) ?
- Quelle est votre position quant à l'incorporation de matière première OGM dans l'alimentation animale (très favorable, favorable, plutôt défavorable, pas favorable du tout) ?
- Avez-vous déjà participé à une manifestation contre les OGM (oui, non) ?
- Estimez-vous que les médias communiquent suffisamment sur le sujet (oui, non) ?
- Faites-vous vous même la démarche de vous informer sur le sujet (oui, non) ?
- Pensez-vous que les OGM puissent permettre la réduction d'usage de fongicides (oui, non) ?
- Pensez-vous que les OGM puissent permettre la réduction des problèmes de famine dans le monde (oui, non) ?

- Pensez-vous que les OGM puissent permettre l'amélioration des conditions de vie des agriculteurs (oui, non) ?
- Pensez-vous que les OGM puissent permettre de futurs progrès scientifiques (oui, non) ?
- Pensez-vous que les OGM représentent un éventuel danger pour notre santé (oui, non) ?
- Pensez-vous que les OGM représentent une menace pour l'environnement (oui, non) ?
- Pensez-vous que les OGM représentent un risque économique pour les agriculteurs (oui, non) ?
- Pensez-vous que les OGM représentent un procédé scientifique inutile (oui, non) ?
- Pensez-vous que nos grands-parents avaient une alimentation plus saine (oui, non) ?
- Pour vous, quels sont les plus gros risques alimentaires à l'échelle mondiale pour les années à venir (OGM, usage intensif de fongicides, pénurie alimentaire) ?

Cette enquête comporte pour terminer 5 questions de signalétique : sexe ; âge ; CSP ; exercice d'un métier en rapport avec l'agriculture, l'agroalimentaire ou la pharmaceutique ; parti politique.

N'hésitez pas à commenter votre script *R* en commençant le commentaire par un dièse. Les noms des fonctions à utiliser sont indiqués en gras.

1. Enregistrer le jeu de données OGM sur votre ordinateur (le jeu se situe dans l'espace numérique de travail).
2. Charger la librairie *Rcmdr* à l'aide de la commande **library** ou de l'onglet *Packages*.
3. Après l'avoir visualisé dans Excel, importer le jeu de données *ogm.xls*.
4. A l'aide de la fonction **summary**, vérifier que le jeu de données a bien été importé. Modifier l'argument *maxsum* de la fonction **summary** (pour obtenir l'aide, taper *?summary*).
5. Convertir la variable *Sexe* en facteur. Remplacer 1 par Homme et 2 par Femme (option *nom des niveaux*).
6. Créer une nouvelle variable *AgeQuali* en convertissant la variable quantitative *Age* en facteur à 4 modalités : moins de 25, 25-40, 41-60, plus de 60 (utiliser

- l'aide de la fenêtre “recoder des variables”).
7. Fusionner les 2 modalités INSEA et Etudiant de la variable *CSP*.
  8. Représenter les variables suivantes à l'aide d'un diagramme en bâtons (**barplot**) : *Concerné*, *Position.Culture*. Modifier la taille de la police des modalités pour avoir toutes les modalités affichées.
  9. Représenter les variables suivantes à l'aide d'un diagramme en camembert (**pie**) : *CSP*, *Position.Al.H*, *Position.Al.A*.
  10. Représenter la variable quantitative *Age* à l'aide d'un histogramme (**hist**) et d'une boîte à moustache (**boxplot**).
  11. Sélectionner dans votre jeu de données uniquement les étudiants. Résumer la variable *Age* pour ces individus.  
À l'aide de l'interface *Rcmdr*, résumer automatiquement (**numSummary**) la variable *Age* selon les modalités de la variable *CSP*. Représenter cela à l'aide d'une boîte à moustaches.
  12. Construire le tableau de contingence (**table**) croisant les modalités des variables *Sexe* et *Concerné*.  
À l'aide de l'interface *Rcmdr*, calculer les pourcentages en ligne (**rowPercents**). Interpréter ces pourcentages.  
À l'aide de l'interface *Rcmdr*, calculer les pourcentages en colonne (**colPercents**). Interpréter ces pourcentages.  
Représenter ce croisement par des diagrammes en bâtons (**barplot** : arguments **beside** et **legend.text**).
  13. Rappeler les hypothèses du test du  $\chi^2$ .  
Tester l'indépendance des variables *Sexe* et *Concerné* à l'aide d'un test du  $\chi^2$  (**chisq.test**).  
Quels éléments rend la fonction **chisq.test** (**names**) ?  
Vérifier que la probabilité critique fournie par le test est bien la bonne (**pchisq**).

## 2 Les rouges à lèvres

Les données que vous allez analyser sont issues d'une expérience réalisée par des étudiants de la “*spé stat*”. Au cours de cette expérience on demande à des consommatrices de réaliser une tâche d'association. Pour cela elles disposent de 8 rouges à lèvres d'une part, de 21 images d'autre part. Les rouges à lèvres sont des produits premium (rose, brun, rouge, prune ; mat et brillant), les images correspondent à des états émotionnels (amoureuse, romantique, sauvage,...). Pour un rouge à lèvres donné, il s'agit tout simplement de lui associer une ou plusieurs images, en fonction de ce que le produit évoque.

Après avoir importé le jeu de données *Rouges à lèvres*, réaliser une analyse des correspondances du tableau de contingence qui croise les rouges à lèvres d'une part, les émotions d'autre part. Pour cela, on utilisera la librairie *FactoMineR*. Commenter.

1. Jouer avec les options de la fonction **plot.CA** du package *FactoMineR* pour représenter les colonnes seulement.
2. Jouer avec les options de la fonction **plot.CA** du package *FactoMineR* pour représenter les lignes seulement.
3. Jouer avec les options de la fonction **plot.CA** du package *FactoMineR* pour représenter les 10 colonnes les plus contributives.
4. Avec la fonction **dimdesc** du package *FactoMineR* décrire les deux premières dimensions issues de l'AC du jeu de données.
5. Quelle est l'inertie du premier axe factoriel ?
6. Jouer avec les packages *explor* et *Factoshiny* pour explorer vos plans factoriels de façon interactive. A votre avis, en quoi ces deux packages sont-ils complémentaires ?
7. Regarder ce que propose le package *factoextra* (aller voir sur internet).

### 3 Que lisent les étudiants dans le journal *Ouest-France* ?

Aujourd'hui, vous allez jouer avec vos propres données. Pour cela, il vous faudra d'abord importer le jeu de données *OuestFrance.txt*.

1. Pour savoir si l'intérêt que l'on porte à certains articles est fonction de l'intérêt que l'on porte à la vie politique, vous allez utiliser les fonctions **textual** et **descfreq** du package *FactoMineR*. Pour cela, créer le tableau de contingence croisant les articles en colonnes, l'intérêt que l'on porte à la vie politique en lignes, à l'aide de la fonction **textual**. Décrire les lignes en fonctions des colonnes avec la fonction **descfreq**.
2. Réaliser une analyse des correspondances de ce même tableau de contingence. Commenter.
3. Réaliser une classification des articles.