

Exercice 1 (Dépenses des ménages)

Les données. Le jeu de données (fichier `base_conso.csv`) est issu d'une enquête *budget de famille* menée par l'INSEE en 2006. Ces enquêtes permettent de connaître le poids des grands postes de consommation dans le budget des ménages. Chaque ligne du tableau correspond à une classe d'âge et chaque colonne correspond à une rubrique de dépense. Nous disposons de 30 variables quantitatives (26 correspondent à différentes rubriques, trois correspondent à des totaux partiels et une à la dépense totale). On dispose également de la répartition de la dépense annuelle pour l'ensemble de la population française, ainsi que de la répartition des dépenses en fonction du revenu des ménages. Les revenus sont regroupés par décile: le décile 1 correspond aux 10% des ménages ayant les revenus les plus faibles, le décile 2 concerne les personnes ayant les revenus entre 10 et 20% les plus faibles, ...

Problématique. On s'intéresse à l'évolution du profil des dépenses selon l'âge. Ainsi, on veut obtenir une typologie des classes d'âge selon les différents postes de dépenses. On veut également obtenir un bilan des liaisons entre les postes de dépenses.

Exercice 2 (Causes de mortalité des français)

Les données. On dispose, pour chaque année de 1979 à 2006 du tableau de contingence croisant, pour la population française, les causes de décès et l'âge subdivisé en tranches d'âge (fichier `deces.csv`). Pour simplifier, on considère principalement les deux tableaux correspondant aux années 1979 et 2006 ainsi que leur somme. On considère enfin le tableau croisant, toujours pour la population française mais cette fois sur l'ensemble de la période allant de 1979 à 2006, et toutes causes confondues, les années et les tranches d'âge.

Problématique. On souhaite étudier la liaison entre l'âge et la cause du décès.

Exercice 3 (Perception des OGM)

Les données. On considère une enquête conduite par deux étudiants d'Agrocampus sur 135 personnes visant à avoir une vue d'ensemble de leurs différentes prises de position concernant les OGM (fichier `ogm.csv`). Il leur a été posé un ensemble de 21 questions fermées que nous avons réparties en deux groupes.

Un premier groupe composé de seize questions en lien direct avec le rapport aux OGM qu'ont les personnes interrogées:

- Vous sentez-vous concerné par la polémique sur les OGM.
- Quelle est votre position quand à la culture d'OGM en France.
- Quelle est votre position quant à l'incorporation de matière première OGM dans les produits alimentaires destinés à l'alimentation humaine.
- Quelle est votre position quant à l'incorporation de matière première OGM dans les produits alimentaires destinés à l'alimentation animale.
- Avez-vous déjà participé à une manifestation contre les OGM.
- Estimez-vous que les médias communiquent suffisamment sur le sujet.
- Faites-vous vous-même la démarche de vous informer sur le sujet.
- Pensez-vous que les OGM puissent permettre la réduction d'usage de fongicides.
- Pensez-vous que les OGM puissent permettre la réduction des problèmes de famine dans le monde.
- Pensez-vous que les OGM puissent permettre l'amélioration des conditions de vie des agriculteurs.

- Pensez-vous que les OGM puissent permettre de futurs progrès scientifiques.
- Pensez-vous que les OGM représentent un éventuel danger pour notre santé.
- Pensez-vous que les OGM représentent une menace pour l'environnement.
- Pensez-vous que les OGM représentent un risque économique pour les agriculteurs.
- Pensez-vous que les OGM représentent un procédé scientifique inutile.
- Pensez vous que nos grands-parents avaient une alimentation plus saine.

Un second groupe composé de cinq variables de signalétique au sens large:

- Sexe
- Catégorie soci-professionnelle
- Tranche d'âge
- Exercez vous des études ou un métier en rapport avec l'agriculture, l'agroalimentaire ou la pharmacutique
- A quel parti politique vous identifiiez vous le plus

Problématique. À travers ce questionnaire, on cherche à obtenir une typologie des personnes interrogées en fonction de leur rapport aux OGM; à voir si cette typologie n'est pas sans lien avec les variables de signalétique d'autre part.

Exercice 4 (Les états des USA)

Les données. De nombreuses données concernant les 50 états des USA sont disponibles dans R en faisant `data(state)`. Les variables `state.center`, `state.x77` et `sate.abb` correspondent respectivement aux centres géographiques des états (longitude et latitude), à diverse statistiques les concernant (population, espérance de vie, etc...). et aussi à l'abréviation de leur nom.

Problématique. L'objectif est de regrouper les états des USA en classes homogènes de sorte que les états d'une même classe présentent des caractéristiques semblables (population, espérance de vie, etc...).