
TP Analyse bivariée

Analyse factorielle des correspondances

L'objectif de ce TP est d'étudier le lien éventuel entre deux séries de données qualitatives et de le représenter graphiquement.

Exercice 1

Données : Marketing.txt

Une entreprise veut choisir un nom pour un nouveau produit. Une série de noms et de qualificatifs sont proposées à un panel de consommateurs. L'objectif est de déterminer le nom qui fasse apparaître le nouveau produit comme de qualité supérieur, prestigieux, luxueux, convenant à un public masculin, connaisseur, raffiné, distingué, de niveau socio-économique supérieur,... (Benzécri, 1980).

- La variable 1 est le qualificatif avec 11 modalités : *Vieillot, Nouveau riche, Sobre/Elégant, Cocasse, Racé, Mièvre, Distingué, Vulgaire, Pour homme, Pour femme, Pour petite nature.*
- La variable 2 est le nom avec 12 modalités : *Orly, Alezan, Corsaire, Directoire, Ducat, Fontenoy, Icare, Zodiaque, Pavois, Cocker, Escale, Hôtesse.*

1) Etude des différents tableaux avec un tableur

(a) Profils lignes

- Quel est le profil moyen des qualificatifs ?
- Quels sont les qualificatifs qui s'éloignent le plus du profil moyen ?

(b) Profils colonnes

- Quel est le profil moyen des noms ?
- Quels sont les noms qui s'éloignent le plus du profil moyen ?

(c) Tableau des contributions

- Calculer la distance du chi-deux.
- Quelles cases contribuent le plus à la distance du chi-deux ?

2) Analyse factorielle des correspondances

- (a) Quel est le nombre d'axes ? Combien d'axes permettent de restituer un pourcentage suffisant de la distance du chi-deux ?

(b) Etude des qualificatifs sur l'axe 1

Si on considère le nuage de points des qualificatifs. Quels sont les individus dont la contribution à la distance du chi-deux (inertie) est supérieure à la moyenne ? Pour chacun d'eux, préciser le signe de sa coordonnée sur l'axe. En déduire une interprétation de l'axe 1.

(c) Etude des noms sur l'axe 1

Faire la même analyse avec le nuage de points des noms. Quels sont les noms bien représentés sur cet axe ? Que peut-on en dire ?

(d) Faire le même travail sur l'axe 2.

(e) Y-a-t'il des noms éloignés du profil moyen et qui sont mal représentés sur le plan principal ? Si oui, trouver sur quels axes ils sont bien représentés, dire si on peut les interpréter et conclure.

Exercice 2

Données : Nobel.txt

Le fichier contient la répartition des prix Nobel par discipline et par pays. Il s'agit des prix Nobel attribués depuis 1945 et jusqu'à 2013. Seuls les pays ayant obtenus plus de 10 prix Nobel sont concernés. Les données sont issues du site officiel des prix Nobel : <http://www.nobelprize.org/>

- La variable 1 est la discipline avec 6 modalités : *Physique, Chimie, Médecine, Littérature, Paix, Economie*.
- La variable 2 est le pays avec 9 modalités : *Allemagne, France, UK, US, Suède, Russie, Canada, Italie, Japon*.

1) Etude des profils lignes et colonnes

- (a) Quels sont les pays qui respectent le profil colonne moyen ? Quels sont ceux qui s'en éloignent et de quelle manière ?
- (b) Quelles sont les disciplines qui respectent le profil ligne moyen et quelles sont celles qui s'en éloignent ?

2) Quel est le nombre d'axes ? Combien d'axes permettent de restituer un pourcentage suffisant de la distance du chi-deux ?

3) Etudier les résultats de l'AFC.

Exercice 3

Données : IndustrieInfoAFC.txt

Le tableau de contingence comptabilise les 275 plus grandes entreprises de l'industrie du numérique en terme de capital en 2005 (source : Eurostat).

- La variable 1 est le continent de localisation de l'entreprise avec 3 modalités : *Asie, Europe, Amérique*.
- La variable 2 est le secteur d'activité avec 5 modalités : *Hardware, software, Internet, Mobile-Telecom, Services*.

Faire une étude entre les deux variables en suivant le plan défini dans les exercices précédents.