

**TP n°2 – Analyse Factorielle des Correspondances : AFC:  
avec R ou Rstudio**

**ENQUETE DE LA REGIE FRANCAISE DES TABACS**

Cette étude de cas est extraite de l'ouvrage de J.P. Benzecri "Analyse des données. Tome 2: l'analyse des correspondances". Une enquête a été effectuée auprès de cent fumeurs afin de choisir les noms de deux nouvelles marques de cigarettes. La première marque est destinée à une clientèle à une clientèle masculine: l'homme ciblé est un connaisseur distingué, raffiné mais viril, de niveau socio-économique élevé. La seconde vise un public féminin, élégant, assuré, dynamique.

Douze marques possibles ont été retenues:

I = {Orly (ORLY), Alezan (ALEZ), Corsaire (CORS), Directoire (DIRE), Ducat (DUCA), Fontenoy (FONT), Icare(ICAR), Zodiac(ZODI), Pavois (PAVI), Cocker(COCK), Escalé (ESCA), Hôtesse (HOTE)}.

Pour évaluer leur image auprès du public, onze attributs ont été proposés aux fumeurs:

J = {Vieillot-désuet (viel), Nouveau-Riche(Nouv), Sobre-élégant(Sobr), Cocasse-Ridicule(Coca), Racé (racé), Mièvre(miev), Distingué(dist), Vulgaire-Commun (vulg), Pour un Homme (hom), Pour une femme (fem), Pour une petite nature (petit)}.

Chaque personne interrogée devait successivement pour chaque attribut, désigner une ou plusieurs marques lui correspondant. Le tableau de données est le suivant:

libellé	ORLY	ALEZ	CORS	DIRE	DUCA	FONT	ICAR	ZODI	PAVI	COCK	ESCA	HOTE
viel	1.000	2.000	14.000	38.000	18.000	10.000	9.000	5.000	9.000	4.000	0.000	1.000
nouv	20.000	9.000	1.000	11.000	10.000	9.000	1.000	1.000	20.000	9.000	7.000	12.000
sobr	9.000	23.000	1.000	15.000	7.000	11.000	6.000	2.000	7.000	12.000	3.000	17.000
coca	1.000	3.000	15.000	15.000	6.000	5.000	12.000	18.000	4.000	25.000	2.000	2.000
racé	4.000	33.000	7.000	8.000	3.000	6.000	6.000	4.000	5.000	15.000	5.000	3.000
miev	3.000	9.000	1.000	7.000	7.000	5.000	12.000	9.000	6.000	9.000	6.000	13.000
dist	11.000	9.000	1.000	17.000	4.000	21.000	6.000	1.000	5.000	4.000	5.000	27.000
vulg	4.000	4.000	32.000	2.000	6.000	0.000	9.000	7.000	3.000	10.000	12.000	7.000
hom	9.000	12.000	23.000	4.000	7.000	13.000	5.000	5.000	10.000	5.000	13.000	9.000
fem	9.000	3.000	9.000	8.000	4.000	2.000	6.000	8.000	1.000	6.000	23.000	33.000
petit	7.000	5.000	2.000	7.000	11.000	2.000	6.000	11.000	9.000	24.000	10.000	5.000

Les valeurs dans le tableau sont des nombres entiers. Chaque valeur du tableau correspond au nombre de personnes ayant choisi simultanément les deux modalités (lignes et colonnes) considérées. Par exemple la valeur 1 de la première et de la première colonne signifie qu'une seule personne parmi l'échantillon associe les modalités VIEL et ORLY. Alors 20 personnes interrogées associent NOUV à ORLY etc. On veut mettre en œuvre l'Analyse Factorielle des Correspondances (AFC) en utilisant le logiciel R. La démarche est la même qu'en ACP.

**Première partie :**

Créer la base de données en format txt puis importer cette base de données R.

**Deuxième partie :**

Mettre en œuvre la méthode AFC à l'aide de R:

**Troisième partie :** Analyse des résultats

- 1 – Combien d'axes peut-on retenir et pourquoi ?
- 2 – Interpréter les quatre premiers axes en fonction des marques
- 3 - Interpréter les quatre premiers axes en fonction des attributs
- 4 – Projeter les marques et attributs dans le premier plan factoriel et interpréter.

**Quatrième partie :**

Mettre en illustratifs Corsaire, Distingué et Vulgaire et interpréter les analyses, les graphiques,....