

Analyse de données – TP n° 5 Analyses Factorielles des Correspondances (AFC)

1 AFC des données utilisées en cours

Il s'agit ici d'apprendre à utiliser la fonction `dudi.coa` du package `ade4`, afin d'effectuer l'AFC de la table de contingence utilisée en cours, qui croise la couleur des yeux et des cheveux de $n = 592$ femmes¹, et dont l'analyse a donc déjà été faite.

(1) Chargez la table à partir du fichier `couleursYeuxCheveux.txt`. Remarquez qu'il ne s'agit pas de la table habituelle au format individus-caractères (table qui ferait quelles dimensions?).

Donnez les profils ligne et colonne, ainsi que l'histogramme des profils (révision).

(2) Consulter l'aide `?dudi.coa`, puis effectuez l'AFC de cette table. Analyser les sorties, et calculer la décomposition de l'inertie avec `inertia.dudi` (la même fonction que pour l'ACP). Remarque : la méthode `summary` pour l'AFC donne aussi ce résultat.

(3) Vérifiez pour cet exemple, avec R, le résultat général montrée en cours liant l'inertie I du nuage et la statistique du test du χ^2 d'indépendance des deux variables.

(4) La fonction `?scatter.coa` donne les trois représentations simultanées présentées en cours, mais son aide n'est pas très explicite : quelle méthode correspond à quelle représentation de l'AFC ?

(5) Écrivez un code R qui donne la représentation simultanée avec coloration par variable et tailles des points proportionnelles aux \cos^2 . Il faudra pour cela comprendre ce que la procédure d'AFC a retourné dans les sorties `$li`, `$co`, etc.

1. source : Cohen 1980, et Lebart et. al. 1997

2 AFC enquête *Emploi des cadres de l'agriculture*

On veut étudier le tableau croisé ci-dessous, issu d'une enquête de l'Association pour l'Emploi des Cadres de l'Agriculture, en 1975. Cette table donne la répartition par secteur et par classe d'âge des 76155 cadres de l'agriculture ayant répondu au questionnaire. Elle permet d'étudier la structure d'âge des différents secteurs.

SECTEURS/AGES	20–29	30–39	40–49	50–59	60–65
Secteur public (SP)	1480	1805	2603	1606	576
Secteur public autonome (SPA)	1409	1375	1251	650	164
Organ. professionnelles (OP)	2916	6875	4943	2099	489
Industries agricoles (IA)	1049	4870	4494	2380	1124
Exploitations (Exp)	1058	3385	4979	2341	1534
Enseignement privé (EP)	3864	1278	703	370	151
Para-agricole (PA)	700	2189	2153	700	256
Outre-mer (OM)	430	810	533	160	40
Divers (Div)	55	276	81	42	20
Agents spécialisé (AS)	1061	1676	816	309	27

Les questions typiques que l'on se pose à partir d'un tel tableau croisé sont :

1. la structure diffère-t-elle beaucoup d'un secteur à l'autre ?
2. y-a-t-il des secteurs ayant des structures semblables ?
3. quels sont les secteurs “jeunes”, les secteurs “vieux” ?
4. dans quels secteurs les jeunes sont-ils entrés durant les dix dernières années ?
5. dans quels secteurs y aura-t-il beaucoup de départs à la retraite au cours des dix prochaines années ?

(6) Le fichier pour cette table est `agriculture.txt`. Affichez avec R la table des profils lignes. Quelles distributions sont ici représentées ? Auxquelles des questions ci-dessus cette table permet-elle de répondre le mieux ?

(7) Même question avec les profils colonne.

(8) Effectuez l'AFC de cette table avec la fonction `dudi.coa` du package `ade4`, et répondez aux questions ci-dessus à l'aide des sorties de l'analyse, en adaptant les codes détaillés dans l'exemple précédent.