



الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة وهران للعلوم والتكنولوجيا محمد بوضياف
كلية الرياضيات والإعلام الآلي - قسم الإعلام الآلي



Fiche TD n°2 : Analyse factorielle des correspondances

Ingénieur3- Spécialité : Data science

Analyse de données avancée

Exercice 1 :

Soit le tableau de contingence (variable/variable) Couleur yeux\Cheveux suivant :

Yeux\Cheveux	Blonds	Marrons	Noirs
Vert	50	10	3
Bleu	43	8	0
Marron	8	66	30
Noir	2	44	50

1. Donner tous les tableaux de contingences et les nuages de points $N(I)$ et $N(J)$ correspondants.
2. Représenter graphiquement les deux nuages ($N(I)$, $N(J)$) sur le même espace de représentation
3. Interpréter les résultats

Exercice 2 (à faire) :

Soit le tableau de contingence suivant entre le niveau d'études (Bac, Licence, Master) et le secteur d'emploi (Public, Privé, Indépendant) :

Niveau \ Secteur	Public	Privé	Indépendant
Bac	25	40	10
Licence	35	60	20
Master	20	30	15

1. Appliquer une AFC
2. Interpréter les résultats

Analyse Factorielle des Correspondances (AFC) :

L'Analyse Factorielle des Correspondances (AFC) est une méthode statistique exploratoire appartenant à la famille des analyses factorielles. Elle est spécifiquement conçue pour analyser les relations entre variables catégorielles présentées dans un tableau de contingence

Tableau de données :

- 1) Tableau de contingence (variable/variable) : K_{ij} avec $K = \sum_i \sum_j K_{ij}$
- 2) Tableau de fréquence : $f_{ij} = \frac{K_{ij}}{K}$
- 3) Tableau de profils lignes : $F_j^i; F_j^i = \frac{f_{ij}}{f_{i.}}$ tel que : $f_{i.} = \sum_{j=1}^p f_{ij}$ fréquence marginale ligne
- 4) Tableau des profils colonnes $F_i^j; F_i^j = \frac{f_{ij}}{f_{.j}}$ tel que : $f_{.j} = \sum_{i=1}^n f_{ij}$ fréquence marginale colonne

Le nuage	$N(I) = \{(F_j^i, f_{i.}), i = 1 \dots n\}$ F_j^i : Individu $f_{i.}$: Poids	$N(J) = \{(F_i^j, f_{.j}), j = 1 \dots p\}$ F_i^j : Individu $f_{.j}$: Poids
La métrique utilisée	$M = D_{\frac{1}{f_{i.}}}$	$M = D_{\frac{1}{f_{.j}}}$
Distance entre deux profils : χ^2	$d^2(F_j^i, F_j^{i'}) = \ F_j^i - F_j^{i'}\ _{D_{\frac{1}{f_{i.}}}}^2$ $= \sum_{j=1}^p \frac{(F_j^i - F_j^{i'})^2}{f_{i.}}$	$d^2(F_i^j, F_i^{j'}) = \ F_i^j - F_i^{j'}\ _{D_{\frac{1}{f_{.j}}}}^2$ $= \sum_{i=1}^n \frac{(F_i^j - F_i^{j'})^2}{f_{.j}}$
Centre de gravité du nuage $g = \frac{\sum_{i=1}^n p_i \cdot x_i}{\sum_{i=1}^n p_i}$	$g_{N(I)} = \frac{\sum_{i=1}^n f_{i.} \cdot F_j^i}{\sum_{i=1}^n f_{i.}} = \frac{\sum_{i=1}^n f_{i.} \cdot \frac{f_{ij}}{f_{i.}}}{\sum_{i=1}^n f_{i.}}$ $= \frac{\sum_{i=1}^n f_{ij}}{\sum_{i=1}^n f_{i.}}$ $\frac{f_{.j}}{\sum_{i=1}^n \sum_{j=1}^p f_{ij}} = \frac{f_{.j}}{\sum_{i=1}^n \sum_{j=1}^p \frac{K_{ij}}{K}}$ $= \frac{f_{.j}}{\frac{f_{.j}}{\frac{1}{K} \sum_{i=1}^n \sum_{j=1}^p K_{ij}}} = \frac{f_{.j}}{\frac{1}{K} * K}$ $= f_{.j}$	$g_{N(J)} = \frac{\sum_{j=1}^p f_{.j} \cdot F_i^j}{\sum_{j=1}^p f_{.j}} = \frac{\sum_{j=1}^p f_{.j} \cdot \frac{f_{ij}}{f_{.j}}}{\sum_{j=1}^p f_{.j}}$ $= \frac{\sum_{j=1}^p f_{ij}}{\sum_{j=1}^p f_{.j}}$ $\frac{f_{i.}}{\sum_{j=1}^p \sum_{i=1}^n f_{ij}} = \frac{f_{i.}}{\sum_{j=1}^p \sum_{i=1}^n \frac{K_{ij}}{K}}$ $= \frac{f_{i.}}{\frac{f_{i.}}{\frac{1}{K} \sum_{j=1}^p \sum_{i=1}^n K_{ij}}} = \frac{f_{i.}}{\frac{1}{K} * K}$ $= f_{i.}$
L'inertie du nuage : $I = \sum_{i=1}^n p_i d^2(x_i, g)$	$I_{N(I)} = \sum_{i=1}^n f_{i.} d^2(F_j^i, f_{.j})$ $d^2(F_j^i, f_{.j}) = \ F_j^i - f_{.j}\ _{D_{\frac{1}{f_{i.}}}}^2$ $= \sum_{j=1}^p f_{i.} \sum_{j=1}^p \frac{(F_j^i - f_{.j})^2}{f_{i.}}$	$I_{N(J)} = \sum_{j=1}^p f_{.j} d^2(F_i^j, f_{i.})$ $d^2(F_i^j, f_{i.}) = \ F_i^j - f_{i.}\ _{D_{\frac{1}{f_{.j}}}}^2$ $= \sum_{j=1}^p f_{.j} \sum_{i=1}^n \frac{(F_i^j - f_{i.})^2}{f_{i.}}$

Corrigé exercice 2 :

1. Le même principe que l'exercice 1

Tableau de contingence :

Yeux\Cheveux	Blond	Marron	Noir
Vert	50	10	3
Bleu	43	8	0
Marron	8	66	33
Noir	2	44	50

Tableau de fréquence : $f_{ij} = \frac{K_{ij}}{K}$ $K = \sum_i \sum_j K_{ij} = 317$

\Cheveux Yeux	Blond	Marron	Noir	fi.
Vert	50/317	10/317	3/317	63/317
Bleu	43/317	8/317	0	51/317
Marron	8/317	66/317	33/317	107/317
Noir	2/317	44/317	50/317	96/317
f.j	103/317	128/317	86/317	1.0

\Cheveux Yeux	Blond	Marron	Noir	fi.
Vert	0.1577	0.0315	0.0095	0.2
Bleu	0.1356	0.0252	0.0	0.16
Marron	0.0252	0.2082	0.1041	0.34
Noir	0.0063	0.1388	0.1577	0.3
f.j	0.32	0.4	0.27	1.0

Tableau de profils Ligne : $F_j^i: F_j^i = \frac{f_{ij}}{f_i}$ tel que : $f_i = \sum_{j=1}^p f_{ij}$

\Cheveux Yeux	Blond	Marron	Noir
Vert	50/63	10/63	3/63
Bleu	43/51	8/51	0
Marron	8/107	66/107	33/107
Noir	2/96	44/96	50/96
f.j	103/317	128/317	86/317

\Cheveux Yeux	Blond	Marron	Noir
Vert	0.4928	0.07875	0.0352
Bleu	0.42375	0.063	0.0
Marron	0.07875	0.5205	0.3856
Noir	0.0197	0.347	0.5841
f.j	0.32	0.4	0.27

Tableau de profils colonne : $F_i^j: F_i^j = \frac{f_{ij}}{f_j}$ tel que : $f_j = \sum_{i=1}^n f_{ij}$

\Cheveux Yeux	Blond	Marron	Noir	fi.
Vert	50/103	10/128	3/86	63/317
Bleu	43/103	8/128	0.0	51/317
Marron	8/103	66/128	33/86	107/317
Noir	2/103	44/128	50/86	96/317

\Cheveux Yeux	Blond	Marron	Noir	fi.
Vert	0.7885	0.1575	0.0475	0.2
Bleu	0.8475	0.1575	0.0	0.16
Marron	0.0741	0.612	0.3063	0.34
Noir	0.021	0.463	0.5257	0.3

2. Appliquez l'algorithme de AFC

5) Calculer $X = D_i^{-\frac{1}{2}} \cdot (f_{ij} - f_i \cdot f_j^t) \cdot D_j^{-\frac{1}{2}}$ Avec $D_i^{-\frac{1}{2}} = D_{\frac{1}{\sqrt{f_i}}}$ et $D_j^{-\frac{1}{2}} = D_{\frac{1}{\sqrt{f_j}}}$

$D_i^{-\frac{1}{2}} =$

2.242	0.0	0.0	0.0
0.0	2.492	0.0	0.0
0.0	0.0	1.720	0.0
0.0	0.0	0.0	1.817

$D_j^{-\frac{1}{2}} =$

1.754	0.0	0.0
0.0	1.573	0.0
0.0	0.0	1.921

$(f_{ij} - f_i \cdot f_j^t) =$

0.09302	-0.0489	-0.0444
0.0833	-0.0399	-0.0436
-0.0847	0.07165	0.0125
-0.0922	0.0164	0.0756

$X = D_i^{-\frac{1}{2}} \cdot (f_{ij} - f_i \cdot f_j^t) \cdot D_j^{-\frac{1}{2}} =$

0.3658	-0.1724	-0.1913
0.3641	-0.1562	-0.2089
-0.2554	0.1939	0.0413
-0.2937	0.0468	0.2638

6) Calculer $S = X^t \cdot X$

$S = X^t \cdot X$

0.4178	-0.183	-0.2341
-0.1832	0.09393	0.08599
-0.2341	0.08599	0.15151

7) Calculer les valeurs propres et vecteurs propres λ_k et U_k de S

Valeurs propres : $\lambda = 0.62988$ $\lambda = 0.000001$ $\lambda = 0.033405$

Trier les valeurs propres : $\lambda_1 = 0.62988$ $\lambda_2 = 0.033405$ $\lambda_3 = 0.000001$

Pour une présentation graphique nous fixons le nombre de valeurs propres a 2 (2 axes de présentation)

Donc les valeurs propres : $\lambda_1 = 0.62988$ $\lambda_2 = 0.033405$

Les vecteurs propres :

U_1	U_2
0.814	0.111
-0.352	-0.686
-0.462	0.71921

8) Calculer les composantes principales

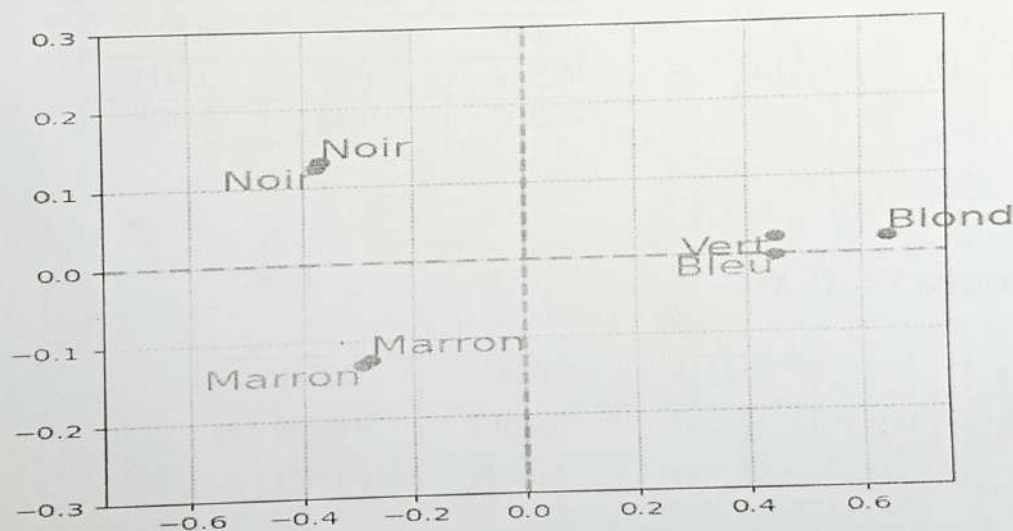
- Profils lignes : $C_K^I = X \cdot U_k$

Yeux	C_1	C_2
Vert	0.4469	0.0213
Bleu	0.4478	-0.0027
Marron	-0.2953	-0.1316
Noir	-0.3774	0.125

- Profils colonne $C_k^J = \sqrt{\lambda_k} \cdot U_k$

Cheveux	C_1	C_2
Blond	0.646	0.0203
Marron	-0.2797	-0.1254
Noir	-0.3664	0.1314

9) Présentation graphique :



Interprétation des résultats :

Pour l'axe 1, on observe une opposition entre les couleurs de cheveux **noirs et blonds** d'un côté, et la couleur **marron** de l'autre. Pour l'axe 2, une opposition se dessine entre les couleurs de cheveux **noirs et marron** d'une part, et la couleur **blonde** d'autre part.

À partir de ces oppositions, nous pouvons déduire que les individus ayant des cheveux noirs ont généralement des yeux noirs, ceux ayant des cheveux marrons ont des yeux marrons, et ceux ayant des cheveux blonds ont des yeux verts ou bleus.