

Analyse Factorielle des Correspondances Un exemple numérique complet

Considérons le tableau de contingence N à n lignes et p colonnes obtenu en ventilant une population de 592 femmes suivant leurs couleurs des yeux et des cheveux.

		couleur des cheveux				
		brun	châtain	roux	blond	Total ($n_{i.}$)
Couleur Des Yeux	marron	68	119	26	7	220
	noisette	15	54	14	10	93
	vert	5	29	14	16	64
	bleu	20	84	17	94	215
Total ($n_{.j}$)		108	286	71	127	592

En lignes : la variable "couleur des yeux" à $n = 4$ modalités (ou catégories) et en colonnes est donnée la variable "couleur des cheveux" à $p = 4$ modalités.

A l'intersection d'une ligne et d'une colonne, nous avons le nombre n_{ij} de femmes ayant simultanément la couleur i des yeux et la couleur j de cheveux.

Le total marginal $n_{i.}$ est le nombre de femmes ayant les yeux de couleur i , alors que le total marginal $n_{.j}$ est le nombre de femmes ayant les cheveux de couleur j .

Question : Y- a- t- il indépendance entre la couleur des yeux et celle des cheveux ? Sinon quels types d'associations existent entre ces couleurs ?

1- Hypothèse d'indépendance

On pose l'hypothèse d'indépendance H_0 énoncé comme suit :

« Les deux variables sont indépendantes » à un risque d'erreur égal à 1 % (0.01)

Calcul du Khi-deux (écart à l'indépendance) :
$$d^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n} \right)^2}{\frac{n_{i.} n_{.j}}{n}}$$
 à partir de la matrice de

contingence. On trouve $d^2=138.29$

D'autre part, on détermine le degré de liberté d'une loi Khi-deux, $ddl = (n-1)(p-1)=9$, puis on cherche la valeur du khi-deux à partir de la table Khi-deux selon le $ddl=9$ et le risque d'erreur=0.01, on trouve cette valeur égale à 21.67

On a $d^2=138.29 > 21.67$, par conséquent on rejette l'hypothèse H_0

Conclusion : les deux variables sont dépendantes. Il y a lieu de faire une AFC.

2- Transformations du tableau de contingence

Chaque élément du tableau profils- lignes est obtenu par la formule: $n_{ij}/n_{i.}$

Chaque élément du tableau profils- colonnes est obtenu par la formule: $n_{ij}/n_{.j}$

Tableau Profils- lignes (pourcentages- lignes arrondis multipliés par 100)

		couleur des cheveux			
		brun	châtain	Roux	blond
Couleur Des Yeux	marron	31	54	12	3
	noisette	16	58	15	11
	vert	8	45	22	25
	bleu	9	39	8	44

- Nuage des n lignes

L'ensemble des profils- lignes forme un nuage de n points dans l'espace des p colonnes. Chaque point i a pour coordonnées dans R^p : $n_{ij}/n_{i.}$; $j= 1,2, \dots,p$.

Il est affecté d'un poids $n_{i.}$ (marge en lignes).

Le centre de gravité de ce nuage est la moyenne des profils- lignes affectés de leurs poids. Sa j ème

composante vaut : $\sum_{i=1}^n n_{i.} \frac{n_{ij}}{n_{i.}}$ C'est la marge en colonnes.

Tableau Profils- colonnes (pourcentages- colonnes arrondis multipliés par 100)

		couleur des cheveux			
		brun	Châtain	Roux	blond
Couleur	marron	63	42	37	6
Des	noisette	14	19	20	8
Yeux	vert	5	10	20	13
	bleu	19	29	24	74
	total	100	100	100	100

- Nuage des n lignes

L'ensemble des profils- colonnes forme un nuage de p points dans l'espace des n lignes. Chaque point j a pour coordonnées dans R^n : $n_{ij}/n_{.j}$; $i= 1,2, \dots,n$.

Il est affecté d'un poids $n_{.j}$ (marge en colonnes).

Le centre de gravité de ce nuage est la moyenne des profils- colonnes affectés de leurs poids. Sa i ème

composante vaut : $\sum_{j=1}^p n_{.j} \frac{n_{ij}}{n_{.j}}$ C'est la marge en lignes.

3- ACP des deux nuages :

- Valeurs propres :

Le nombre des valeurs propres est égal à $\min(n-1)(p-1)=3$

La valeur propre égale à 1 associée au vecteur propre centre de gravité est éliminée

On trouve les valeurs propres suivantes : 0.2088 0.0222 0.0026

Total des valeurs propres = inertie totale = $d^2/n = 0.2336$

On choisira un plan composé des deux premiers axes, ce qui donne un pourcentage d'inertie expliquée égal à 98.99%.

ACP des profils-lignes :

Facteurs principaux u_k = vecteurs propres de la matrice (profils-colonnes X profils-lignes)= composantes principales de l'ACP des profils-colonnes.

Composantes principales : $a_k = Xu_k$ avec X tableau profils-lignes

Contribution d'un point-ligne i à l'axe k : $\frac{n_{i.} (a_{ki})^2}{n \lambda_k}$

Qualité de représentation (ou cosinus carré) d'un point-ligne sur l'axe k : $\frac{(a_{ki})^2}{\sum_{k=1}^q a_{ki}^2}$ (q le nombre d'axes)

LIGNES	COORDONNEES			CONTRIBUTIONS			COSINUS CARRES		
YEUX	1	2	3	1	2	3	1	2	3
y.marron	- 0.49	0.09	-0.02	43,1	13	6.7	0.97	0.03	0.00
y.noisette	- 0.21	-0.17	0.10	3,4	19,8	61.1	0.54	0.34	0.12
y.vert	0.16	-0.34	-0.09	1,4	55,9	31.9	0.18	0.77	0.05
y.bleu	0.55	0.08	0.00	52,1	11,2	0.3	0.98	0.02	0.00

ACP des profils-colonnes :

Facteurs principaux v_k = vecteurs propres de la matrice (profils-lignes X profils-colonnes) = composantes principales de l'ACP des profils-lignes.

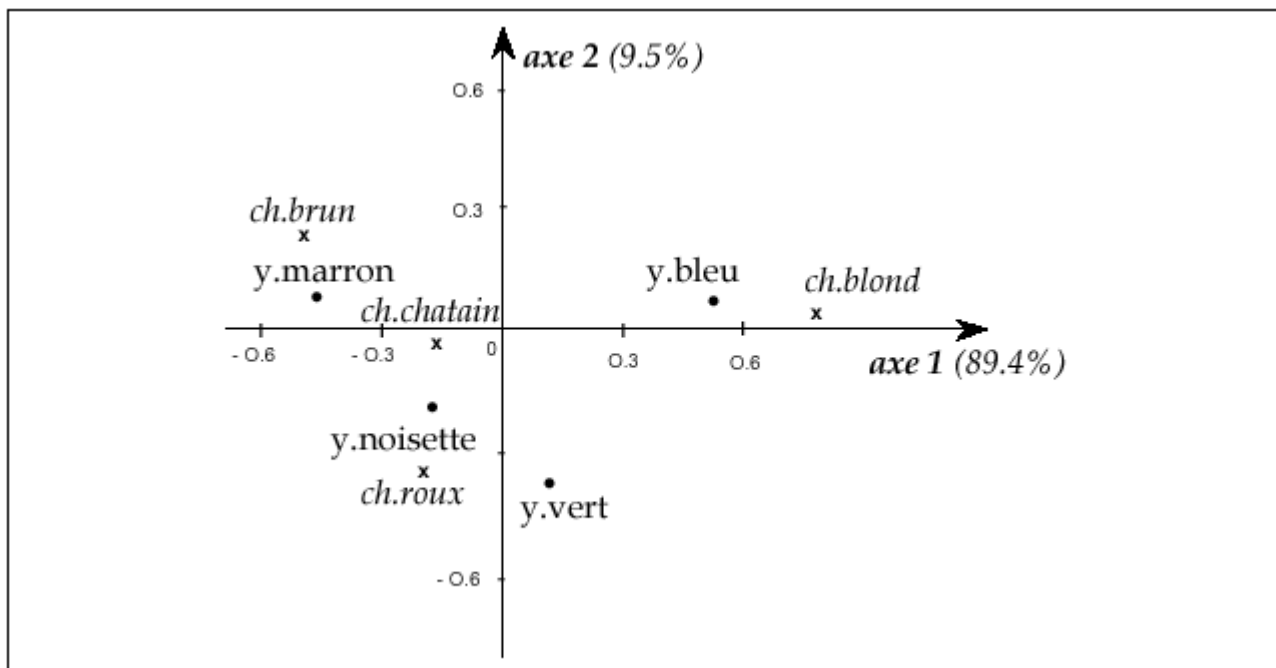
Composantes principales : $b_k = Xv_k$ avec X tableau profils-colonnes

Contribution d'un point-colonne j à l'axe k : $\frac{n_{.j}}{n} \frac{(b_{kj})^2}{\lambda_k}$

Qualité de représentation (ou cosinus carré) d'un point-colonne sur l'axe k : $\frac{(b_{kj})^2}{\sum_{k=1}^q b_{kj}^2}$ (q le nombre d'axes)

COLONNES	COORDONNEES			CONTRIBUTIONS			COSINUS CARRES		
CHEVEUX	1	2	3	1	2	3	1	2	3
Ch.Brun	-0.50	0.21	-0.06	22.2	37,9	21,6	0.84	0.15	0.01
Ch.châtain	-0.15	-0.03	0.05	5.1	2,3	44,3	0.86	0.04	0.09
Ch.roux	-0.13	-0.32	-0.08	1.0	55,1	31,9	0.13	0.81	0.05
Ch.blond	0.84	0.07	-0.02	71.7	4,7	2,2	0.99	0.01	0.00

4- Représentation graphique simultanée des lignes et des colonnes



5- Quelques interprétations :

ACP profils-lignes :

- 1er axe construit presque exclusivement par les yeux "marrons" et "bleus" (contributions de 43.1% et 52.1%).

- Second axe surtout lié aux yeux "verts".

ACP profils-colonnes :

- Sur 1^{er} axe, cheveux "blond" s'oppose à tous les autres. Le point "blond" a une contribution de 71.7% et un cosinus carré de 0.99 : il est pratiquement sur cet axe. Le point "roux" a une contribution très faible sur le premier axe (1.0%).

- Second axe essentiellement construit par la couleur "roux" (55.1 %) qui s'oppose simultanément à "brun" et "blond". La couleur "roux" est le seul point bien représenté sur cet axe (cosinus carré de 0.81).