TD RD5 - AFC

Exercice 1

Dans la DVS de $(X = D_I^{-1}F, Q = D_J^{-1}, D = D_I)$ des profils lignes, nous avons trouvé :

- la matrice d'inertie $X^TDXQ = F^TD_I^{-1}FD_J^{-1}$,
- les vecteurs propres V_L vérifiant $X^TDXQV_L = F^TD_I^{-1}FD_J^{-1}V_L = V_L\Sigma^2$, avec $V_L^TD_J^{-1}V_L = I_r$
- et $F_L = XQV_L = D_I^{-1}FD_J^{-1}V_L$.

1. Retrouver par symétrie pour le triplet $(X = D_J^{-1}F^T, Q = D_I^{-1}, D = D_J)$ des profils colonnes les résultats V^C et F^C : r = le nb de valeurs propres
   - la matrice d'inertie $X^TDXQ = (D_J^{-1}F)^T D_J \,.\, D_J^{-1} F D_I^{-1} = F^T \cancel{D_J^{-1}} \cancel{D_J} D_J^{-1} F D_I$
   - les vecteurs propres V^C vérifiant $X^TDXQV^C = (D_J^{-1}F)^T D_J (D_J^{-1}F) D_I^{-1} V^C = V^C\Sigma^2$, avec $V^{C\,T} D_I^{-1} V^C = I_r$
   - et $F^C = XQV^C = D_J^{-1} F D_I^{-1} V^C$

2. On pose $X = D_I^{-1/2}FD_J^{-1/2}$. $n f_{ij} = n_{ij}$ $n_{i.} = n f_{i.}$
   (a) Montrer que $X = \left(\dfrac{n_{ij}}{\sqrt{n_{i.}n_{.j}}}\right)$

$$X = D_I^{-1/2} F D_J^{-1/2} = \left(\frac{1}{\sqrt{f_{i.}}} f_{ij} \frac{1}{\sqrt{f_{.j}}} \right)_{ij} = \frac{1}{n} \frac{n_{ij}}{\sqrt{\frac{n_{i.}\, n_{.j}}{n^2}}} = \frac{n_{ij}}{\sqrt{n_{i.}\, n_{.j}}}$$

   (b) Réaliser la DVS simple de X. On note V, U et Σ les matrices de cette DVS.

$$X^TX = D_J^{-1/2} F^T D_I^{-1/2} \times D_I^{-1/2} F D_J^{-1/2} = D_J^{-1/2} F^T D_I^{-1} F D_J^{-1/2}$$

$X^TXV = V\Sigma^2$ → Axes factoriels ET $V^TV = I_r$ (car normé).

$\Sigma^2 = \mathrm{diag}(d_i)$

(c) Montrer que l'on retrouve les mêmes valeurs propres que dans les DVS des profils lignes et colonnes.

$$F^T D_I^{-1} F D_J^{-1} V_L = V_L \Sigma^2$$

$$D_J^{-1/2} F^T D_I^{-1} F D_J^{-1/2} V = V \Sigma'^2 \xRightarrow[\times D_J^{1/2}]{} F^T D_I^{-1} F D_J^{-1} D_J^{1/2} V = D_J^{1/2} V \Sigma'^2$$

<u>CCL</u> : Les deux DVS ont les mêmes valeurs propres Σ^2 et $D_J^{1/2} V$ et V_L sont colinéaires (colonne)

(d) Montrer que $V_L = D_J^{1/2} V$.

Est-ce que leurs normes sont égaux (car ils sont déjà colinéaires)

$V_L^T D_J^{-1} V_L = I_r$ et $V^T V = I_r$

$$(D_J^{1/2} V)^T D_J^{-1} (D_J^{1/2} V) = V^T (D_J^{1/2} D_J^{-1} D_J^{1/2}) V = V^T V = I_r$$

(e) En déduire que $F_L = L D_J^{-1/2} V$.

$$F_L = D_I^{-1} F D_J^{-1} V_L = \underbrace{D_I^{-1} F}_{L} \underbrace{D_J^{-1} D_J^{1/2}}_{D_J^{-1/2}} V = L D_J^{-1/2} V$$

(f) En déduire que $F^C = D_J^{-1/2} V \cancel{\Lambda}$. Σ

symétrie des colonnes DVS X^T avec $XX^T U = U\Sigma^2$
Profil colonnes = profil ligne de X^T

$$X^T X V = V \Sigma^2$$

$$X X^T U = U \Sigma^2 \longrightarrow X^T X \underbrace{X^T U} = \underbrace{X^T U} \Sigma^2$$

$X^T U$ et V colinéaire

$$U^T X X^T U = U^T U \Sigma^2 = \Sigma^2$$

$$V = X^T U \Sigma^{-1}$$

$$U = X V \Sigma^{-1}$$

$$F^C = X Q U^C = D_J^{-1} F^T D_I^{-1} U^C \underset{d)}{=} D_J^{-1} F^T D_I^{-1} D_I^{1/2} U = D_J^{-1} F^T D_I^{-1/2} D_I^{-1/2} F D_J^{-1/2} \times V \Sigma^{-1}$$

$$= D_J^{-1/2} \underbrace{D_J^{-1/2} F^T D_I^{-1} F D_J^{-1/2}}_{V \Sigma^2} V \Sigma^{-1} = D_J^{-1/2} V \Sigma$$

Exercice 2

On considère le tableau de contingence

6 effectifs $= n$

$$N = \begin{array}{c} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \end{array}$$

On appellera A, B, C, D les modalités ligne et X, Y, Z celles colonne.

1. Calcul des fréquences et profils

   (a) Calculer le tableau des fréquences relatives, F, et les fréquences marginales F_I et F_J.

   $D_J = \operatorname{diag}(F_J)$.

   $F = \frac{1}{n} N = \frac{1}{6} N$

   $$F = \begin{pmatrix} 1/6 & 0 & 0 \\ 1/6 & 1/6 & 0 \\ 1/6 & 0 & 1/6 \\ 0 & 0 & 1/6 \end{pmatrix}, \quad \underbrace{F_I}_{\text{lignes}} = \begin{pmatrix} 1/6 \\ 2/6 \\ 2/6 \\ 1/6 \end{pmatrix}, \quad \underbrace{F_J}_{\text{colonnes}} = \begin{pmatrix} 3/6 = 1/2 \\ 1/6 \\ 2/6 = 1/3 \end{pmatrix} = \begin{pmatrix} f_{\cdot 1} \\ f_{\cdot 2} \\ f_{\cdot 3} \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/6 \\ 1/3 \end{pmatrix}$$

   (b) les profils ligne L et colonne C.

   de N

   $\frac{\text{ligne}}{\Sigma \text{ ligne}}$ de N

   $$L = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}, \quad C^T = \begin{pmatrix} 1/3 & 0 & 0 \\ 1/3 & 1 & 0 \\ 1/3 & 0 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}$$

   $\frac{\text{colonne}}{\Sigma \text{ colonne}}$ de N.

2. Calculer les distances entre les modalités A, B, C et D de la première variable. Les résultats seront représentés sous forme d'un tableau.

   $d^2_{\chi^2}(A, D) = \frac{(0-1)^2}{1/2} + \frac{(0-0)^2}{1/6} + \frac{(1-0)^2}{2/6} = 2 + 3 = 5.$

   $d^2_{\chi^2}(B, D) = 2(0-\frac{1}{2})^2 + 6(0-\frac{1}{2})^2 + 3(1-0)^2 = 1/2 + 3/2 + 3 = 5.$

d^2	B	C	D
A	2	5/4	5
B		9/4	5
C			5/4

3. Construire la matrice $X = D_I^{-1/2} F D_J^{-1/2}$ puis $X^T X$.

   $D_J = \operatorname{diag}(F_J)$ ← fréquences marginales colonnes

   $$X = \begin{pmatrix} 1/\sqrt{3} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 0 & 1/2 \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix} \quad \underbrace{X^T X}_{\text{symétrique}} = \begin{pmatrix} 2/3 & 1/\sqrt{12} & 1/\sqrt{24} \\ 1/\sqrt{12} & 1/2 & 0 \\ 1/\sqrt{24} & 0 & 3/4 \end{pmatrix}$$

4. Valeurs propres et inertie

(a) Pourquoi une valeur propre sera 1 ? Calculer les deux autres valeurs propres. Que représentent-elles ?

$$\det\left(X^TX - \lambda I_3\right) = \begin{vmatrix} 2/3-\lambda & 1/\sqrt{12} & 1/\sqrt{24} \\ 1/\sqrt{12} & 1/2-\lambda & 0 \\ 1/\sqrt{24} & 0 & 3/4-\lambda \end{vmatrix} = 1/24 \times \left(\frac{1}{2}-\lambda\right) + \left(3/4-\lambda\right)\times\left[\left(\frac{2}{3}-\lambda\right)\left(\frac{1}{2}-\lambda\right) - \frac{1}{12}\right]$$

$$= \frac{1}{24}\left(\frac{1}{2}-\lambda\right)\left[24\times\left(\frac{3}{4}-\lambda\right)\left(\frac{2}{3}-\lambda\right) - 2\right]$$

$\lambda_1 = 2/3$ et $\lambda_2 = 1/4$ ce sont l'inertie projetée selon l'axe.

(b) Quel est l'inertie totale du nuage ? En déduire la statistique du χ^2.

$$I_T = \lambda_1 + \lambda_2 = \frac{11}{12} \qquad D^2_n = n\, I_T = \frac{11}{2} \qquad n = 6.$$

(c) En déduire les % d'inertie projetée sur les différents axes.

$$\tau_1 = \frac{2/3}{11/12} = \frac{8}{11} \qquad \tau_2 = \frac{1}{4}\times\frac{12}{11} = \frac{3}{11}$$

5. Vecteurs propres et composantes principales

(a) Calculer les vecteurs propres normés pour la norme usuelle de X^TX. On donne $v_2 = (\sqrt{6/15}, -\sqrt{8/15}, -\sqrt{1/15})$

$$X^TX\, v_1 = \lambda_1 v_1 \iff \begin{cases} \cancel{2/3\, x} + 1/\sqrt{12}\, y + 1/\sqrt{24}\, z = \cancel{2/3\, x} \\ 1/\sqrt{12}\, x + 1/2\, y = 2/3\, y \\ 1/\sqrt{24}\, x + 3/4\, z = 2/3\, z \end{cases}$$

$$\iff \begin{cases} 1/\sqrt{12}\, y + 1/\sqrt{24}\, z = 0 \\ 1/\sqrt{24}\, x + \frac{1}{12} z = 0 \\ 6/\sqrt{12}\, x = y \end{cases} \iff \begin{cases} 6/12\, x + 1/\sqrt{24}\, z = 0 \\ -12/\sqrt{24}\, x = z \\ \sqrt{3}\, x = y \end{cases} \qquad v_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ \sqrt{3} \\ -\sqrt{6} \end{pmatrix}$$

(b) En déduire les composantes principales F_L. On donne $F_L^2 \approx (0.89, -0.45, 0.22, -0.45)$

$D_S = \mathrm{diag}(F_S)$.

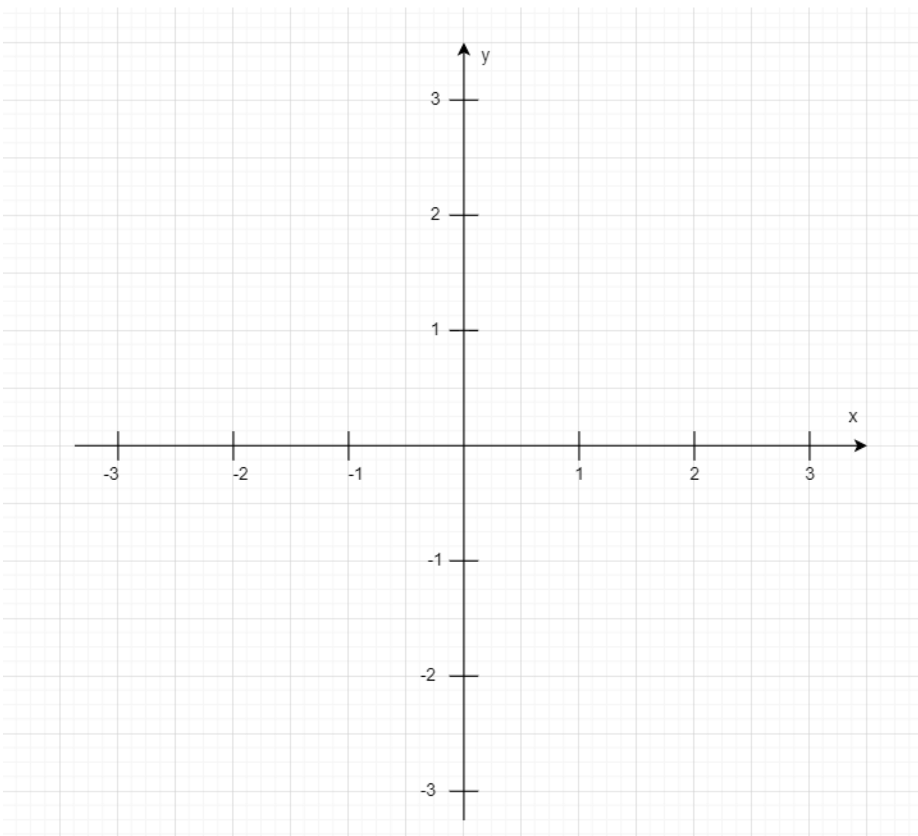
$$F_L^1 = L\, D_S^{-1/2}\, v_1 = \frac{\sqrt{2}}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ -1 \\ -2 \end{pmatrix}$$

$$F_C^1 = D_S^{-1/2}\, v_1 \sqrt{\lambda_1} = \sqrt{\frac{2}{3}}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix}\begin{pmatrix} 1 \\ \sqrt{3} \\ -\sqrt{6} \end{pmatrix} \times \frac{1}{\sqrt{10}}$$

$$= \sqrt{\frac{1}{15}}\begin{pmatrix} \sqrt{2} \\ \sqrt{18} \\ -\sqrt{18} \end{pmatrix} = \sqrt{\frac{2}{15}}\begin{pmatrix} 1 \\ 3 \\ -3 \end{pmatrix}$$

(c) En déduire les composantes principales colonnes F^C. On donne $F^2_C \approx (0.45, -0.89, -0.22)$

(d) Représenter les profils ligne et colonne dans le premier plan factoriel en distinguant ligne et colonne par des couleurs différentes.



6. Retrouver les résultats avec R. On pourra utiliser la commande `CA` de `FactoMineR` sur le tableau N ou reprendre les calculs pas à pas..

Exercice 3 Êtes-vous sympathique ?

Pour le savoir, répondre aux questions suivantes puis en faire une synthèse.

qu'est-ce qu'être sympathique ?

(JM Lasgouttes Paris I)

1 Les données

Il s'agit d'une recherche sur la représentation sociale. Les personnes interrogées appartenaient à 8 catégories professionnelles différentes : paysan (`PAYS`), ouvrier (`OUVR`), vendeur (`VEND`), commerçant (`COMM`), employé (`EMPL`), technicien (`TECH`), universitaire (`UNIV`), profession libérale (`LIBE`).

Elles avaient à choisir les 3 qualités les plus appropriées à une personne sympathique, parmi une liste de 9 : sérieuse (`seri`), généreuse (`gene`), gaie (`gai`), honnête (`honn`), intelligente (`intl`), serviable (`serv`), courageuse (`cour`), compréhensive (`comp`), discrète (`disc`).

Le tableau suivant indique, pour chaque groupe professionnel, le nombre de fois où chaque qualité a été associée à la représentation d'une personne sympathique.

	seri	gene	gai	honn	intl	serv	cour	comp	disc	total	
PAYS	20	9	9	27	10	16	20	4	8	123	÷41
OUVR	42	10	22	51	18	28	38	12	22	243	÷81
VEND	11	2	5	14	8	7	5	8	6	66	
COMM	8	9	12	23	14	16	14	12	12	120	
EMPL	19	10	16	52	32	25	22	25	30	231	÷ 3 = 77
TECH	10	5	12	23	20	13	11	13	10	117	
UNIV	2	8	7	6	15	6	6	9	4	63	
LIBE	8	42	23	24	46	22	22	34	16	237	
TOTAL	120	95	106	220	163	133	138	117	108	1200	

$1200 \div 3 = 400$

Question 1 *Combien de personnes ont été interrogées pour cette enquête ? Quelle est la proportion des employés pour qui être honnête rend sympathique ? Quelle est la proportion d'employés parmi les gens qui pensent qu'être honnête rend sympathique ?* $\frac{52}{220}$ $\frac{52}{77}$

Question 2 *Commentez l'assertion « il est beaucoup plus courant pour un ouvrier que pour un paysan de penser qu'une personne sérieuse est sympathique ». On se restreindra à une seule interprétation.*

$\frac{20}{41} \approx \frac{42}{81}$ donc non

2 Analyse de correspondances

L'analyse des correspondances du tableau de contingence produit les valeurs propres ci-dessous :

```
[1] 0.098 0.022 0.005 0.003 0.001 0.001 0.000
```

2 axe suffisant.

On fournit ci-dessous, pour les profils lignes et les profils colonnes, les poids des modalités (en %) et, sur les 3 premiers axes, les coordonnées des modalités, leurs contributions aux axes (en %) et la qualité de leur représentation par les axes factoriels (en % aussi).

fréquence marginale/relative — coordonées composantes principales — contribution — cos 2

	Poids
PAYS	10.2
OUVR	20.2
VEND	5.5
COMM	10.0
EMPL	19.2
TECH	9.8
UNIV	5.3
LIBE	19.8

	Axis1	Axis2	Axis3
PAYS	-0.303	-0.213	-0.034
OUVR	-0.357	-0.111	0.018
VEND	-0.191	0.120	0.189
COMM	0.015	0.046	-0.105
EMPL	-0.060	0.215	-0.056
TECH	0.015	0.147	0.075
UNIV	0.461	0.001	0.092
LIBE	0.498	-0.114	-0.006

	Axis1	Axis2	Axis3
PAYS	9.6	21.4	2.5
OUVR	26.3	11.6	1.3
VEND	2.1	3.6	40.3
COMM	0.0	1.0	22.6
EMPL	0.7	40.8	12.6
TECH	0.0	9.7	11.4
UNIV	11.4	0.0	9.1
LIBE	49.9	11.8	0.2

	Axis1	Axis2
PAYS	63.5	31.5
OUVR	89.6	8.7
VEND	35.9	14.1
COMM	1.0	9.9
EMPL	6.6	83.3
TECH	0.6	58.4
UNIV	90.4	0.0
LIBE	94.4	5.0

% % %

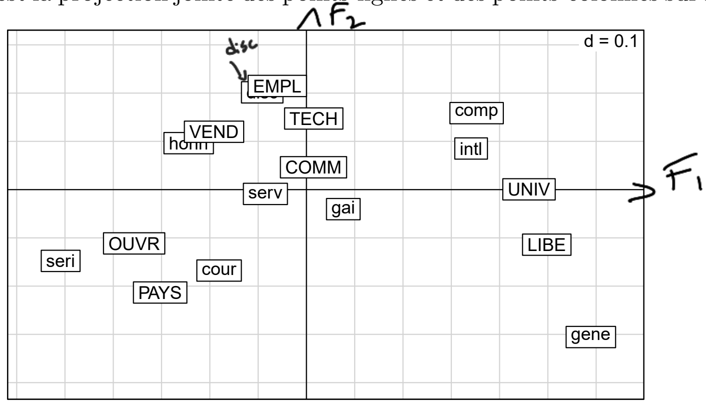
	Poids
seri	10.0
gene	7.9
gai	8.8
honn	18.3
intl	13.6
serv	11.1
cour	11.5
comp	9.8
disc	9.0

	Comp1	Comp2	Comp3
seri	-0.509	-0.148	0.133
gene	0.589	-0.306	-0.063
gai	0.076	-0.040	0.036
honn	-0.244	0.096	-0.039
intl	0.341	0.085	0.070
serv	-0.085	-0.009	-0.046
cour	-0.181	-0.168	-0.056
comp	0.352	0.158	0.062
disc	-0.092	0.202	-0.091

	Axis1	Axis2	Axis3
seri	26.4	10.0	36.2
gene	28.0	34.0	6.6
gai	0.5	0.7	2.3
honn	11.1	7.8	5.7
intl	16.1	4.5	13.8
serv	0.8	0.0	4.9
cour	3.8	14.9	7.4
comp	12.4	11.3	7.6
disc	0.8	16.8	15.4

	Axis1	Axis2
seri	85.7	7.2
gene	77.0	20.7
gai	25.4	7.1
honn	82.6	12.9
intl	87.0	5.4
serv	51.6	0.6
cour	48.1	41.4
comp	78.8	15.9
disc	13.1	63.4

Le diagramme ci-dessous est la projection jointe des points-lignes et des points-colonnes sur le premier plan factoriel.



Question 3 *Pourquoi y a-t-il 7 valeurs propres ?* min(n,p) - 1 = min(8,9) - 1 = 8-1=7

Question 4 *Quelles sont les modalités qui définissent le premier axe factoriel ? Et le deuxième ? On précisera sur quel(s) critère(s) on se fonde.*

Question 5 *Quelles sont les modalités (lignes et colonnes) qui sont particulièrement mal représentées par le premier plan factoriel ?* F1 + F2

donc gai et comm

Question 6 *Que peut-on déduire du fait que* OUVR *et* PAYS *sont proches sur le graphique ? Même question pour* VEND *et* honn*.* 1) Ils ont des profils d'une personne très proches. 2) La plupart des vendeurs répond souvent honnête.

Axe 1 :

	−	+
	ouvr.	libe
	seri	gene
		intel

Axe 2 :

	−	+
	pays	empl.
	gene	disc