Solution de l'exercie N°3 de la Série N°2 AFC

Exercice 2. Suite de l'exercice 1, et 2.

1-Déterminer les composantes principales des profils-lignes et déduire celles des profils-colonnes.

2-Déduire les coordonnées des lignes de X_r et X_c par rapport aux axes principaux.

3-Quelle sont les valeurs de l'espérance et la variance de chaque composante principale pour les deux profils? Quelles sont les valeurs les covariances entre les composantes principales? Que peut-on déduire?

4-Déterminer les contributions absolues et relatives, de chaque ligne, aux inerties des axes principaux.

5.Dans le même plan définit, par les deux premiers axes principaux, représenter les deux nuages de point associés à X_r et X_c.
6-Analyser et discuter les résultats obtenus.

* **

Solution

Rappel: la matrice des effectifs observés est

$$N^* = \begin{pmatrix} 50 & 280 & 120 & 20 \\ 8 & 29 & 210 & 350 \\ 150 & 230 & 100 & 40 \end{pmatrix}.$$

La matrice des fréquences observées est

$$N = \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix}.$$

Le centre de gravité des profils-lignes:

$$g_r = (0.131\,06, 0.339\,63, 0.270\,95, 0.258\,35)^t.$$

Matrice diagonale des profils-linges

$$D_r = \begin{pmatrix} 0.296\,16 & 0 & 0 \\ 0 & 0.376\,18 & 0 \\ 0 & 0 & 0.327\,66 \end{pmatrix}.$$

Matrice diagonale des profils-colonnes

$$D_c = \begin{pmatrix} 0.131\,06 & 0 & 0 & 0 \\ 0 & 0.339\,63 & 0 & 0 \\ 0 & 0 & 0.270\,95 & 0 \\ 0 & 0 & 0 & 0.258\,35 \end{pmatrix}.$$

Le centre de gravité des profils-colonnes:

$$g_c = (0.296\,16, 0.376\,18, 0.327\,66)^t\,.$$

Matrice des profils-linges

$$X_r = \begin{pmatrix} 0.106\,38 & 0.595\,74 & 0.255\,32 & 4.\,255\,3\times10^{-2} \\ 0.013\,4 & 4.\,857\,6\times10^{-2} & 0.351\,76 & 0.586\,27 \\ 0.288\,46 & 0.442\,31 & 0.192\,31 & 7.\,692\,4\times10^{-2} \end{pmatrix}.$$

Les valeurs propres de $V_r D_c^{-1}$ sont

$$\lambda_1 = 0.479\,66, \lambda_2 = 5.\,310\,9\times10^{-2}, \lambda_3 = 0, \lambda_4 = 0,$$

et les vecteurs directeurs des axes principaux associés sont

$$u_1 = \begin{pmatrix} 0.132\,46 \\ 0.325\,89 \\ -9.\,079\,1\times10^{-2} \\ -0.367\,58 \end{pmatrix},\ u_2 = \begin{pmatrix} 0.309\,57 \\ -0.267\,64 \\ -0.106\,4 \\ 6.\,445\,7\times10^{-2} \end{pmatrix},$$

$$u_3 = \begin{pmatrix} 2.\,242\,0\times10^{-2} \\ -0.215\,51 \\ 0.421\,87 \\ -0.228\,77 \end{pmatrix},\ u_4 = \begin{pmatrix} 0.131\,06 \\ 0.339\,63 \\ 0.270\,95 \\ 0.258\,35 \end{pmatrix}$$

****

1) Nous allons traités les questions que pour les profils-colonnes, car celles des profils-lignes se traitent de la même façon. Les composantes principales des profils-colonnes sont définies par:

$$c_k = X_r w_k,\ k = 1,2,3,4.$$

où $w_k = D_c^{-1} u_k$ sont les facteurs principaux. Donc

$$\begin{aligned} c_1 &= \begin{pmatrix} 0.106\,38 & 0.595\,74 & 0.255\,32 & 4.\,255\,3\times10^{-2} \\ 0.013\,4 & 4.\,857\,6\times10^{-2} & 0.351\,76 & 0.586\,27 \\ 0.288\,46 & 0.442\,31 & 0.192\,31 & 7.\,692\,4\times10^{-2} \end{pmatrix} \\ &\times \begin{pmatrix} 0.131\,06 & 0 & 0 & 0 \\ 0 & 0.339\,63 & 0 & 0 \\ 0 & 0 & 0.270\,95 & 0 \\ 0 & 0 & 0 & 0.258\,35 \end{pmatrix}^{-1} \begin{pmatrix} 0.132\,46 \\ 0.325\,89 \\ -9.\,079\,1\times10^{-2} \\ -0.367\,58 \end{pmatrix} \\ &= \begin{pmatrix} 0.533\,06 \\ -0.891\,86 \\ 0.542\,07 \end{pmatrix}, \end{aligned}$$

et

$$c_2 = \begin{pmatrix} 0.106\,38 & 0.595\,74 & 0.255\,32 & 4.\,255\,3\times10^{-2} \\ 0.013\,4 & 4.\,857\,6\times10^{-2} & 0.351\,76 & 0.586\,27 \\ 0.288\,46 & 0.442\,31 & 0.192\,31 & 7.\,692\,4\times10^{-2} \end{pmatrix}$$

$$\times \begin{pmatrix} 0.131\,06 & 0 & 0 & 0 \\ 0 & 0.339\,63 & 0 & 0 \\ 0 & 0 & 0.270\,95 & 0 \\ 0 & 0 & 0 & 0.258\,35 \end{pmatrix}^{-1} \begin{pmatrix} 0.309\,57 \\ -0.267\,64 \\ -0.106\,4 \\ 6.\,445\,7\times10^{-2} \end{pmatrix}$$

$$= \begin{pmatrix} -0.307\,83 \\ 1.\,509\,8\times10^{-3} \\ 0.276\,47 \end{pmatrix}.$$

Les deux dernières composantes principales c_3 et c_4 sont des vecteurs nuls car ces valeurs propres associées sont nulles. En résumé nous avons

$$c_1 = \begin{pmatrix} 0.533\,06 \\ -0.891\,86 \\ 0.542\,07 \end{pmatrix},\ c_2 = \begin{pmatrix} -0.307\,83 \\ 1.\,509\,8\times10^{-3} \\ 0.276\,47 \end{pmatrix},$$

$$c_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},\ c_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

2) La matrice des profils-lignes X_r dans la nouvelle base $\{u_{1,}u_2, u_3, u_4\}$ devient

$$\mathbf{X}_r = \begin{pmatrix} 0.533\,06 & -0.307\,83 & 0 & 0 \\ -0.891\,86 & 1.\,509\,8\times10^{-3} & 0 & 0 \\ 0.542\,07 & 0.276\,47 & 0 & 0 \end{pmatrix}.$$

Les lignes de cette matrice représentent les coordonnées des modalités du vecteur ligne dans la nouvelle base.

3) Théoriquement la moyenne de chaque composante principale est nulle et les variances égalent aux valeurs propres associées. Mais dû aux erreurs d'arrondies des calculs, vous allez trouvez des résultats approximatives.

4) Les contributions absolues et relatives, de chaque ligne, aux inerties des axes principaux sont définies respectivement par:

$$\mathbf{ca}_k(i) = c_k^2(i)\, f_{i\cdot} \text{ et } \mathbf{cr}_k(i) = \frac{c_k^2(i)}{\lambda_k} f_{i\cdot}.$$

Donc la contribution absolue et la contribution relative de la première ligne par rapport au premier axe principal sont:

$$\mathbf{ca}_1(1) = c_1^2(1)\, f_{1\cdot} = (0.533\,06)^2\, 0.296 = 8.\,410\,9\times10^{-2}$$

et

$$\mathbf{cr}_1(1) = \frac{c_1^2(1)}{\lambda_1} f_{1\cdot} = \frac{(0.533\,06)^2}{0.479\,66} 0.296 = 0.175\,35 \simeq 17\%.$$

La contribution absolue et la contribution relative de la première ligne par rapport au deuxième axe principal sont:

$$\mathbf{ca}_2(1) = c_2^2(1)\, f_{1\cdot} = (-0.307\,83)^2\, 0.296 = 2.\,804\,9\times10^{-2}$$

et

$$\mathbf{cr}_2(1) = \frac{c_2^2(1)}{\lambda_2} f_{1\cdot} = \frac{(-0.307\,83)^2}{5.310\,9 \times 10^{-2}} 0.296 = 0.528\,14 \simeq 53\%.$$

Ainsi on peut calculer les contributions pour le reste des lignes par rapport aux axes principaux. Je souligne que les contributions absolues et relatives par rapport aux axes principaux associées aux valeurs propres nulles sont nulles.

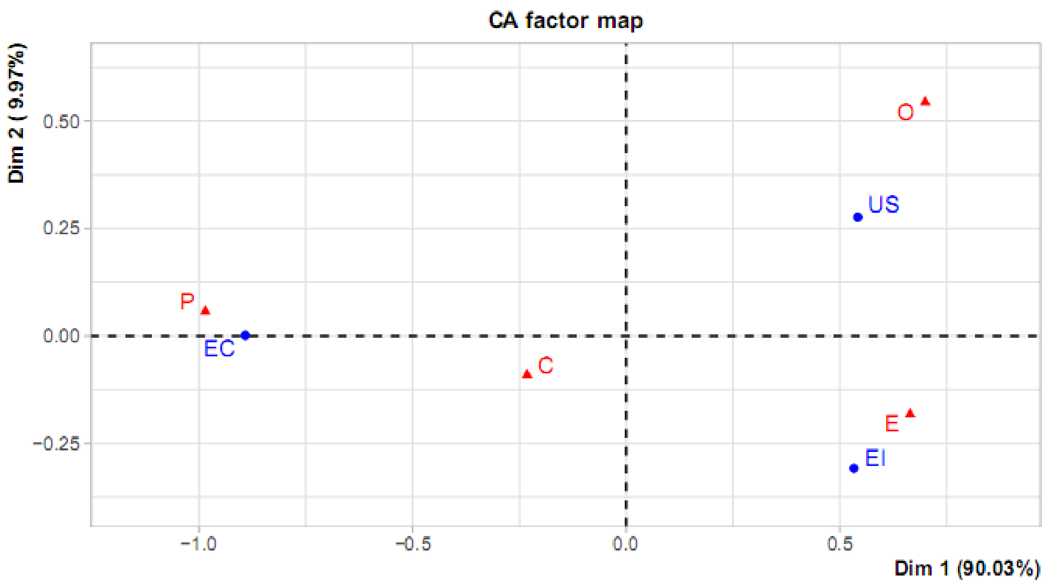
5) Pour simplifier les notations, nous allons renommer les modalités comme suit:

Écoles d'ingénieurs ⟶	EI
Écoles de commerce ⟶	EC
Universités Scientifiques ⟶	US

et

Ouvriers ⟶	O
Employés ⟶	E
Cadres ⟶	C
Professions libérales ⟶	P

La représentation des deux nuages de points de X_r et X_c sur le même plan principal associé aux deux premiers axes principaux est donnée par la figure suivante:



6) On remarque que les étudiants de Écoles de commerce (EC) leurs pères exerces des professions libérales (P). Les étudiants d'Universités Scientifiques leurs pères sont des ouvriers. Les étudiants des Écoles d'ingénieurs leurs pères sont des Employés. On peut dire que les étudiants ayant des pères qui sont des cadres ne figurent pas, en général, parmi les trois filières de formation (Écoles de commerce, Universités Scientifiques, Écoles d'ingénieurs).

```
################################################
Voici le code R réalisant cette dernière figure:
##########################################
df <- data.frame(O= c(50,8,150), E=c(280,29,230), C= c(120,210,100),
P=c(20,350,40),
row.names = c("EI","EC","US"))
library(FactoMineR)
library(factoextra)
res=CA(df)
#################################################
```