Faculté de Maths | USTHB | 24/05/17

Master MIND — Examen : Analyse des Données

Durée : 90 Mns

Important : Chaque réponse doit porter le numéro de la question correspondante.

Exercice 1 (06 points): Soit X, le tableau de dimension (6,2) suivant :

$$ {}^{t}X = \begin{pmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{pmatrix} $$

Effectuer une ACP normée permettant d'expliquer une inertie d'au moins 80% de l'inertie totale, en justifiant chaque étape.

Exercice 2 (08 points): Interpréter l'ACP effectuée sur les données présentées dans le document distribué.

Exercice 3 (06 points) : On mesure deux variables quantitatives V (1ère variable) et W (2ème variable) sur une population de taille n. On suppose que les variables sont centrées réduites et de coefficient de corrélation r >0. Montrer que la composante principale d'ordre k peut s'écrire sous la forme :

$$ \Psi_k = \pm \frac{1}{\sqrt{2}}(V + (-1)^{k+1} W) $$

N.B : La copie doit être soigneusement présentée et les réponses aux questions doivent être claires et précises.

Corrig de l'examen Anal-Donnes 2016-2017
MIND M1-S2

Exercice 1. L'ACP à effectuer étant normée, nous devons normer les variables qui possèdent des écart-types égaux à $\sqrt{4/6}$=$\sqrt{2/3}$. Le tableau étant centré, car $g = \binom{0}{0}$. Donc le tableau centré réduit est le suivant:

$$^tX = \sqrt{3/2}\begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{bmatrix}$$

1) La matrice à diagonaliser dans ce cas est la matrice des corrélations $R =^t XD_pX = \frac{1}{6}{}^tXX =$

$$\frac{1}{6}\times\frac{3}{2}\begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & -1 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

2) La matrice $R = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$ possède les éléments propres $\lambda_1 = \frac{3}{2}$, $u_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ et $\lambda_1 = \frac{1}{2}$, $u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

3) Le taux d'inertie expliqué par les deux axes principaux ensemble est 100%.On doit prendre les deux axes car la part d'inertie expliquée par le premier axe uniquement est 75%.

4) Corrélations des variables avec les composantes principales:

On sait que le coefficient de corrélation entre une variable X^j et une composante principale ψ^k est donné par $r(X^j, \psi^k) = \sqrt{\lambda_1}u_k^j$. D'où:

$r(X^1, \psi^1) = \sqrt{\lambda_1}u_1^1 = \sqrt{\frac{3}{2}} \times \frac{-1}{\sqrt{2}} = -\frac{1}{2}\sqrt{3}$

$r(X^2, \psi^1) = \sqrt{\lambda_1}u_1^2 = \sqrt{\frac{3}{2}} \times \frac{1}{\sqrt{2}} = \frac{1}{2}\sqrt{3}$

$r(X^1, \psi^2) = \sqrt{\lambda_2}u_2^1 = \sqrt{\frac{1}{2}} \times \frac{1}{\sqrt{2}} = \frac{1}{2}$

$r(X^2, \psi^2) = \sqrt{\lambda_1}u_2^2 = \sqrt{\frac{1}{2}} \times \frac{1}{\sqrt{2}} = \frac{1}{2}$

5) Calcul des composantes principales:

a) La première composante principale ψ^1 est donnée par la relation

$$\psi^1 = Xu_1 = \sqrt{\frac{3}{2}} \times \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & -1 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} \\ -\sqrt{3} \\ \sqrt{3} \\ -\frac{1}{2}\sqrt{3} \end{bmatrix}$$

b) La deuxième composante principale ψ^2 est donnée par la relation

$$\psi^2 = Xu_2 = \sqrt{\frac{3}{2}} \times \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & -1 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} \\ 0 \\ 0 \\ -\frac{1}{2}\sqrt{3} \end{bmatrix}.$$

Il reste à représenter les points dans le plan.

Exercice 3. Il suffit d'écrire le tableau $X = (V, W)$.Dans ce cas la matrice à diagonaliser est $R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$ (matrice des corrélations) car les données sont centrées réduites.

Les valeurs et vecteurs propres sont $\lambda_1 = 1 + r, u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ et $\lambda_2 = 1 - r, u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

D'où

$\psi^1 = (V, W) \times \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{-1}{\sqrt{2}}(V - W)$ et $\psi^2 = (V, W) \times \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}(V + W)$. D'où le résultat.

Exercice 2. L'interprétation de l'ACP sur l'étude dont les individus sont les fromages.

1) Le tableau est de type $Ind \times Var$. Les variables sont toutes quantitatives et par conséquent on peut appliquer une ACP sur les données.

2) On remarque que les variances sont significativement différentes et donc on doit réduire les données et cela permet d'effectuer une ACP normées.

3) D'aprés le tableau qui représente la matrice des corrélations, nous avons les variables H2s, goût et Acla significativement et positivement corrélées entre elles. Ce constat nous permet d'affirmer l'existence de deux axes principaux.

4) Le tableau 3 confirme notre constat en examinant le pourcentage d'inertie expliquée par le premier plan (85.52%).

5) Le tableau 4 indique une forte corrélation entre les quatres variables et l'axe 1. De plus cette corrélation est de même signe et cela permet de conclure que le premier axe est un axe d'échelle (Il ordonne les fromages selon leurs caractéristiques).

6) L'examen du tableau 6 permet de suggérer 5 classes de fromages.

Un premier groupe se concentre autour du centre de gravité (2,3,6,7,10,18,19,23 et 29). Ce groupe peut être qualifié de groupe de fromages qui ont des valeurs pour l'ensemble des variables les moyennes de l'ensemble des variables.

Ailleurs, on peut distinguer les quatre groupes suivants:

-(4,11,15 et 24): ce groupe est caractérisé par les fortes valeurs des variables qui se trouvent du même côté de l'axe.

-(8,12,16 et 27): ce groupe est caractérisé par la variable Acac.

-(26,28 et 30): ce groupe possède de faibles valeurs pour les variables hormis Acac.

-(1,5,9,13,17 et 21): ce groupe est caractérisé par les faibles valeurs de l'ensemble des variables.