Université Hassiba Benbouali Chlef
Faculté des Sciences exactes et informatique
Département de Mathématiques

Année universitaire 2021/2022
Master I: Math.Appli & Stat
Module : Analyse de données

Hom work 01–Corrigé

Exercice 01:

On considere le tableau de données, noté X, qui est defini par : $X = \begin{pmatrix} 2 & 1 & -1 & 3 & -3 \\ 0 & 3 & -1 & 2 & -2 \\ 1 & 2 & -1 & 1 & -1 \end{pmatrix}$

ou la i^{eme} ligne designe l'individu x_i et la j^{eme} colonne designe la variable x^j. Chaque individu possede un poids égal à 1/3. On considere les résultats de l'ACP du tableau X lorsque R^5 est muni de la metrique identité $M = I$.

1. Determiner le tableau centré Y.et le tableau centré réduit Z ?
   $Y = \begin{pmatrix} 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$, Z n'existe pas.

2. Sans calculer la matrice variance V du tableau X, combien existe-t-il d'axes factoriels ?
   Le rang de Y est 2 puisque la colonne 1 est l'oppose de la colonne 2, la colonne 4 est l'oppose de la colonne 5 et la colonne 3 est nulle. Enfin la premiére et la cinquième colonne ne sont pas colineaires. Donc le rang de Y est 2 donc celui de V aussi. Il y a donc deux axes factoriels non triviaux.

3. Calculer V. En deduire l'inertie totale du nuage étudié.
   $V = \frac{1}{3}Y'Y = \frac{1}{3}\begin{pmatrix} 2 & -2 & 0 & 1 & -1 \\ -2 & 2 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 2 & -2 \\ -1 & 1 & 0 & -2 & 2 \end{pmatrix}$ L'inertie totale du nuage etudié est la trace de V soit 8/3.

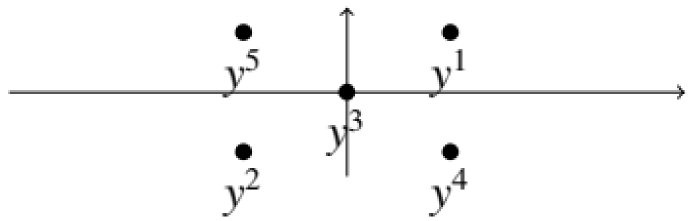
4. soit $U^t = (1 \ \ -1 \ 0 \ 1 \ \ -1)^t$
   (a) Verifier que U est un vecteur propre de V associe à la valeur propre 2.
   On a Vu = 2u. ‘
   (b) En deduire les valeurs propres et vecteurs propres de V.
   Puisqu'il n'y a que deux axes factoriels non triviaux, il y a trois valeurs propres nulle, 2 et une derniere valeur propre. Or la trace vaut 8/3 donc la derniere valeur propre est 2/3.

5. Calculer les deux premieres composantes principales , notée Ψ_1 et Ψ_2 .
   $\Psi_1 = Yu_1 = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}$ *et* $\Psi_2 = Yu_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$

6. Representer les trois individus dans le plan factoriel constitué des deux premiers axes.

7. Calculer la contribution relative de chaque individu a l'inertie du premier axe.
   La contribution de l'individu 1 a l'axe 1 est 2/3, a l'axe 2 est 0.
   La contribution de l'individu 2 a l'axe 1 est 1/6, a l'axe 2 est 1/2.
   La contribution de l'individu 3 a l'axe 1 est 1/6, a l'axe 2 est 1/2.

8. Representer les 5 variables dans le plan factoriel constitué des deux premiers axes.



Exercice 02:

On considere le tableau de données, noté X, et defini par : $X = \left(\begin{array}{c|cccccc} & j_1 & j_2 & j_3 & j_4 & j_5 & j_6 \\ \hline i_1 & 1 & 1 & 2 & 0 & 3 & 1 \\ i_2 & 0 & 1 & 2 & 1 & 1 & 3 \\ i_3 & 1 & 2 & 4 & 1 & 4 & 4 \end{array}\right)$

ou la i^{eme} ligne désigne la variable x_i et la j^{eme} colonne désigne l'individu x^j. Par la suite, on considere les résultats de l'ACP sur matrice variance du tableau X.

1. Calculer les coordonnees du centre de gravité g du nuage $\mathcal{M}$ constitue des vecteurs colonnes de X (munis du meme poids 1/6), et en déduire le tableau Y centre qui est associe a X. On presentera Y sous la forme $Y = 1/3.Y_1$ où Y_1 est une matrice a coefficients ' entiers.

   $g' = (4/3 \;\; 4/3 \;\; 8/3)'$. D'ou' $Y = \frac{1}{3}\begin{pmatrix} -1 & -1 & 2 & -4 & 5 & -1 \\ -4 & -1 & 2 & -1 & -1 & 5 \\ -5 & -2 & 4 & -5 & 4 & 4 \end{pmatrix}$

2. Soit V la matrice variance du tableau X. Completer les valeurs manquantes dans l'expression de la matrice V ci-dessous :

   $V = \frac{1}{18}\begin{pmatrix} 16 & 1 & 17 \\ 1 & 16 & 17 \\ 17 & 17 & 34 \end{pmatrix}$

3. Expliquer pourquoi le nombre d'axes factoriels non triviaux est égal a 2.
   La troisieme ligne de Y est la somme des deux premieres, donc le rang de Y est au plus 2. Par ailleurs, les lignes n'etant pas toutes proportionnelles, ce rang n'est pas égal à 1. Donc le rang de Y, qui est aussi celui de V, vaut 2. Donc le nombre d'axes factoriels non triviaux vaut 2.

4. Calculer l'inertie totale du nuage etudié.
   $I_T = tr(VM) = tr(V) = \frac{16+16+34}{18} = \frac{66}{18} \simeq 3.6667.$

5. Montrer que $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ est un vecteur directeur d'un axe factoriel.

$$V \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \frac{1}{18} \begin{pmatrix} 51 \\ 51 \\ 102 \end{pmatrix} = \frac{17}{6} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

6. Calculer le pourcentage d'inertie expliquee par l'axe factoriel déterminé a la question 5. Cet axe est-il le premier ou le second axe factoriel ?

   D'apres ce qui précéde, nous avons le pourcentage $\tau = \frac{17/6}{11/3} = \frac{17}{22} = 0.7727(77.27\%)$
   Comme $\tau > 0.5(50\%)$, il s'agit du premier axe factoriel.

7. Determiner les coordonnées du premier vecteur axial factoriel, (on choisira sa premiere coordonnée de façon a ce qu'elle soit positive).

$$u_1 = \frac{1}{\sqrt{1^2+1^2+2^2}} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

8. Calculer la premiere composante principale de l'individu $j2$, notée Ψ_1^{j2}.

$$\Psi_1^{j2} = (y^{j2})' u_1 = \frac{1}{3\sqrt{6}}(-1 \ -1 \ -2) \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = -\frac{\sqrt{6}}{3} = -\sqrt{\frac{2}{3}}.$$

9. Calculer la contribution de l'individu j^2 a l'inertie du premier axe, notée $CTR_1(j^2)$.

$$CTR(i,\alpha) = \frac{w_i(\Psi_i^\alpha)^2}{\lambda_\alpha} \Longrightarrow CTR_1(j2) = \frac{1}{6}\frac{(\Psi_1^{j2})^2}{\lambda_1} = \frac{1}{6}\frac{6}{17}\frac{6}{17} = \frac{2}{51} = 0.0392 = 3.92\%.$$

10. Calculer la qualite de représentation de l'individu j^2 sur le premier axe, notée $QLT_1(j^2)$.

    - Avec la matrice $Z \Longrightarrow QLT_1(\alpha) = \frac{(\Psi_i^\alpha)^2}{\|Z_i\|^2}$
    - Avec la matrice $Y \Longrightarrow QLT_1(\alpha) = \frac{(\Psi_i^\alpha)^2}{\|y_i\|^2}$

$$QLT_1(j2) = \frac{\Psi_1^{j2})^2}{\|y_{j2}\|^2} = \frac{2/3}{(-1/3)^2 + (-1/3)^2 + (-2/3)^2} = \frac{2/3}{6/9} = 1$$

11. Calculer la contribution de la variable i_1 a l'inertie du premier axe, notée $CTR_1(i_1)$.

$$CTR_1(i_1) = \frac{(\phi_i^1)^2}{\lambda_1} = \frac{\sqrt{(\lambda_1 u_1^1)^2}}{\lambda_1} = (u_1^1)^2 = 0.1667$$

Exercice 03:

On considére le tableau X de données suivant : $X = \begin{pmatrix} 1 & 2 & 0 & -1 \\ 1 & 0 & -2 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$

On rappelle que le terme général de X, noté x_i^j, indique la valeur prise par la i eme variable (avec i$\in$ 1,2,3) pour le j eme individu (avec j$\in$ 1,2,3,4). Dans ce qui suit, on examine les résultats de l'ACP sur matrice variancex du tableau X.

1. Déterminer la valeur de c telle que l'on ait $x_1^j + x_2^j + cx_3^j = 0$ pour tout j $\in$ 1,2,3,4.
   On vérifie que $x_1^j + x_2^j - 2x_3^j = 0$ pour tout j : par exemple, pour j = 1, on a 1 + 1 - 2 = 0. Donc $c = -2$.

2. Determiner le tableau centré Y. et le tableau centré réduit Z ?
   Soit g le c.d.g. du nuage des individus. On a $g = \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix}$ Le tableau Y des données centrées est donc $Y = \begin{pmatrix} 1/2 & 3/2 & -1/2 & -3/2 \\ 3/2 & 1/2 & -3/2 & -1/2 \\ 1 & 1 & -1 & -1 \end{pmatrix}$.

3. Calculer la matrice variance V du tableau X.
   Il en résulte que $V = \frac{1}{4}Y'Y = \begin{pmatrix} 5 & 3 & 4 \\ 3 & 5 & 4 \\ 4 & 4 & 4 \end{pmatrix}$

4. Calculer la matrice de corrélation R du tableau X.

5. Montrer que le vecteur $v = (1, 1, 1)^t$ dirige un axe factoriel non trivial du nuage des individus, associé a une valeur propre λ de matrice variance V dont on précisera la valeur. Expliquer pourquoi cet axe est le premier axe factoriel.
   On a $V.v = \frac{12}{4}v = 3v$. Donc v dirige un axe factoriel non trivial associé à la valeur propre $\lambda = 3$. Par ailleurs, on sait que la somme des valeurs propres est égale à la trace de V , donc à $\frac{14}{4} = 3 + \frac{1}{2}$. Comme il y a au plus 3 axes factoriels car le nombre p de variables est égal à 3, et que deux d'entre-elles sont égales à 3 et 0, il existe un seul autre axe factoriel non trivial qui est associé à la valeur prore 1/2. On en conclut que l'axe dirigé par v, associé à la valeur propre 3, est bien le premier axe factoriel du nuage des individus.

6. Calculer un vecteur directeur du second axe factoriel du nuage des individus et déterminer l'inertie projetée sur cet axe.
   Soit $w = (x \;\; y \;\; z)'$ un vecteur directeur de cet axe. D'après ce qui précède, $w \perp v; \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$.
   D'où $x+y+z = 0$ et $x+y-2z = 0$. On en déduit $z = -x-y$ et $3x+3y = 0$. Finalement, $y = -x$ et $z = 0$. Donc pour w, on prend $w = (1 \; -1 \; 0)$, et on vérifie que $V.w = \frac{1}{2}w$, puisque l'on sait que la seconde valeur propre vaut 1/2. L'inertie projetée sur le second axe factoriel non trivial est donc égale à 1/2.

7. Soit Ψ_1 la première composante principale du nuage des individus. Calculer Ψ_j^1 pour tout j $\in$ 1,2,3,4.
   On a $\|v\|^2 = 3$. D'où $\Psi_1 = Y'\frac{v}{\|v\|} = \frac{1}{\sqrt{3}}Y'.v = \frac{1}{\sqrt{3}}(3 \; 3 \; -3 \; -3)' = \sqrt{3}(1 \; 1 \; -1 \; -1)'$.

8. Calculer les coordonnées des trois variables sur le premier axe factoriel de l'espace des variables.

Les coordonnées des trois variables sur le premier axe sont les coordonnées du vecteur $\phi = \sqrt{\lambda}\frac{v}{\|v\|} = v = (1 \;\; 1 \;\; 1)'$.

9. Calculer la qualité de représentation de chaque individu sur le premier axe factoriel.

   Pour tout j, on a $QLT_1(j) = \frac{(\Psi_1^j))^2}{\|y_j\|^2} = \dfrac{3}{(1/4)+(9/4)+(1)} = \dfrac{6}{7}$

   Pour tout j, on a $CTR_1(j) = \frac{(1/4)(\Psi_1^j))^2}{\lambda} = \dfrac{3/4}{3} = \dfrac{1}{4}$