ANALYSE DES DONNEES (Partiel)

Master M1 MMD, 14 mars 2014

Calculatrice autorisée, documents autorisés : 2 feuilles recto-verso.

Barême approximatif : exercice 1 (10 pts) ; exercice 2 (10 pts).

Les exercices 1 et 2 de ce sujet peuvent être traités de façon indépendante.

Exercice 1

On considère un tableau X de données, comportant p lignes et n colonnes, et de terme général noté x_i^j avec $i \in I = \{1, \ldots, p\}$ et $j \in J = \{1, \ldots, n\}$. Pour tout $k \in \mathbb{N} \setminus \{0\}$, on note $\mathbb{1}_k$ le vecteur de $\mathbb{R}^k$ dont les coordonnées sont toutes égales à 1. On suppose que la condition suivante, notée (1), est vérifiée :

(1) Il existe une constante c telle que pour tout $j \in J$, on a $x_1^j + \ldots + x_p^j = c$.

1 - Soit $\mathcal{M}$ le nuage des n points x^j (j ème vecteur colonne de X) avec $j \in J$, chaque point x^j étant muni du poids $\dfrac{1}{n}$. On note g_X le centre de gravité de $\mathcal{M}$. Pour quelle valeur de la constante γ a-t'on $g_X = \gamma\, X\, \mathbb{1}_n$?

2 - Soit Y le tableau obtenu après avoir centré X. On rappelle l'égalité $Y = X - g_X \mathbb{1}_n'$. En utilisant cette égalité, et en notant V_X la matrice variance relative au nuage $\mathcal{M}$, montrer que $V_X = \dfrac{1}{n} XX' - g_X g_X'$.

3 - Montrer que (1) est équivalent à l'existence d'une constante c telle que $X' \mathbb{1}_p = c\, \mathbb{1}_n$.

4 - Montrer que le vecteur $\mathbb{1}_p$ dirige un axe factoriel de l'ACP sur matrice variance du tableau X. Préciser la valeur de l'inertie du nuage projeté sur cet axe.

5 - On pose $Z = Y'$. Soit $\mathcal{N}$ le nuage des p points y_i (i ème vecteur colonne de Y') avec $i \in I$, chaque point y_i étant muni du poids $\dfrac{1}{p}$. On note g_Z le centre de gravité du nuage $\mathcal{N}$. Montrer que $g_Z = 0$.

6 - Considérons les deux ACP sur matrice variance précédentes, c.-à-d. celle du tableau X et celle du tableau Z. Expliquer pourquoi ces deux ACP admettent le même nombre d'axes factoriels non triviaux.

7 - Soit u^α un vecteur axial factoriel (non trivial) de l'ACP sur matrice variance du tableau X. On note λ_α la valeur propre associée à u^α. Montrer que $Y' u^\alpha$ dirige un axe factoriel (non trivial) de l'ACP sur matrice variance du tableau Z. Préciser la valeur de l'inertie du nuage $\mathcal{N}$ projeté sur cet axe factoriel.

8 - Soit F_α (resp. G_α) la α ème composante principale de l'ACP sur matrice variance du tableau X (resp. Z). Exprimer G_α en fonction de n, λ_α, Y et F_α.

Exercice 2

On considère le tableau de données, noté X, et défini par :

$$X = \begin{array}{|c|cccccc|}\hline & j_1 & j_2 & j_3 & j_4 & j_5 & j_6 \\ \hline i_1 & 1 & 1 & 2 & 0 & 3 & 1 \\ i_2 & 0 & 1 & 2 & 1 & 1 & 3 \\ i_3 & 1 & 2 & 4 & 1 & 4 & 4 \\ \hline \end{array}$$

où la i ème ligne désigne la variable x_i et la j ème colonne désigne l'individu x^j.
Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X.

1 - Calculer les coordonnées du centre de gravité g du nuage $\mathcal{M}$ constitué des vecteurs colonnes de X (munis du même poids $1/6$), et en déduire le tableau Y centré qui est associé à X. On présentera Y sous la forme $Y = \frac{1}{3}Y_1$ où Y_1 est une matrice à coefficients entiers.

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

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

3 - Expliquer pourquoi le nombre d'axes factoriels non triviaux est égal à 2.

4 - Calculer l'inertie totale du nuage étudié.

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

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

7 - Déterminer les coordonnées du premier vecteur axial factoriel, noté u^1 (on choisira sa première coordonnée de façon à ce qu'elle soit positive).

8 - Calculer la première composante principale de l'individu j_2, notée $\Psi_1^{j_2}$.

9 - Calculer la contribution de l'individu j_2 à l'inertie du premier axe, notée $CTR_1(j_2)$.

10 - Calculer la qualité de représentation de l'individu j_2 sur le premier axe, notée $COR_1(j_2)$.

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