

Université de Carthage
Ecole Supérieure de la Statistique et de l'Analyse de l'Information

Devoir surveillé d'Analyse des données

1 ère année du cycle de formation d'ingénieurs

Durée de l'épreuve : 1 heure 30 - Documents non autorisés

Nombre de pages : 2 - Date de l'épreuve : 13 avril 2021

Exercice 1 : On considère le tableau de données, noté X , qui est défini par :

$$X = \begin{array}{c|ccc} & j_1 & j_2 & j_3 \\ \hline i_1 & 3 & 2 & 0 \\ i_2 & 2 & 0 & 5 \\ i_3 & 5 & 0 & 2 \\ i_4 & 0 & 2 & 3 \\ i_5 & 2 & 1 & 3 \\ i_6 & 0 & 1 & 5 \end{array}$$

où la $j^{\text{ème}}$ colonne désigne la variable x^j et la $i^{\text{ème}}$ ligne désigne l'individu x_i . Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X (i.e., ACP de X lorsque \mathbb{R}^3 est muni de la métrique identité et chaque individu possède un poids égal à $1/6$).

2 pts 1) Calculer les coordonnées du centre de gravité g du nuage des individus, et en déduire le tableau Y centré qui est associé à X .

2 pts 2) Soit V la matrice variance du tableau X . Montrer que

$$V = \frac{1}{3} \begin{pmatrix} 9 & -2 & -5 \\ -2 & 2 & -2 \\ -5 & -2 & 9 \end{pmatrix}$$

3 pts 3) Montrer que le vecteur $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ est un vecteur propre de V . Que peut-on en déduire pour la représentation du nuage des individus ?

2 pts 4) Sachant que les coordonnées des individus sur le premier axe sont données par : $\Psi_{11} = \Psi_{31} = 2\sqrt{2}$, $\Psi_{21} = \Psi_{41} = -\sqrt{2}$, $\Psi_{51} = 0$ et $\Psi_{61} = -2\sqrt{2}$, en déduire λ_1 et λ_2 , les deux valeurs propres associées aux deux premiers axes factoriels.

2 pts 5) Calculer les pourcentages d'inertie expliqués par chacun des axes factoriel.

2 pts 6) Evaluer la qualité de représentation de l'individu i_2 sur chacun des deux premiers axes.

2 pts 7) Calculer la covariance entre la première variable et la première composante et en déduire la qualité de représentation de la première variable sur l'axe 1.

Exercice 2 : On a testé des chaises hautes pour enfants. Les chaises hautes ont été évaluées sur 6 caractéristiques (sécurité, commodité, ceinture, plateau, nettoyage et rangement) selon une échelle de 1 à 5 (1=excellent, ..., 5=médiocre). Les résultats moyens obtenus sont donnés dans le tableau suivant :

Marque	sécurité	commodité	ceinture	rangement	nettoyage	plateau
M1	2.0	2.4	1.7	2.3	3.3	3.8
M2	3.0	4.0	4.4	4.9	3.9	3.8
M3	1.5	3.4	3.8	4.7	2.3	2.6
M4	1.2	3.6	3.9	4.3	1.3	1.3
M5	1.9	1.6	3.4	4.0	1.2	1.1
M6	3.5	4.8	4.5	4.6	3.9	4.1
M7	3.3	4.1	3.9	4.9	3.6	3.9

A partir du tableau ci-dessus, on a réalisé une Analyse en Composantes Principales normée.

2 pts 1) Expliquer l'intérêt de cette ACP.

2 pts 2) Indiquer la démarche à suivre lors de l'interprétation des résultats d'une ACP.

2 pts 3) Sachant que les trois premières valeurs propres de cette ACP normée sont respectivement 3.64, 1.90 et 0.32, déterminer le nombre d'axes à retenir en vous basant sur trois critères différents.