Corrigé d'Examen : Analyse de données

Exercice 01 : (03 pts)

1. Que signifié l'analyse de données ?
   L'Analyse des Données recouvre les techniques ayant pour objectif la description statistique et graphique des grands tableaux (n lignes, p colonnes,). Elles insistent sur les représentations graphiques en particulier de celles des individus qui sont considérés au même titre que les variables.

2. Rappeler briévement ce qu'est un Tableau disjonctif complet, puis préciser la dimension associée.
   Le TDC est un tableau binaire croisant les modalités des variables actives avec les unités(individus). Chaque cellule est constituée de la valeur 1 ou 0 selon que l'individus a choisi ou non la modalité correspondante. Si on dénombre un total de n individus actives sur lesquelles on a observé p variables actives correspondant à un total de m modalités actives, le TDC est de dimension $n \times m$.

3. Citer un exemple dans lequel la technique de l'AFCM est particulièrement bien adaptée.
   L'AFCM est adaptée à l'étude d'un tableau de données comportant plus de 2 variables qualitatives , c'est une technique complémentaire de l'ACP qui traite uniquement des variables quantitatives et de l'AFC qui traite uniquement les tableaux de contingence , *elle est particulièrement adaptée à l'étude des résultats d'une enquête.*

4. Est ce que dans le cadre de l'AFCM , L'inertie total des nuages dépend uniquement du nombre de variables et de modalités actives? Justifier?
   Oui dans l'AFCM , L'inertie total des nuages dépend uniquement du nombre de variables et de modalités actives puisqu' elle est égale $(\frac{\text{total modalités}}{\text{total variables}} - 1) = \frac{m}{p} - 1)$

Exercice 02 : (06 pts)

Soit 3 individus prenant pour les variables ξ_1, ξ_2 et ξ_3 les valeurs respectives suivantes
Individu 1: $(-2, 3, -1)$ Individu 2: $(-1, 1, 0)$ Individu 3: $(2, -1, -1)$. Individu 4: $(1, -3, 2)$.
Pour l'étude , on effectuera une ACP centrée avec des poids équirépartis et la métrique identité.

1. Déterminer les données centrés que l'on notera Y .

$$X = \begin{pmatrix} -2 & 3 & -1 \\ -1 & 1 & 0 \\ 2 & -1 & -1 \\ 1 & -3 & 2 \end{pmatrix} \Longrightarrow Y = X - I_g = \begin{pmatrix} -2 & 3 & -1 \\ -1 & 1 & 0 \\ 2 & -1 & -1 \\ 1 & -3 & 2 \end{pmatrix}$$

2. Calculer V la matrice de variance-covariance .

$$V = \frac{1}{n}{}^tY.Y = \frac{1}{4}\begin{pmatrix} 10 & -12 & 2 \\ -12 & 20 & -8 \\ 2 & -8 & 6 \end{pmatrix} = \begin{pmatrix} 5/2 & -3 & 1/2 \\ -3 & 5 & -2 \\ 1/2 & -2 & 3/2 \end{pmatrix}$$

3. Vérifier que V admet une valeur propre $\lambda_3 = 0$.
   Pour cela montrons que $det(V - \lambda_3 I_3) = 0 \Longrightarrow det(V) = 0 \Longrightarrow \lambda_3 = 0$ valeur propre de V .

4. Le tableau suivant fournit les valeurs propres ainsi qu'un certain nombre d'indicateurs liés à l'ACP. Compléter le tableau. On a

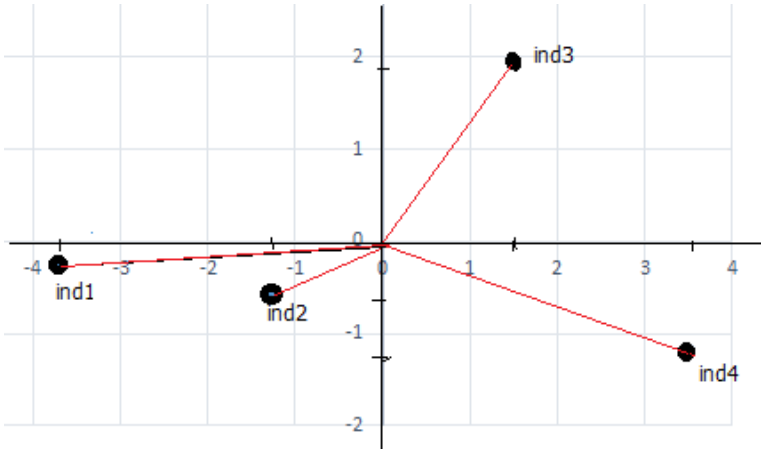
$$\begin{cases} \sum \lambda_\alpha = tr(V) \Longrightarrow \lambda_1 + \lambda_2 = 9 \\ \frac{\lambda_2}{\sum \lambda_\alpha = 9} \times 100 = 15.22 \Longrightarrow \lambda_2 \simeq 1.37 = \frac{5.5}{4} \Longrightarrow \lambda_1 = 7.63 = \frac{30.5}{4} \end{cases}$$

λ_α		%Inertie	%Inertie cumulée
7.63		**84.78%**	**84.78%**
1.37		15.22%	**100%**
0		**0**	**100%**

5. On donne, les vecteurs propres associés aux valeurs propres, $u_1 = \begin{pmatrix} 0.5 & -0.8 & 0.3 \end{pmatrix}^t$, $u_2 = \begin{pmatrix} 0.65 & 0.11 & -0.75 \end{pmatrix}^t$ Calculer les composantes principales.

$$\Psi = Y.U = \begin{pmatrix} \psi^1 & \psi^2 \\ \hline \frac{-37}{10} & \frac{-11}{50} \\ \frac{-13}{10} & \frac{-27}{50} \\ \frac{3}{2} & \frac{97}{50} \\ \frac{7}{2} & \frac{-59}{50} \end{pmatrix} = \begin{pmatrix} \psi^1 & \psi^2 \\ \hline -3.7 & -0.22 \\ -1.3 & -0.54 \\ 1.5 & 1.94 \\ 3.5 & -1.18 \end{pmatrix}$$

6. Représenter les individus dans le plan principal.



Exercice 03 : (11 pts)

On a relevé sur $n = 10$ individus deux variables qualitatives, la variable X à 4 modalités $\{A, B, C, D\}$ et la variable Y à trois modalités $\{\alpha, \beta, \gamma\}$. Les résultats sont regroupés dans le tableau suivant qui donne sous forme d'une $(\star)$, les modalités relevées sur un individu.

Ind		A	B	C	D		α	β	γ
1		$\star$					$\star$		
2		$\star$							$\star$
3				$\star$				$\star$	
4				$\star$					$\star$
5			$\star$				$\star$		
6		$\star$					$\star$		
7		$\star$						$\star$	
8				$\star$				$\star$	
9					$\star$				$\star$
10				$\star$			$\star$		

I : Réaliser une (AFC) sur ces données, en répondant aux questions suivantes :

1. Pourquoi en peut utiliser l'AFC pour traiter ce tableau des données ?
   L'ACP traite uniquement des variables quantitatives et l'AFC traite uniquement les tableaux de contingence, i.e. deux variables qualitatives simultanément est on a le cas deux variables qualitatives et 4 et 3 modalités réspectivement .
2. Donner la table de contingence associée.

$H =$

	α	β	γ		$n_{i.}$
A	2	1	1		4
B	1	0	0		1
C	1	2	1		4
D	0	0	1		1
$n_{.j}$	4	3	3		**10**

3. Construire le tableau des profils-lignes noté H_1 et profils-colonnes noté H_2 ?
   - profils-lignes noté H_1 $H_1 = \frac{f_{ij}}{f_{i.}} = \frac{n_{ij}}{n_{i.}} \Longrightarrow H_1 =$

	α	β	γ		$(H_1)_{i.}$
A	1/2	1/4	1/4		1
B	1	0	0		1
C	1/4	1/2	1/4		1
D	0	0	1		1
$(H_1)_{.j}$	7/4	3/4	6/4		**4**

   - profils-colonnes noté H_2

$H_2 = \frac{f_{ij}}{f_{.j}} = \frac{n_{ij}}{n_{.j}} \Longrightarrow H_2 =$

	α	β	γ		$(H_2)_{i.}$
A	1/2	1/3	1/3		7/6
B	1/4	0	0		1/4
C	1/4	2/3	1/3		15/12
D	0	0	1/3		1/3
$(H_2)_{i.}$	1	1	1		**3**

4. Calculer $H_1'.H_2$?

$$\begin{pmatrix} \frac{1}{2} & 1 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 1 \end{pmatrix} \times \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{9}{16} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{5}{12} & \frac{1}{4} \\ \frac{3}{16} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

5. Dans quel espace sera représenté le nuage des profils-lignes et des profils-colonnes ?
   Respectivement dans les espaces R^3 et R^4.

6. L'AFC de ce tableau de contingence et la diagonalisation de la atrice a diagonalisé S permet d'obtenir les résultats suivants : $\lambda_0 = 1$, $\lambda_1 = \frac{5}{16}$, $\lambda_2 = \frac{1}{6}$.
   Que pouvez-vous dire sur la valeur propre λ_0 , En conséquence, sur quelles valeurs propres allez-vous concentrer votre analyse ?

   $\lambda_0 = 1$ est un valeur propre triviale , on concentre l'analyse sur les deux autres.

7. Calculer l'inertie total des profils lignes ?

$$I_T = (\lambda_1 + \lambda_2) = \frac{46}{96}$$

8. Calculer la distance du χ_2 entre les profils-ligne (modalité A et de modalité B).

$$d^2(i,i') = \sum_{j=1}^{p} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \Longrightarrow d^2(A,B) = \Big(\ - \ \Big)^2 + \Big(\ - \ \Big)^2 =$$

II: Réaliser une (AFCM) sur ces données, en répondant aux questions suivantes :

1. Quelle est la défférence entre l' AFC et AFCM ?
   Le but de l'Analyse Factorielle des Correspondances Multiple (AFCM), comme celui de l'AFC, est de détecter des liens entre variables qualitatives, et de positionner les individus par rapport à ces liens. La méthode AFC vue dans le cours ne traite formellement que du lien entre 2 variables qualitatives. La méthode AFCM permet de généraliser cette étude à autant de variables qualitatives que l'on souhaite.

2. Pourquoi en peut utilises AFCM pour traiter ce tableau des données ?

   L'ACM est une technique complémentaire de l'ACP qui traite uniquement des variables quantitatives et de lŠAFC qui traite uniquement les tableaux de contingence, L'ACM est adaptée à l'étude d'un tableau de données comportant plus de 2 variables qualitatives avec des individus est c'est le cas .

3. Diterminer le tableau disjonctif complet ?

Ind		A	B	C	D		α	β	γ
1		1	0	0	0		1	0	0
2		1	0	0	0		0	0	1
3		0	0	1	0		0	1	0
4		0	0	1	0		0	0	1
5		0	1	0	0		1	0	0
6		1	0	0	0		1	0	0
7		1	0	0	0		0	1	0
8		0	0	1	0		0	1	0
9		0	0	0	1		0	0	1
10		0	0	1	0		1	0	0

4. Calculer l'inertie totale des deux nuages en présence. ?

$$I_T = \frac{m}{p} - 1 = \frac{7}{2} - 1 = \frac{5}{7}$$

Fin