TD2 : Analyse d'une table de contingence

Exercice 1 Des élections professionnelles ont lieu dans une entreprise. Il y a 5 candidats, et les électeurs sont répartis en 3 catégories. Les résultats par catégories sont donnés dans la table T ci-contre.

	Albert	Berthe	Charles	Djoe	Elsa
Chefs	120	32	8	32	8
Grands-chefs	48	13	13	13	13
Super-chefs	6	1	1	1	1

Question 1.1 Quel est le nombre total N d'électeurs ? Quels pourcentages de voix ont obtenus les 5 candidats ? Quel pourcentage d'électeurs a-t-on dans chaque chaque catégorie ?

Question 1.2 D'après les pourcentages calculés à la question précédente, combien de chefs auraient voté pour Albert si la répartition des votes étaient indépendante de la catégorie ?

Question 1.3 Calculer la matrice F des fréquences ($F = T/N$). Comparer à la matrice des fréquences qu'on aurait si la répartition des votes étaient indépendante de la catégorie : quelles cases du tableau concentrent le plus d'écart à l'indépendance ?

On s'intéresse maintenant aux « individus » que sont les catégories d'électeurs. La matrice des *profils-lignes* L est obtenue en divisant chaque ligne de T par son effectif total.

Question 1.4 Calculer L. Quels catégories d'électeurs vous semblent avoir les votes les plus similaires ?

Question 1.5 Démontrer qu'on peut aussi obtenir ces profils-lignes à partir de la matrice F des fréquences.

Question 1.6 Calculer les distances euclidiennes entre le profil-ligne « Super-chefs » et les deux autres ? Comment expliquer ce résultat ?

Question 1.7 Pour donner des importances similaires à toutes les colonnes, calculer une matrice L_s obtenue en divisant chaque colonne de L par la racine du total de la colonne correspondante de F (ce qui revient à utiliser une distance du χ^2) : $L_{s\,ij} = L_{ij}/\sqrt{F_{\bullet j}}$ où $F_{\bullet j} = \sum_i F_{ij}$.

Question 1.8 Calculer à nouveau les distances euclidienne entre le profil-ligne « Super-chefs » et les deux autres, cette fois-ci en utilisant L_s. Qu'est-ce qui a changé ?

On a « de bonnes raisons » de diagonaliser non pas directement la matrice $L_s'L_s$, mais la matrice $X'X$, où X est obtenue en multipliant chaque ligne de L_s par la racine de la somme de la ligne correspondante dans F : $X_{ij} = L_{s\,ij}\sqrt{F_{i\bullet}}$, où $F_{i\bullet} = \sum_j F_{ij}$. Cela revient à affecter chaque profil-ligne L_i d'une masse $F_{i\bullet}$.

Question 1.9 Exprimer X_{ij} en fonction de F_{ij}, $F_{i\bullet}$ et $F_{j\bullet}$.

Pour les candidats, on peut raisonner comme on l'a fait pour les catégories, en partant de la transposée de T.

Question 1.10 Exprimer, en fonction de F, une matrice C dont les lignes sont les profils-colonnes de T, et une matrice C_s similaire à L_s.

Question 1.11 De manière similaire à ce qu'on a vu en question 1.9, quelle matrice va-t-on diagonaliser pour l'étude du nuage des candidats ?

Question 1.12 Démontrer que si U est vecteur propre de $X'X$, alors XU est vecteur propre de XX'. Exprimer $U'X'XU$ en fonction de U. Si U est de norme 1, quelle est la norme de XU ? Comment peut-on définir à partir de U un vecteur propre de XX' de norme 1 ?

Question 1.13 Compléter les tableaux suivants, où L_p et C_p sont les projections des profils-lignes L_s et des profils-colonnes C_s sur les axes factoriels, et où E_{val} est le vecteur des valeurs propres de $X'X$ et E_{vec} la matrice dont les colonnes sont les vecteurs propres :

	Terme général (fonction de $F_{ij}, F_{i\bullet}, F_{\bullet j}$)	Expression R (fonction de F)
L_s		
C_s		
X		

	Expression R (fonction de X, L_s, C_s, E_{vec} et de E_{val})
L_p	
C_p	

Question 1.14 Vérifier les propriétés suivantes :

- Selon chaque axe, la moyenne des coordonnées des profils-lignes (respectivement profils-colonnes) pondérée par les masses $F_{i\bullet}$ (respectivement $F_{\bullet j}$), est nulle.
- Selon chaque axe, la moyenne des carrés des coordonnées des profils-lignes (respectivement profils-colonnes) pondérée par les masses, est égale à la valeur propre correspondante.