

DEVOIR ÉCRIT N°2 – MAI 2010

Sans calculatrice, ni document.

Le sujet comporte 5 pages dont 4 d'annexes à rendre avec la copie.

EXERCICE N°1 : question de cours

On rappelle quelques notations utilisées en ACM.

- n : nombre d'individus ou d'unités statistiques
- q : nombre de variables
- m : nombre total de modalités
- $X_j(i)$: valeur de la variable indicatrice X_j pour l'individu i .
- n_i : effectif de la modalité i

1. Rappeler l'expression de l'inertie totale de l'AFC du tableau disjonctif complet.
2. T^* désignant la matrice à diagonaliser lors de l'AFC du tableau disjonctif complet, montrer que :

$$T^* = \left[\frac{1}{q\sqrt{n_i n_j}} \sum_{k=1}^n X_i(k) X_j(k) \right]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq m}}$$

3. Démontrer la formule énoncée dans la question 1.

EXERCICE N°2 :

Le tableau figurant en annexe donne les résultats concernant des variables de différents types pour certains pays du monde en 2007.

1. Expliquer pourquoi on ne peut pas faire une ACP de ces données. Quel est le nom de la méthode d'analyse des données adéquate dans ce cas ?
2. Construire un tableau de Burt pertinent pour rendre compte de ces données. Quelle est l'inertie totale de ces données ? Vous utiliserez le tableau fourni en annexe. Attention, sa dimension n'est peut-être pas correcte !!

EXERCICE N°3 :

On désire effectuer une analyse des résultats du premier tour des élections présidentielles de 2002 en France métropolitaine, par région.

Les données nécessaires sont sur l'annexe.

Certains regroupements de candidats idéologiquement proches ont été effectués.

1. Quel est le nom du type d'analyse de données que nous allons effectuer ? Pourquoi ?
2. Va-t-on utiliser les profil-lignes ou les profil-colonnes ? Pourquoi ?
3. Identifiez les tableaux de profil-lignes et de profil-colonnes en expliquant votre démarche et complétez-les.
4. Faire une analyse rapide des profil-lignes et des profil-colonnes.
5. Quel est le nombre de valeurs propres de la matrice à diagonaliser ?
6. Calculer l'inertie totale ; en déduire la valeur du χ^2 de ce tableau de données. Faites en une interprétation. On admet que l'effectif total est de 39 100 220. On se contentera d'une extrapolation à partir de la table fournie en annexe.
7. Quels sont les axes à conserver ? Pourquoi ?
8. Quelle est la signification de « cr » ? Quelle est son utilité ?
9. Après avoir sélectionné et en expliquant votre choix des régions et des candidats intéressants, faire une tentative d'analyse de ces données.