

RATTRAPAGE D'ANALYSE DES DONNEES II

N.B. Aucun document n'est autorisé

Durée 1 h

Les questions sont indépendantes, elles peuvent être traitées dans n'importe quel ordre.

A Choix multiples

1. Un tableau de Burt est dans ses dimensions :
 - a) indépendant du nombre de variables
 - b) indépendant du nombre des modalités de chaque variable
 - c) égal au nombre de modalités moins une unité
 - d) dépendant du nombre de modalités pour chaque variable.
2. Les analyses de classification sont cohérentes avec le cadre méthodologique suivant :
 - a. regrouper des variables suivant leurs covariances
 - b. substituer des groupes de variables à une variable synthétique
 - c. combiner les éléments afin de minimiser les variances intraclasses
 - d. grouper des observations dans des modalités pré-établies
3. Un dendrogramme permet :
 - a. d'éliminer les variables à faible dispersion
 - b. d'apprécier des proximités entre variables
 - c. de rechercher les valeurs propres doubles
 - d. de rechercher des proximités entre observations

B Répondre de façon concise

4. L'Analyse Factorielle Discriminante (AFD) est une Analyse en Composantes Principales (ACP) particulière. Expliquer ?
5. Le problème de l'AFD revient à maximiser :
$$\frac{\mu'VB\mu}{\mu'W\mu}$$
ou à maximiser de façon équivalente :
$$\frac{\mu'VB\mu}{\mu'V\mu}$$
 - a) Quelles matrices représentent B, V et W ? Donner leurs expressions respectives ? Rappeler la relation entre ces trois matrices ?
 - b) Résoudre ce problème de l'AFD ? En déduire que cela revient à faire une ACP sur des points à préciser avec une métrique bien connue dont on rappellera le nom ?
6. En présence d'un grand nombre d'individus ($>10^3$), est-il possible d'utiliser directement les méthodes de classification hiérarchique. Quel est l'algorithme idéal de classification dans ce cas de données volumineuses. Le décrire brièvement.