

*Master 1 Economie (Toutes spécialités)***DEVOIR D'ANALYSE DE DONNEES**

Toutes les étapes d'un calcul doivent figurer sur la copie. Les copies doivent être rendues au plus tard le **Samedi 9 mars 2024 à 12 heures**. Les groupes doivent être composés de 10 étudiants au maximum. Les noms des membres des groupes doivent être classés par ordre alphabétique. La spécialité et le numéro du groupe doivent obligatoirement apparaître sur la page de présentation de la copie. Les copies sur lesquelles les noms ne sont pas classés par ordre alphabétique ne seront pas corrigées, et la note correspondante est évidemment 00/20.

Exercice 1

NB : Pour chaque question, la syntaxe de Stata, le résultat fourni par cette syntaxe, et éventuellement la réponse précise à la question posée doivent être présentés. Le nom du fichier résultat relatif à cet exercice doit être de la forme « M1eco_groupe(#) ».

La base de données « m1eco.dta » est une base au format STATA, qui contient des informations sur les régions françaises (sauf la Corse) et des caractéristiques démographiques et économiques relatives à ces régions.

Légende:

POPUL: Population de la région, en milliers d'individus

TACT: Taux d'activité (population active/population totale de la région)

SUPERF: Superficie de la région, en kilomètres carrés

NBENTR: Nombre d'entreprises dans la région

NBBREV: Nombre de brevets déposés au cours de l'année

CHOM: Taux de chômage, en pourcentage

TELEPH: Nombre de lignes téléphoniques en place dans la région, en milliers.

- 1) Réaliser une ACP (analyse en composantes principales) sur l'ensemble des variables de la base, à l'aide du logiciel STATA, et interpréter les résultats.
- 2) Réaliser une ACP sur l'ensemble des variables, en permutant les positions de la première et de la deuxième variable de la base.
- 3) Analyser les différences dans les résultats des questions 1 et 2. Commenter.
- 4) Donner la syntaxe de STATA pour présenter dans un tableau, après réalisation de l'ACP, les statistiques descriptives (moyenne, écart-type, minimum, maximum) des variables utilisées. Le nombre d'observations ne doit pas apparaître dans le tableau.
- 5) Exécuter la commande « screeplot » après réalisation de l'ACP, et interpréter le résultat obtenu.

Exercice 2

On dispose d'une variable V_1 ayant 3 modalités A, B, C, et d'une variable qualitative V_2 ayant comme modalités 1, 2 et 3. Les informations relatives à ces 2 variables sont collectées sur un échantillon de 50 individus, et le croisement donne le tableau de contingence ci-dessous. Arrondir tous les résultats au 100^{ème} près.

	1	2	3	Somme
A	8	3	12	23
B	7	9	x	
C	1	y	5	
Somme	16			

x et y sont non nulles et liées par la relation $4xy - y^2 = 0$

V_1 est la tranche d'âge et V_2 est le niveau de diplôme le plus élevé du répondant. Les modalités sont : A : 15-24 ans ; B : 25-44 ans ; C : 45 ans et plus.

1 : Inférieur au Bac ; 2 : Bac-Bac+3 ; 3 : Master et plus.

- 1) Calculer les profils-lignes, présenter les résultats dans un tableau, et les interpréter.
- 2) Calculer les profils-colonnes, présenter les résultats dans un tableau, et les interpréter.
- 3) Calculer les contributions au χ^2 .
- 4) Effectuer le test d'indépendance de V_1 et V_2 . On donne $\chi^2_{5\%}(4) = 9,48$;
 $\chi^2_{5\%}(6) = 12,59$; $\chi^2_{5\%}(8) = 15,51$.

Exercice 3

- 1) On considère le tableau de données X suivant.
$$X = \begin{pmatrix} -1 & 4 & -1 \\ -1 & 2 & -1 \\ 3 & 4 & -3 \\ 3 & 2 & -3 \end{pmatrix}$$

- a) Préciser l'espace des individus et l'espace des variables.
 - b) Calculer la matrice R des corrélations.
- 2) Donner l'analogie d'une part, et la différence d'autre part, entre une variable discriminante et une composante principale.
 - 3) Démontrer que la diagonale principale d'une matrice des corrélations est un vecteur unitaire.
 - 4) Quelle est l'utilité du test d'indépendance du Chi-deux (χ^2) dans la mise en œuvre de l'analyse factorielle des correspondances ?

BONNE CHANCE