



**Epreuve : ANALYSE DES DONNEES**

**Date : LE 26 DECEMBRE 2002**

**Durée : 1h30**

**Directives pédagogiques :**

- Deux pages de documents sont autorisées.
  - Il sera tenu compte de la clarté des raisonnements.
- Toute tentative de fraude sera sanctionnée par la note zéro.

**Lisez attentivement les recommandations données à la fin de l'épreuve.**

**Nom :** .....

**Prénom :** .....

**PARTIE 1 : Statistique descriptive**

**Exercice 1 :** Une compagnie de téléphone a fait enquête auprès de 12 ménages pour déterminer combien d'appareils téléphoniques il y avait par ménage.

**a)** Remplissez dans le tableau suivant les fréquences des résultats de l'enquête suivants : 2, 5, 4, 3, 4, 3, 1, 3, 3, 2, 3, 4

Nombre d'appareils téléphoniques	1	2	3	4	5
Fréquence (f)					

**b)** Quel est le résultat le plus fréquent ? .....

**Exercice 2 :** Un vent a soufflé pendant 40 jours. Voici les vitesses enregistrées :

15, 22, 14, 12, 21, 34, 19, 11, 13, 0, 16, 4, 23, 8, 12, 18, 24, 17, 14, 3, 10,  
12, 9, 15, 20, 5, 19, 13, 17, 11, 16, 19, 24, 12, 7, 14, 17, 10, 14, 23

**a)** S'agit-il de variables discrètes ou continues? .....

**b)** Choisissez un intervalle de classe approprié et présentez ces données à l'intérieur d'un tableau de distribution de fréquences.

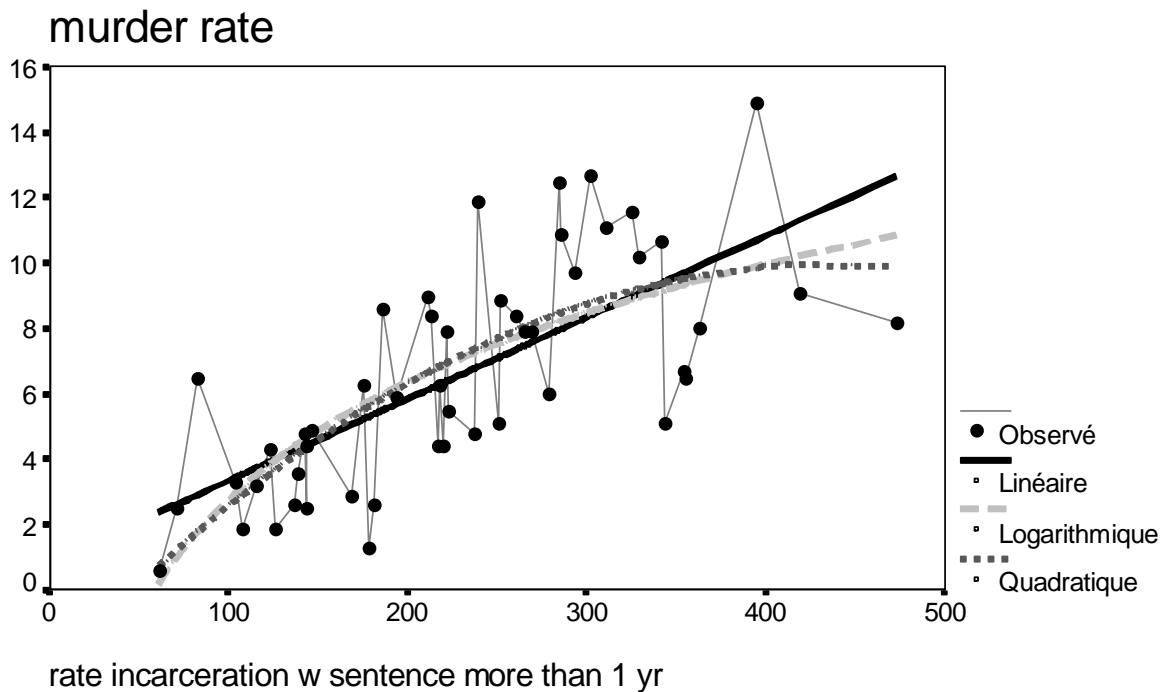
**c)** Quel est l'intervalle de classe le plus fréquent? .....

**d)** Créez un tableau de distribution de fréquences incluant des colonnes pour la fréquence relative et la fréquence en pourcentage des données.

**e)** Quelles conclusions peut-on tirer des tableaux ? Expliquez.

## PARTIE 2 : Fonction d'ajustement d'un nuage de point

Nous illustrons dans la figure suivante la relation entre la proportion de meurtres aux Etats-Unis (**murder**) et la proportion d'individus incarcérés (**incarate**).



Plusieurs fonctions d'ajustement sont proposées. Laquelle est la plus appropriée ? justifiez

.....

.....

.....

D'autre part, il est toujours utile de regarder les corrélations entre les variables explicatives et expliquées du modèle.

	INCARATE	MURDER
INCARATE	1.0000 ( 50) P= .	.7032 ( 50) P= .000
MURDER	.7032 ( 50) P= .000	1.0000 ( 50) P= .

Quel est le degré de liaison entre ces deux variables ? .....

Justifier votre réponse : .....

.....

.....

Rappelons ici qu'un modèle linéaire simple se formalise de la façon suivante :

$$Y = aX + b + \varepsilon$$

où y représente la variable que l'on cherche à expliquer, X est la variable explicative et a, b et c sont les coefficients que l'on va chercher à estimer,  $\varepsilon$  étant l'aléas (ou bruit blanc) c'est-à-dire, la partie que l'on ne peut contrôler donc qu'il n'est pas possible de prévoir.

L'objectif est, alors, de calculer une équation qui va refléter le mieux possible la réalité. On aura ainsi :  $\hat{Y} = \hat{a}X + \hat{b}$  Les erreurs ont disparu du modèle de prévision.

La technique utilisée afin d'estimer ces coefficients est **la méthode des moindres carrés ordinaires**.

### Estimation des coefficients de régression

----- Variables in the Equation -----					
Variable	B	SE B	Beta	T	Sig T
INCARATE	.024962	.003642	.703246	6.853	.0000
(Constant)	.852739	.904850		.942	.3507

Donnez l'équation de la droite linéaire donnée par la figure ci-dessus.

.....

.....

.....

Les coefficients de cette droite sont-ils significatifs ? .....

.....

.....

.....

Justifiez votre réponse.

.....

.....

.....

**PARTIE 3 : Analyse factorielle** (cf. sorties correspondantes dans pages 1 à 4) **les retourner avec vos feuilles de réponses commentées et avec une légende.**

Les questions suivantes concernent une analyse factorielle effectuée sur des données décrivant les indicateurs macro économiques prélevés sur 802 entreprises marocaines. Le but de cette étude est de dégager une certaine typologie sur l'ensemble des entreprises sur la base des indicateurs les plus significatifs parmi le Chiffre d'Affaire, la Valeur ajoutée, l'Exportation, la Production, l'Investissement réalisé et les Importations globales.

On choisit de représenter chaque individu-entreprise  $i$  par le vecteur des valeurs prises par les variables  $x_1, \dots, x_6$  centrées et réduites pour les entreprises.

**1) Justifier les opérations de centrage et de réduction.**

.....

.....

.....

**2) L'analyse factorielle est-elle pertinente ? justifier votre réponse en utilisant tous les critères possibles.**

.....

.....

.....

.....

**3) a) En effectuant une analyse bivariée sur les variables, proposez un scénario possible de regroupement de variables. Vous verrez par la suite si ce regroupement est maintenu ou pas après l'extraction des facteurs.**

.....

.....

.....

.....

**4) Quelles sont les variables qui vont le plus contribuer à la construction des facteurs ? justifier**

.....

.....

.....

**5) L'extraction de facteurs a été faite à l'aide de l'Analyse en composantes principales**

**a) Comparer la qualité de représentation fournie par l'ACP obtenue par l'extraction de 3 et de 4 facteurs.**

.....

.....

**b) Quels sont les pourcentages d'inerties expliquées ?**

- à 3 facteurs : ..... à 4 facteurs : .....

**c) quels nombres de facteurs retenez vous? ..... justifiez**

.....

**d) Est-il nécessaire de faire une rotation pour la méthode d'extraction retenue dans 5.c). ?**

**Justifiez : .....**

.....

.....

**6) Selon votre réponse à la question 5) d) , donnez une interprétation des facteurs que vous retenez (sans ou avec rotation) de la méthode retenue en terme des activités des entreprises?**

.....

.....

.....

**7) a) A quoi correspondent les graphiques de la page 4 ?**

.....

.....

.....

**b) En analysant ces graphiques,**

- **Quel est le secteur ayant le plus grand taux d'investissement ?**

.....

.....

- **Quel est le secteur ayant le plus grand taux d'importation ?**

.....

**c) En ce basant sur le Cube OLAP, comment pouvez vous caractériser chacun des secteurs ?**

.....

.....

.....

.....

## **RECOMMANDATIONS**

- Avant de répondre, prenez le temps de lire l'ensemble des questions ;
- Il y a un enchaînement logique des questions afin que l'analyse soit complète à la fin de l'épreuve. Ne vous précipitez donc pas à répondre en sautant des questions. Le barème établi ne favorise pas forcément ceux qui répondent au maximum de questions mais plutôt ceux qui analyseront correctement les données de manière logique ;
- Pour justifier vos réponses, référer dans le texte aux sorties correspondantes (la figure ou le tableau de SPSS) après les avoir numéroté dans les pages de sortie et commenté en y incluant une légende.



Bonne Chance  
I. BERRADA