

EXAMEN DE STATISTIQUES DESCRIPTIVES - PLPSTA01
Licence de Psychologie 1^{ère} année
16 juin 2008

Tous les documents sont interdits. Un formulaire succinct est en page 2.

Les réponses doivent être justifiées.

Les deux exercices sont indépendants. Durée : deux heures.

Exercice 1 (Sujet tiré du devoir)

On étudie le devenir des déchets dans 21 régions de France métropolitaine en 2004 (la région Ile de France a été retirée de l'étude)¹.

La variable "Décharge" désigne les quantités de déchets mis en décharges autorisées, elle est exprimée en milliers de tonnes (notation kt).

La variable "Biologique" désigne les quantités de déchets traités de façon biologique, exprimée aussi en milliers de tonnes (kt).

On note X la variable "Biologique" et Y la variable "Décharge".

Région	Décharge	Biologique
CORSE	162	0
LIMOUSIN	156	19
CHAMPAGNE-ARDENNE	530	20
AUVERGNE	692	53
FRANCHE-COMTE	256	99
BASSE NORMANDIE	681	104
MIDI-PYRENEES	793	168
LORRAINE	1084	171
PROVENCE-ALPES-COTE D'AZUR	2364	178
POITOU-CHARENTES	634	183
BOURGOGNE	941	194
LANGUEDOC-ROUSSILLON	851	203
PICARDIE	1122	217
HAUTE NORMANDIE	402	219
ALSACE	367	243
AQUITAINE	914	261
CENTRE	905	341
BRETAGNE	288	355
RHONE-ALPES	1970	359
PAYS DE LOIRE	1633	392
NORD-PAS-DE-CALAIS	1602	393

1. Donnez la population, sa taille, la nature des variables étudiées.

2. Dans cette question, on étudie les données individuelles de la variable "Biologique".

(a) Calculez la moyenne et l'écart type de la variable "Biologique". On utilisera $\sum_{i=1}^{21} x_i = 4172$ et

$\sum_{i=1}^{21} x_i^2 = 1\,127\,190$. En déduire le tonnage centré réduit pour les deux régions suivantes : Corse, Provence-Alpes-Cotes d'Azur.

(b) Calculez la médiane.

¹Données de l'Institut Français de l'Environnement, souce ADEME

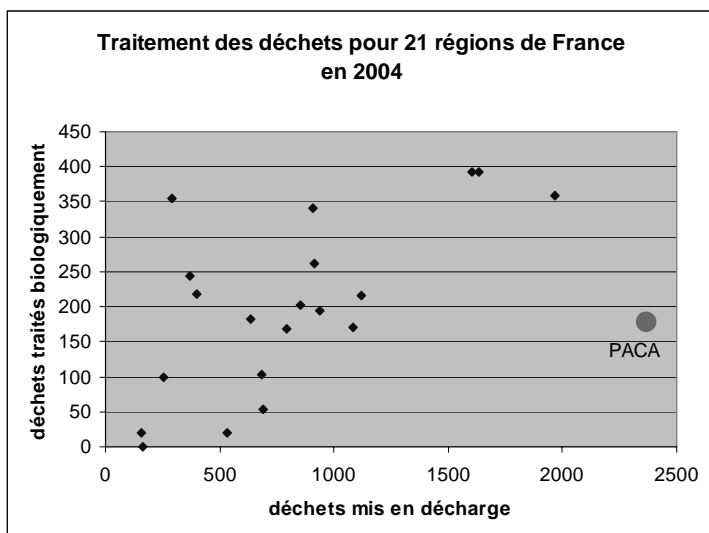
- (c) Calculez la proportion des régions dont la quantité de déchets traités de façon biologique (en kt) est comprise dans l'intervalle $[100, 300]$.
3. Dans cette question, on étudie les données regroupées de la variable "Biologique", selon le découpage en classes suivant :

Déchets traités biologiquement X	$[0, 150]$	$]150, 200]$	$]200, 250]$	$]250, 400]$
Effectifs	6	5	4	6

- (a) Calculez la moyenne de la variable "Biologique" à partir des données regroupées.
- (c) Calculez les proportions cumulées.
- (d) En utilisant la fonction de répartition, calculez la proportion des régions dont la quantité de déchets traités de façon biologique (en kt) est comprise dans l'intervalle $[100, 300]$.
4. On considère à présent les données individuelles du couple de variables "Décharge" et "Biologique".
- (a) Calculez le coefficient de corrélation linéaire entre ces deux variables (on donnera une précision de trois décimales). On utilisera le résultat de la question 2)a) ainsi que les données ci-dessous :

$$\mu_Y = 873,7 \quad \sigma_Y = 582,3 \quad \sum_{i=1}^{21} x_i y_i = 4\,404\,161$$

- (b) On donne ci-dessous la représentation graphique du couple de variables. Identifiez la région Provence-Alpes-Cote d'Azur (PACA) sur ce graphique. Sans faire de calcul, répondez à la question suivante : comment va évoluer le coefficient de corrélation linéaire si on enlève cette région de l'étude ?



Exercice 2

On a relevé le type de milieu de vie (variable X) et le temps de mémorisation (en secondes) de mots difficiles (variable Y) de 297 enfants de CM1. La distribution jointe est donnée dans le tableau ci-dessous :

$X \setminus Y$	$]0,10]$	$]10,20]$	$]20,25]$
Zone urbaine	41	70	42
Milieu rural	22	9	6
Milieu semi-urbain	45	50	12

- Déterminez la population, sa taille, et le type des variables étudiées.
- On considère le temps de mémorisation pour les enfants vivant en zone urbaine.
 - Donnez la distribution des proportions de cette variable.

- (b) Représentez graphiquement cette distribution. Quel est son mode? (justifiez)
3. On considère le temps de mémorisation pour les enfants vivant en milieu rural.
- (a) Donnez la distribution des proportions de cette variable.
- (b) Représentez graphiquement cette distribution, en prenant la même échelle que le graphique de la question 3. Quel est son mode? (justifiez)
4. Comparez les deux distributions précédentes. Que pouvez-vous conclure?
5. Calculez le pourcentage d'élèves qui ont un temps de mémorisation supérieur à 20 secondes.
6. Parmi les enfants vivant en milieu urbain ou semi urbain, quelle est la proportion de ceux qui ont un temps de mémorisation inférieur à 10 secondes?
7. Quelle est la proportion d'enfants qui ont un temps de mémorisation supérieur à 20 secondes et qui viennent de milieu rural?
8. On considère les enfants qui ont un temps de mémorisation inférieur à 10 secondes; quelle est la proportion de ceux vivant en milieu rural?

RAPPEL DE FORMULAIRE

Calcul de proportions cumulées :

$$\text{Si } x \in]a, b], \quad F(x) = F(a) + (F(b) - F(a)) \frac{x - a}{b - a}$$

Covariance pour des données individuelles

$$\text{Cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y) = \left(\frac{1}{N} \sum_{i=1}^N x_i y_i \right) - \mu_X \mu_Y$$

Coefficient de corrélation linéaire

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$