

Devoir de Statistique

On étudie le devenir des déchets dans les différentes régions de France métropolitaine en 2004. Sur le site de l'Institut Français de l'Environnement (IFEN), nous avons récupéré les données ci-après (source ADEME). 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).

Région	Décharge	Biologique
ALSACE	366,8	242,6
AQUITAINE	913,8	261,2
AUVERGNE	692,3	52,6
BASSE NORMANDIE	680,5	104,2
BOURGOGNE	941	193,9
BRETAGNE	287,5	355
CENTRE	905	340,5
CHAMPAGNE-ARDENNE	529,8	20,2
CORSE	162,2	0
FRANCHE-COMTE	256,2	98,7
HAUTE NORMANDIE	402,4	218,5
ILE DE FRANCE	2930,4	549,5
LANGUEDOC-ROUSSILLON	850,8	202,5
LIMOUSIN	156,2	19
LORRAINE	1084,3	170,7
MIDI-PYRENEES	792,5	168,2
NORD-PAS-DE-CALAIS	1601,8	392,8
PAYS DE LOIRE	1633,4	391,8
PICARDIE	1121,6	217,4
POITOU-CHARENTES	633,6	183,1
PROVENCE-ALPES-COTE D'AZUR	2363,6	177,8
RHONE-ALPES	1969,6	359,3

1. On s'intéresse aux déchets traités de façon biologique sur tout le territoire.

(a) Classez les données dans l'ordre croissant.

(b) Calculez la moyenne et l'écart type (sur les 22 régions). On utilisera $\sum_{i=1}^{22} x(i) = 4719,5$ et $\sum_{i=1}^{22} x(i)^2 = 1\,428\,203,05$. En déduire le tonnage centré réduit pour les régions suivantes : Corse, Auvergne, Franche-Comté, Rhone Alpes, Pays de Loire et Ile de France.

(c) Parmi les régions citées en b), quelles sont celles dont les tonnages (en kt) s'écartent de plus d'un écart type de la moyenne, à 10^{-1} près ? Quelles sont celles dont les tonnages s'écartent de plus de 2 écarts types de la moyenne ?

(d) En déduire (sans calcul supplémentaire) toutes les régions de France dont les tonnages s'écartent de plus d'un écart type du tonnage moyen (à 10^{-1} près).

Indication : on remarquera qu'avec des données centrées réduites, le classement des régions est le même que celui sur les données de départ.

(e) Pour la question 2, on élimine la région Ile de France de l'étude. Justifiez ce choix.

2. On s'intéresse aux déchets traités de façon biologique sur les 21 régions restantes (Ile de France exclue). On considère le découpage en classes suivant : $[0, 150[$, $[150, 200[$, $[200, 250[$, $[250, 400]$.

(a) Tracez l'histogramme de cette distribution.

(b) Au vu de l'histogramme, et sans faire de calculs, que pensez-vous de la position relative des indicateurs de tendance centrale ?

(c) Calculez ces indicateurs, ainsi que l'écart type, à partir des données individuelles et à partir des données regroupées. Commentez les résultats.

Indication : Pour les données individuelles, on utilisera $\sum_{i=1}^{21} x(i) = 4170$ et $\sum_{i=1}^{21} x(i)^2 = 1\,126\,252,8$.

(d) A l'aide de l'histogramme, calculez la proportion des régions dont la quantité de déchets traités de façon biologique (en kt) est comprise dans l'intervalle $[100, 300]$.

Retrouvez ce résultat en utilisant la fonction de répartition.

Comparez le résultat obtenu avec la proportion observée sur les données individuelles.

(e) Tracez la boîte à moustaches (pour les données regroupées).

3. On réunit les régions pour travailler sur cinq "super" régions :

- Sud-Ouest : Aquitaine, Limousin, Poitou-Charentes, Midi-Pyrénées,
- Sud-Est : Auvergne, Corse, Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Rhône-Alpes,
- Nord-Ouest : Basse-Normandie, Bretagne, Centre, Haute-Normandie, Pays de Loire,
- Nord-Est : Alsace, Bourgogne, Champagne-Ardenne, Franche-Comté, Lorraine, Nord-Pas-de-Calais, Picardie,
- Ile de France

(a) Pour chaque super région, calculez le tonnage moyen (en kt) de déchets traités biologiquement.

(b) En utilisant le résultat de la question a), calculez le tonnage moyen (en kt) de déchets traités biologiquement sur la France.

(c) Comparez le résultat ci-dessus (obtenu en b)) avec ceux de la question 1-b).

4. On s'intéresse au caractère "Respect de l'environnement" des 22 régions. Pour cela, on considère pour chaque région la part des déchets traités biologiquement dans l'ensemble des déchets (mis en décharge et traités biologiquement). On appelle "ratio bio" cette part. Et on classe les régions dans les quatre catégories suivantes selon la valeur de leur "ratio-bio".

La catégorie "Pollueur Extra" regroupe les régions dont le ratio-bio est inférieur à 10%.

La catégorie "Gros Pollueur" regroupe les régions dont le ratio-bio est compris entre 10% et 20%.

La catégorie "Pollueur moyen" regroupe les régions dont le ratio-bio est compris entre 20% et 30%.

La catégorie "Région d'avenir" regroupe les régions dont le ratio-bio est supérieur à 30%.

Par exemple, le ratio-bio de la région Alsace est égal à $\frac{242,6}{(366,8+242,6)} = 0,3981$ soit 39,81%. La région Alsace est dans la catégorie "Région d'avenir".

Réalisez le tableau des effectifs de la variable "Respect de l'environnement", et faites la représentation graphique de cette variable. Quel est son mode ?

5. On considère à présent le couple de variables "Décharge" et "Biologique" (données individuelles région par région sur les 22 régions).

(a) Représentez graphiquement ce couple de variables, la variable "Décharge" étant en abscisse.

(b) Calculez la corrélation entre ces deux variables. Commentez.

(c) Sans faire de calcul, répondez à la question suivante : si on enlève la région PACA, que devient la corrélation ?