

Série N°1 en Fouille de données (Rappels des statistiques)

Exercice n°1 : soit la série statistique suivante représentant la consommation en énergie de foyers par unité de mesure : 26, 150, 5, 6, 150, 5, 42, 150, 6, 13, 150, 5, 72.

1. Quel est le nombre de modalité de consommations ?
2. Calculer la moyenne des valeurs. Quelle est la valeur qui influence la valeur moyenne. Justifier la réponse.
3. Calculer et interpréter le mode, la médiane et l'écart-type de cette série.
4. Représenter graphiquement par un diagramme adéquat et justifier son utilité.

Exercice n°2 : le tableau suivant représente les notes obtenues par quinze candidats dans un concours :

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
10,2	11,23	8,31	12	17	7,63	8,18	10	9,31	10	14,12	16,77	9,87	7,44	10,01

1. Expliquer la raison pour laquelle les valeurs continues ne sont pas adéquates à la représentation graphique par les représentations des valeurs discrètes.
2. Proposer une discrétisation des valeurs de la série. Représenter graphiquement la nouvelle série obtenue.
3. En utilisant la loi de Sturges ($k = 1 + 3,322 \log_{10} n$) où n représente le nombre de valeurs observées, calculer le nombre k de classes de modalités et déterminer ces classes. En déduire la nouvelle distribution des notes.
4. Représenter la nouvelle distribution graphiquement par un graphique adéquat.
5. Comparer le graphique de la question 2 avec celui de la question 4.

Exercice n° 3 : On veut analyser la relation entre le nombre de bateaux observés et le nombre de poissons tués. On se limite à l'étude d'une seule espèce de poissons. Le tableau suivant donne les valeurs entre 1981 et 1990 (référence : econometrie.cnam.fr).

Année	81	82	83	84	85	86	87	88	89	90
Nbr bateaux	516	512	528	560	584	616	644	676	712	720
Nbr poissons	75	83	86	83	99	101	106	111	131	130

1. Dessiner le nuage de points de cette série.
2. Calculer et interpréter la covariance et le coefficient de corrélation linéaire entre les deux variables.

$$\hat{y} = \frac{\text{cov}(x, y)}{V(x)}(x - \bar{x}) + \bar{y}$$

3. Dessiner la droite pour l'ajustement de cette série.
4. Identifier les erreurs possibles de mesure et présenter un tableau correctif.