Feuille d'exercices 8

Exercice 1. On a relevé les revenus d'un échantillon de trente foyers. Les valeurs sont les suivantes (en centaines d'euros), présentés par ordre croissant :

186	201	211	238	260	277	277	281	292	295
318	325	325	328	336	431	432	496	543	567
571	585	623	625	696	747	788	871	944	959

1. Donnez la médiane de cette série.
2. Calculez la moyenne et l'écart-type.
3. Regroupez les données en classes de longueur 100 et représentez les données regroupées sous forme de tableau.
4. Tracez l'histogramme et le polygône des fréquences cumulées.
5. Calculez la médiane, la moyenne et l'écart-type sur les données regroupées en classes, comparez les résultats à ceux obtenus sur les données brutes.

Exercice 2. Pour comparer le nombre d'enfants par famille dans deux villes A et B, on choisit un échantillon de familles par ville et pour chacunes d'elles, on note le nombre d'enfants. Les répartitions des familles des échantillons suivant le nombre d'enfant sont les suivantes :

Echantillon A

nombre k d'enfants	0	1	2	3	4	5	6	7	8
familles avec k enfants	93	146	104	63	47	33	10	4	0

Echantillon B

nombre k d'enfants	0	1	2	3	4	5	6	7	8
familles avec k enfants	56	99	129	95	67	38	10	3	3

1. Construire le diagramme en bâton de la première série statistique.
2. Calculer le nombre moyen d'enfants par famille dans chaque échantillon.
3. Calculer la médiane et les quartiles des séries statistiques.
4. Construire les boites à moustaches et comparer.

Exercice 3. Considérons le tableau suivant donnant la répartition des salaires mensuels dans une entreprise

Salaires	Nbre de personnes
[1000,1500[	11
[1500,2000[	26
[2000,2500[	63
[2500,3000[	81
[3000,3500[	35
[3500,4000[	21
[4500,5000[	13

Tracez la courbe de Lorenz de cette série et calculez son indice de Gini.

Exercice 4. Soit $\left((x_i, y_j, n_{i,j})\right)_{\substack{1\le i\le p\\ 1\le i\le q}}$ une série statistique double. On note $N = \sum_{i,j} n_{i,j}$ la taille de la population, $\bar{x} = \frac{1}{N}\sum_{i,j} n_{i,j}x_i$ $\bar{y} = \frac{1}{N}\sum_{i,j} n_{i,j}y_j$ les valeurs moyennes des deux caractères. Ces deux caractères sont indépendants si

$$\frac{n_{i,j}}{N} = \frac{\sum_{i=1}^{p} n_{i,j}}{N}\frac{\sum_{j=1}^{q} n_{i,j}}{N}.$$

Ils sont non corrélés si le coefficient de corrélation linéaire $\rho_{x,y} = \dfrac{\overline{(x-\bar{x})(y-\bar{y})}}{\sqrt{\overline{(x-\bar{x})^2}}\sqrt{\overline{(x-\bar{x})^2}}}$ est nul. Montrez que des caractères indépendants sont non corrélés.

Exercice 5. Pour vérifier les relations d'allométries entre insectes, on a retenu les deux variables suivantes
X = logarithme de la longueur de l'étytre,
Y = logarithme de la largeur de la tête
Les mesures sur 50 insectes ont fournis les résultats sous forme de tableau $(x_i, y_i)_{1\le i\le 50}$ vérifiant

$$\begin{array}{lll} \sum x_i = 155 & \sum y_i = 125 & \sum x_i y_i = 391.1 \\ \sum x_i^2 = 482.5 & \sum y_i^2 = 320.5 & \sum x_i^2 y_i^2 = 3468.7 \end{array}$$

Calculez

1. La moyenne et l'écart-type du caractère X sur l'échantillon observé.
2. La moyenne et l'écart-type du caractère Y sur l'échantillon observé.
3. La covariance $cov(x, y)$ et le coefficient de corrélation $\rho_{x,y}$.
4. L'équation de la droite de régression de Y en X obtenue à partir de ces données.
5. L'équation de la droite de régression de X en Y obtenue à partir de ces données.
6. La loi d'allométrie exprimant la largeur de la tête en fonction de la longueur de l'élytre.

Exercice 6. On se propose d'étudier l'influence de la température T sur la durée d'incubation des œufs de grenouilles. On choisit six échantillons de deux cents œufs chacun. Le nombre X d'éclosions au vingt-deuxième jour est le suivant :

température t_i d'incubation	6	6,4	6,8	7,2	7,6	8
nombre d'éclosions x_i	131	144	157	170	190	189

1. Dessinez le nuage de points des données. Tracer une droite semblant bien approcher ce nuage.
2. Calculer le coefficient de corrélation de l'observation et donnez l'équation de la droite de régression. Comparer avec la droite précédente.
3. Donnez la nombre prédit par la droite de régression sur un échantillon de 200 œufs à 7,5 degrés Celsius.

Exercice 7. On a mesuré les variables x et y sur dix individus et obtenu

individu i	1	2	3	4	5	6	7	8	9	10
x_i	18	20	19	16	19	16	19	21	15	17
y_i	43	110	70	17	91	29	80	134	15	34

On cherche une relation de la forme $y = a\ln x + b$. Pour cela on pose $z = \ln x$.
– Dessiner le nuage de points $(z_i, y_i)_{1\le i\le 10}$ et calculer la droite de régression.
– Quelle valeur de y peut-on prédire pour un individu présentant pour x la valeur 22 ?