Statistique descriptive

Exercice 1 La répartition de cent familles suivant le nombre d'enfants est représentée par le tableau suivant :

nombre n d'enfants	0	1	2	3	4	5
nombre de familles ayant n enfants	8	23	42	18	6	3

1. Construire les diagrammes en bâtons des fréquences et des fréquences cumulées de cette série statistique.
2. Déterminer le mode et calculer la médiane et la moyenne de la série statistique.
3. Calculer les premier et troisième quartiles, la variance et l'écart-type de la série statistique.

Exercice 2 Dans un standard téléphonique, on a relevé la durée de 500 communications. Le tableau suivant donne le nombre n de communications ayant duré plus de x minutes :

x	1	2	3	4	5	6	7	8	9	10
n	460	395	310	242	195	166	141	118	98	80

x	11	12	13	14	15	16	17	18	19	20
n	63	50	35	23	17	10	5	3	2	0

1. Construire l'histogramme des fréquences et la courbe des fréquences cumulées.
2. Calculer la durée moyenne d'une communication téléphonique, puis l'écart-type de la série des durées des communications téléphoniques.
3. Déterminer la médiane, les quartiles et l'intervalle interquartile de cette série.

Exercice 3

Thème. — On se propose à partir d'exemples très simples, d'une part de mettre en évidence certaines propriétés de minimisation pour la médiane et la moyenne, et d'autre part de considérer la population statistique des moyennes et des variances des échantillons (cette population est à la base de problèmes touchant à la statistique inférentielle).

Énoncé. — Dans ce qui suit, on considère les deux séries statistiques suivantes :

x'_i	2	3	4	6
n_i	2	4	1	1

y_i	3	4	6	7
p_i	1	2	2	1

qui serviront d'exemples.

Partie 1. — Questions sur la médiane.

1. Déterminer les médianes M_x et M_y des deux séries définies ci-dessus (pour la deuxième, il est permis de prendre une valeur arbitraire d'un certain intervalle ou alors de prendre le centre de cet intervalle).
2. On considère les fonctions S_1 et S_2 définies pour les deux séries par : $S(a) = \sum_i n_i|a - x'_i|$ (c'est la somme des écarts absolus des valeurs du caractère au nombre a). Déterminer suivant les intervalles dans lesquels on place a les valeurs de $S(a)$ en fonction de a. Montrer que S_1 et S_2 sont affines par morceaux. Dessiner les graphes de ces fonctions et montrer que chacune de ces fonctions passe par un minimum. Donner une conclusion en comparant la médiane à la valeur de a correspondant à ce minimum.
3. Généralisation : on considère une série dans laquelle les valeurs du caractère sont rangées par ordre croissant, les effectifs partiels étant tous égaux à 1. Démontrer la propriété précédente lorsque l'effectif total est $2p+1$, cas pour lequel la médiane est x_{p+1}. Démontrer ensuite cette même propriété lorsque l'effectif total est $2p$.

Partie 2. — Questions sur la moyenne.

1. Calculer les moyennes m_1 et m_2 des deux séries données.
2. On considère pour la première série par exemple la somme des écarts au carré : $V(a) = \sum_i n_i(a - x'_i)^2$. En étudiant le trinôme du second degré obtenu, montrer que cette somme est minimum lorsque $a = m_1$. Étudier la deuxième série.
3. Dans le cas général, montrer en exprimant $V(a) - nV$ où $V(a)$ est la variance que la somme $V(a)$ est minimum lorsque a est égal à la moyenne de la série statistique étudiée.

Exercice 4

Thème. — Caractère continu. On donne ici un exemple de construction d'un histogramme avec des classes d'amplitude inégale.

Énoncé. — En 1988, la distribution des âges des habitants de la région Bretagne est la suivante (en milliers) :

âge x	effectifs des classes
$0 \leqslant x < 5$	158
$5 \leqslant x < 15$	427
$15 \leqslant x < 25$	432
$25 \leqslant x < 35$	403
$35 \leqslant x < 45$	293
$45 \leqslant x < 55$	306
$55 \leqslant x < 60$	157
$60 \leqslant x < 75$	353
$75 \leqslant x < 100$	173

1. Construire l'histogramme et calculer la moyenne de cette série statistique.
2. On voudrait « rajeunir » la population en favorisant la venue de personnes de la classe d'âge [25,35[. Combien faudrait-il en faire venir pour que la moyenne d'âge soit de 35 ans ? Est-ce réaliste ?
3. Après avoir déterminé les effectifs cumulés croissants, représenter graphiquement la fonction de répartition de la série (courbe cumulative).
4. Déterminer une valeur approchée de la médiane graphiquement et par le calcul.

Exercice 5

Thème. — Ajustement affine : droite de Mayer et droite de régression par la méthode des moindres carrés.

Exposé de la méthode (droite de Mayer). — Étant donné un nuage de n points $M_i(x_i, y_i)$, $i = 1, \ldots, n$, on cherche à l'ajuster par une droite. Une droite étant complètement déterminée par deux points, cela conduit à la méthode suivante :

- On divise le nuage de n points en deux sous-nuages à peu près d'égale importance et de telle sorte que l'abscisse de tous les points du premier soit strictement inférieure à l'abscisse de tous les points du second. On parlera donc de « sous-nuage droit » et de « sous-nuage gauche ».
- On détermine le point moyen G_d du sous-nuage droit et le point moyen G_g du sous-nuage gauche.
- La droite d'ajustement cherchée est alors la droite $(G_g G_d)$.

Énoncé. — Le tableau suivant donne pour un atelier de fabrication le nombre d'heures supplémentaires (NHS) mensuelles imposées par la direction et la production totale mensuelle correspondante (P) :

NHS	0	20	30	50	60	90	100
P	150	162	165	170	172	177	187

1. Tracer le nuage de points dans un repère orthogonal.
2. En prenant pour « nuage gauche » les trois premiers points, déterminer une équation de la droite d'ajustement par la méthode de Mayer.
3. Déterminer une équation de la droite d'ajustement par la méthode des moindres carrés. Comparer le résultat obtenu avec celui de la question précédente.

Exercice 6

Thème. — Ajustement non linéaire mais linéarisable par changement déchelle.

Énoncé. — Dans une réaction enzymatique, on mesure la vitesse initiale en fonction de la température. On représentera par ν_k le produit de 10^7 par la vitesse initiale à la température t_k exprimée en degré Celsius. On a obtenu les résultats suivants :

k	1	2	3	4	5	6	7	8	9
t_k	10	15	20	25	30	35	40	45	50
ν_k	2.818	3.162	3.951	5.012	6.310	8.913	10.000	14.130	15.850

1. Tracer dans un repère orthogonal le nuage de points (t,ν_k), $k = 1, \ldots, 9$. En traçant une courbe d'ajustement « à l'estime », constater qu'un ajustement par une droite n'est pas possible.
2. On effectue un changement d'échelle. A cet effet, sur l'axe des abscisses on porte $x_k = \frac{1}{T_k}$ où T_k représente la température absolue ($T_k = t_k + 273$) et sur l'axe des ordonnées $y_k = \log \nu_k$ (logarithme décimal). Que constate-t-on ?
3. Effectuer un ajustement linéaire par la méthode des moindres carrés sur le nuage (x_k, T_k). En déduire l'équation de la courbe d'ajustement du nuage (t_k, ν_k).