Licence MIASHS - UCO

Contrôle n° 1, Statistiques S3

le 9 novembre 2023

* * *

Durée : 1h15

Documents et calculatrice interdits

Ex. 1. On considère la série suivantes :

classes	effectifs
$10 \leqslant x < 12$	10
$12 \leqslant x < 13$	10
$13 \leqslant x < 14$	40
$14 \leqslant x < 16$	30
$16 \leqslant x < 20$	10

1. Déterminer la moyenne de cette série par interpolation linéaire.
2. Tracer l'histogramme de cette série.
3. Déterminer le premier quartile, la médiane ainsi que le troisième quartile par interpolation linéaire.

Ex. 2. Soient $x_1, x_2, ..., x_n$ les n valeurs d'une série statistique et $\overline{x}$ la moyenne arithmétique de cette série.

1. Démontrer que :
$$\frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})^2 = \frac{1}{n}\sum_{k=1}^{n}x_k^2-\overline{x}^2$$
2. Application.
Une étude statistique $(x_i)_{1\leqslant i\leqslant 9000}$ faite sur 9000 individus donne les résultats suivants :
— Moyenne : $\overline{x} = 10$.
— Ecart type : $\sigma_x = 4$
Une seconde étude statistiques $(y_i)_{1\leqslant i\leqslant 1000}$ du même type et faite sur 1000 individus donne les résultats suivants :
— Moyenne : $\overline{y} = 16$
— Ecart type : $\sigma_y = 2$
   (a) Déterminer la moyenne de l'ensemble des individus composant ces deux séries $(x_i)_{1\leqslant i\leqslant 9000}$ et $(y_i)_{1\leqslant i\leqslant 1000}$.
   (b) Montrer que l'écart type σ de l'ensemble des individus composants ces deux séries $(x_i)_{1\leqslant i\leqslant 9000}$ et $(y_i)_{1\leqslant i\leqslant 1000}$ est strictement supérieur à 4.
   (c) Commet peut-on expliquer un écart type σ supérieur à 4 alors que $\sigma_x = 4$ et $\sigma_y = 2$.

Ex. 3. On considère les séries statistiques $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_n$ donnant les résultats de deux caractères X et Y d'une population de 100 individus tels que pour tout individu ω_i, on a : $X(\omega_i) = x_i$ et $Y(\omega_i) = y_i$.
De plus, on donne :

$$\sum_{i=1}^{100} x_i = 1200 \quad ; \quad \sum_{i=1}^{100} y_i = 1400 \quad ; \quad \sum_{i=1}^{100} x_i^2 = 16900 \quad ; \quad \sum_{i=1}^{100} y_i^2 = 20500$$

$$\sum_{i=1}^{100} x_i y_i = 18000 \quad ; \quad \sum_{i=1}^{100} (y_i - \overline{y})^3 = -1080 \quad ; \quad \sum_{i=1}^{100} (y_i - \overline{y})^4 = 32400$$

1. Calculer $\overline{x}$, $\overline{y}$, l'écart type relatif au caractère X, noté σ_x et l'écart type relatif au caractère Y noté σ_y.
2. Droite de régression linéaire.
   (a) Déterminer la covariance $\mathrm{Cov}(X, Y)$ puis le coefficient de corrélation linéaire des deux caractères X et Y.
   (b) Donner les coefficients $\widehat{a}$ et $\widehat{b}$ de la droite de régression linéaire de y en fonction de x.
   (c) Donner une approximation à 1% du pourcentage de variation du caractère Y par rapport au caractère X.
3. Coefficient d'asymétrie et de kurtosis.
   (a) Calculer les coefficients d'asymétrie et d'aplatissement de kurtosis de la série $(y_i)_{1 \leqslant i \leqslant n}$.
   (b) Faire un schéma simplifié de la distribution du caractère Y par rapport à une distribution normale.

Ex. 4. Soient a un réel et X une variable aléatoire de densité f définie par :

$$f(x) = \begin{cases} \dfrac{2}{a}\left(1 - \dfrac{x}{a}\right) & \text{si } 0 \leqslant x \leqslant a \\ 0 & \text{sinon} \end{cases}$$

1. Vérifier que f définie bien une densité de probabilité.
2. Déterminez la fonction de répartition de X et en déduire la médiane de cette loi.
3. Calculer $E(X)$ et $V(X)$.
4. Soient $X_1, X_2, \ldots, X_n$ des variables indépendantes de même loi que X.
   Déterminer la fonction de répartition puis la densité de la variable aléatoire m_n définie pour tout $n \in \mathbb{N}^*$ par :
   $$M_n = \max\{X_1, \ldots, X_n\}$$