

Probabilités et statistiques

Semestre 3 – UEM2.1 – Cours et TD

A.2 Séries statistiques à une variable

A.2.1 Effectifs, fréquences, cumuls

Soit une variable X de modalités $x_1 < \dots < x_k$ observée sur n individus.

- **Effectif** n_i : nombre d'individus de modalité x_i ; $\sum_i n_i = n$.
- **Fréquence** $f_i = n_i/n$; $\sum_i f_i = 1$; **pourcentage** $100 f_i$.
- **Effectif cumulé croissant** $N_i = n_1 + \dots + n_i$; **fréquence cumulée** $F_i = N_i/n$.

Pour une variable continue, on regroupe en classes $[a_i, a_{i+1}[$ d'amplitude $h_i = a_{i+1} - a_i$ et de centre $c_i = (a_i + a_{i+1})/2$.

A.2.2 Représentations graphiques

Type de variable	Graphique
qualitative	diagramme à bandes, diagramme circulaire (angle = $360^\circ \times f_i$)
discrète	diagramme en bâtons, polygone des effectifs (ou fréquences), courbe cumulative en escalier
continue	histogramme, polygone, courbe cumulative (polygone des fréquences cumulées)

À retenir

Dans un **histogramme**, c'est l'*aire* du rectangle qui est proportionnelle à l'effectif. Si les classes n'ont pas la même amplitude, la hauteur est n_i/h_i (densité d'effectif).

A.2.3 Caractéristiques de position

Définition

- **Moyenne** : $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_i f_i x_i$ (pour des classes, x_i est remplacé par le centre c_i).
- **Mode** : modalité d'effectif maximal (classe modale pour les données groupées). Dans une classe modale $[a, b]$: $M_o = a + h \frac{d_1}{d_1 + d_2}$, où d_1 (resp. d_2) est l'écart entre l'effectif (ou la densité) de la classe modale et celui de la classe précédente (resp. suivante).
- **Médiane** M_e : valeur qui partage la série en deux parties de même effectif ($F(M_e) = 0,5$). Par interpolation linéaire dans la classe médiane $[a, b]$: $M_e = a + h \frac{n/2 - N_{\text{préc}}}{n_{\text{classe}}}$.
- **Quartiles** Q_1, Q_3 : $F(Q_1) = 0,25$ et $F(Q_3) = 0,75$ (même méthode).

A.2.4 Caractéristiques de dispersion

Définition

- **Étendue** : $e = x_{\max} - x_{\min}$; **intervalle interquartile** $Q_3 - Q_1$.
- **Variance** : $V(X) = \frac{1}{n} \sum n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum n_i x_i^2 - \bar{x}^2$ (formule de König-Huygens).

- **Écart-type** : $\sigma = \sqrt{V(X)}$ (même unité que X).
- **Coefficient de variation** : $CV = \sigma/|\bar{x}|$ (sans unité, pour comparer des séries d'échelles différentes).

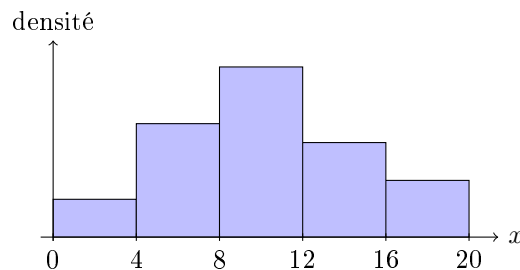
Théorème / Propriété : transformation affine

Si $Y = aX + b$ alors $\bar{y} = a\bar{x} + b$, $V(Y) = a^2V(X)$ et $\sigma_Y = |a|\sigma_X$.

A.2.5 Caractéristiques de forme

Avec $\mu_r = \frac{1}{n} \sum n_i(x_i - \bar{x})^r$ (moment centré d'ordre r) :

- **asymétrie** : coefficient de Pearson $\frac{\bar{x} - M_o}{\sigma}$, coefficient de Fisher $\gamma_1 = \frac{\mu_3}{\sigma^3}$ (> 0 : étalée à droite, 0 : symétrique, < 0 : étalée à gauche) ;
- **aplatissement** : $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$ (0 pour une loi normale ; > 0 plus pointue, < 0 plus aplatie).



Histogramme de l'exercice 2.2 (densités n_i/h_i : 1; 3; 4,5; 2,5; 1,5).

Exemple

$n = 40$ familles, $X =$ nombre d'enfants : $x_i = 0, \dots, 5$; $n_i = 4, 10, 14, 7, 4, 1$. Alors $\bar{x} = 80/40 = 2$, $\sum n_i x_i^2 = 218$, $V = 218/40 - 4 = 1,45$, $\sigma \approx 1,20$, $M_e = M_o = 2$.

Exercices

Exercice A.2.1

Reprendre la série de l'exemple ($x_i = 0, \dots, 5$; $n_i = 4, 10, 14, 7, 4, 1$). Dresser le tableau complet (f_i , N_i , F_i) puis calculer $\bar{x}$, M_e , M_o , V , σ , CV et l'étendue.

Exercice A.2.2

Le temps de préparation (en minutes) de 50 étudiants est réparti ainsi : $[0, 4[: 4$; $[4, 8[: 12$; $[8, 12[: 18$; $[12, 16[: 10$; $[16, 20[: 6$. Tracer l'histogramme, calculer la moyenne, la variance, l'écart-type, la médiane, les quartiles et le mode ; calculer le coefficient d'asymétrie de Pearson et conclure.

Exercice A.2.3

Une série a pour moyenne 5 et pour variance 4. Déterminer moyenne, variance, écart-type et CV de la série $y_i = 3x_i + 1$. Redémontrer la formule de König-Huygens.

Exercice A.2.4

Dans une promotion, un groupe de 30 étudiants a une moyenne de 12 et une variance de 4 ; un second de 20 étudiants a une moyenne de 14 et une variance de 9. Calculer moyenne et variance de la promotion entière.

Exercice A.2.5

La série A a pour moyenne 50 et pour écart-type 5 ; la série B a pour moyenne 200 et pour écart-type 15. Laquelle est la plus homogène ?