Statistique descriptive à une variable

3) Etude du cas d'une variable statistique continue :

Soit X une variable statistique continue définie sur une population de $\boldsymbol{n}$ individus et dont les observations sont regroupés en $\boldsymbol{k}$ classes : $[\boldsymbol{e_0}, \boldsymbol{e_1}[$, $[\boldsymbol{e_1}, \boldsymbol{e_2}[$, …, $[\boldsymbol{e_{k-1}}, \boldsymbol{e_k}[$ dont les effectifs respectifs $n_1, n_2, \dots, n_k$.

Exemple : Un échantillon de 50 poissons de la même espèce a fourni les poids suivants (en g) :

Poids (g)	Ampl. a_i	Centre x_i	Effectif n_i	Fréq. fi
[70; 80[	10	75	8	0, 16
[80; 90[	10	85	$8n_i$	0, 16
[90; 100[	10	95	12	0, 24
[100; 110[	10	105	17	0, 34
[110; 120[	10	115	5	0, 10
total			50	1

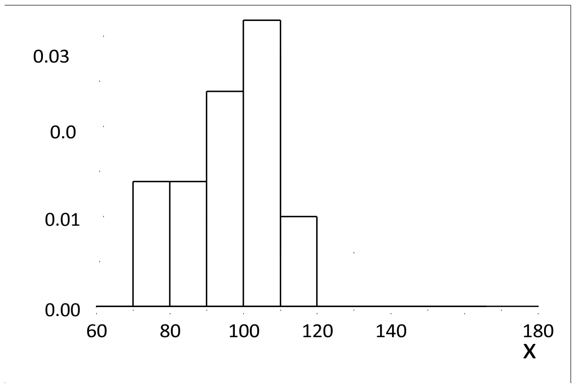
$\boldsymbol{a_i = e_i - e_{i-1}}$: Amplitude de la classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$.

$\boldsymbol{x_i} = \frac{e_i + e_{i-1}}{2}$: centre de la classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$.

$\boldsymbol{n_i} = \frac{n_i}{n}$: effectif de la classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$.

3.1 Représentation graphique : Histogramme

A chaque classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$, on fait correspondre un rectangle de base égale à son amplitude et hauteur proportionnelle à sa fréquence f_i.



Remarque : En reliant les points de coordonnées $(\boldsymbol{x_i}, f_i)$ par des segments de droites dans l'histogramme, on obtient une courbe dite Polygone des fréquences.

3.2 Fréquences cumulées et courbes cumulatives :

$Fi \uparrow = f1 + \dots + fi$ dite fréquence cumulée croissante, correspondante à $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$, est la proportion des observations inférieure à $\boldsymbol{e_i}$.

$Fi \downarrow = f1 + \dots + fi$ dite fréquence cumulée croissante, correspondante à $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$, est la proportion des observations supérieur ou égale à $\boldsymbol{e_{i-1}}$.

COURBES CUMULATIVES :

La courbe cumulative croissante (respectivement décroissante) de la variable statistique X est la représentation graphique de la fonction $F\uparrow(x) : \mathbb{R} \to [0,1]$ (respectivement $F\downarrow(x) : \mathbb{R} \to [0,1]$) définie par

$$\boldsymbol{F\uparrow(x)} = \begin{cases} \boldsymbol{0} & \boldsymbol{si\ x < e_0} \\ \boldsymbol{\frac{f_i}{a_i}(x - e_{i-1}) + F_{i-1}\uparrow} & \boldsymbol{si\ x \in [e_{i-1}, e_i[} \\ \boldsymbol{1} & \boldsymbol{si\ x \geq e_k} \end{cases}$$

(Respectivement $$\boldsymbol{F\downarrow(x)} = \begin{cases} \boldsymbol{1} & \boldsymbol{si\ x \leq e_0} \\ \boldsymbol{\frac{f_i}{a_i}(e_i - x) + F_{i+1}\downarrow} & \boldsymbol{si\ x \in [e_{i-1}, e_i[} \\ \boldsymbol{0} & \boldsymbol{si\ x > e_k}. \end{cases}$$

Remarque :

1) $F\uparrow$ et $F\downarrow$ sont appelées fonctions de répartitions de la variable statistique X.

2) $F\uparrow(x)$: représente la proportion d'observations qui sont inférieure à x.

3) $F\uparrow(e_i) = F_i\uparrow \quad pour\ tout\ i = 1,2, \dots, k.$

4) $F\downarrow(e_i) = F_{i+1}\downarrow \quad pour\ tout\ i = 0,1, \dots, (k-1)$

3.3 Représentation numérique des données :

3.3.1 Paramètres de tendance centrale :

i) Le MODE : noté M_o.

La classe dont la fréquence est la plus élevée est dite classe modale.

Soit $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$ la classe modale, alors $M_o \in [\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$.

La valeur du mode M_o peut-être approchée de deux manières :

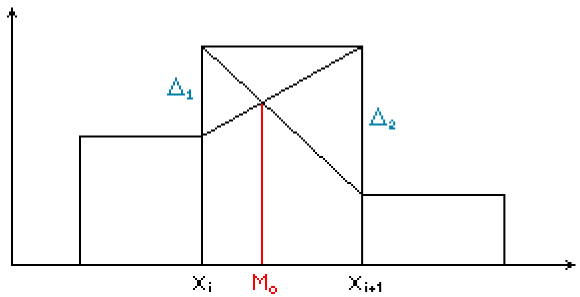
Première méthode : $M_o = \frac{e_{i-1}+e_i}{2}$ centre de la classe modale.

Deuxième méthode : En utilisant la règle de Thales on obtient

$$\boxed{\boldsymbol{M_o = e_{i-1} + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2}}}$$

Où : $\Delta_1 = \boldsymbol{f_i - f_{i-1}}$

$\Delta_2 = \boldsymbol{f_i - f_{i+1}}$.



ii) La MEDIANE : Notée M_e. Elle vérifie $F\uparrow(M_e) = F\downarrow(M_e) = \frac{1}{2}$.

La classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$ est dite classe médiane si $M_e \in [\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$.

Calcul de la médiane : soit $F_{i-1}\uparrow \leq \frac{1}{2} < F_i\uparrow \Rightarrow M_e \in [\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$. D'après la courbe cumulative croissante on obtient : $\frac{F_i\uparrow - F_{i-1}\uparrow}{e_i - e_{i-1}} = \frac{F\uparrow(M_e) - F_{i-1}\uparrow}{M_e - e_{i-1}} \Rightarrow$

$\frac{f_i}{a_i} = \frac{0.5 - F_{i-1}\uparrow}{M_e - e_{i-1}} \Rightarrow$

$$\boxed{\boldsymbol{M_e = \frac{a_i}{f_i}(0.5 - F_{i-1}\uparrow) + e_{i-1}}}$$

En particulier si $\exists \boldsymbol{i \text{ } tel \text{ } que \text{ } F_i\uparrow = 0.5}$ **alors** $\boldsymbol{M_e = e_i}$.

iii) La MOYENNE ARITHMETIQUE : $notée\ \bar{x}$.

$$\boxed{\boldsymbol{\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i}}$$

où $x_i = \frac{e_{i-1}+e_i}{2}$ $centre\ de\ la\ classe$ $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$

et $\boldsymbol{n_i}$: l'effectif correspondant à la classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$.

3.3.2 Paramètre de dispersions :

i) VARIANCE et ECART TYPE :

$$\boxed{\boldsymbol{V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2} \quad \textbf{ou encore} \quad \boldsymbol{V(X) = \left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2\right) - \bar{x}^2}}$$

où $x_i = \frac{e_{i-1}+e_i}{2}$ $centre\ de\ la\ classe$ $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$ et $\boldsymbol{n_i}$: l'effectif correspondant à la classe $[\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$. $\boldsymbol{\bar{x}}$ **:** Moyenne arithmétique de X.

L'écart type de X est défini par :

$$\boxed{\sigma_X = \sqrt{\boldsymbol{V(X)}}.}$$

ii) INTERVALLE INTERQUARTILES :

Les quartiles notés, Q_1, Q_2, Q_3 vérifient :

$F\uparrow(Q_1) = 0.25,\ F\uparrow(Q_2) = F\uparrow(M_e) = 0.5,\ F\uparrow(Q_3) = 0.75.$

L'intervalle interquartile est donné par : $\boldsymbol{[Q_1, Q_3]}$, il mesure la dispersion des observations autour de la médiane $\boldsymbol{M_e}$.

Calcul du quartile Q_1 : soit $F_{i-1}\uparrow \leq 0.25 < F_i\uparrow \Rightarrow Q_1 \in [\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$. En procédant de même que pour le cas de la médiane, on obtient :

$$\boxed{Q_1 = \boldsymbol{\frac{a_i}{f_i}(0.25 - F_{i-1}\uparrow) + e_{i-1}}}$$

Calcul du quartile $\boldsymbol{Q_3}$: soit $F_{i-1}\uparrow \leq 0.75 < F_i\uparrow \Rightarrow Q_3 \in [\boldsymbol{e_{i-1}}, \boldsymbol{e_i}[$. En procédant de même que pour le cas de la médiane, on obtient :

$$\boxed{Q_3 = \boldsymbol{\frac{a_i}{f_i}(0.75 - F_{i-1}\uparrow) + e_{i-1}}}$$