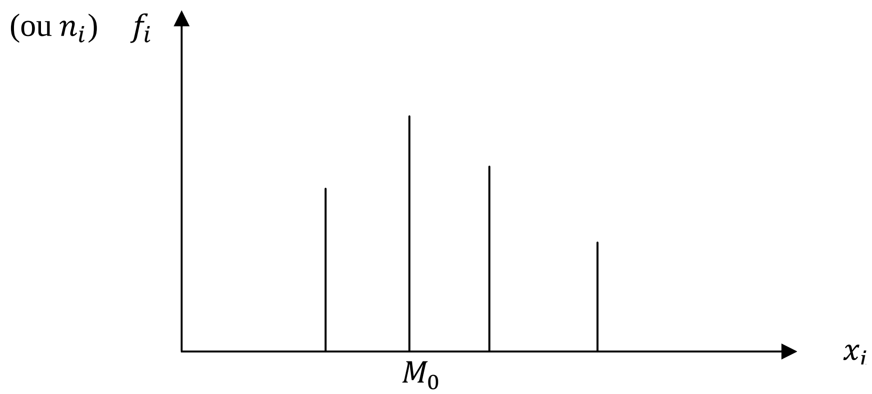
Paramètres de position (ou de tendance centrale)

Les paramètres de position (ou de tendance centrale) sont un ensemble de valeurs caractéristiques qui permettent une représentation condensée de l'information dans la série statistique. Les paramètres de position (la moyenne, la médiane, le mode et les quantiles) donnent l'ordre de grandeur de l'ensemble des mesures.

1. Le Mode (notation M_0)

Le mode d'une distribution statistique est la valeur de la variable qui correspond au plus grand effectif ou à la plus grande fréquence (valeur dominante).

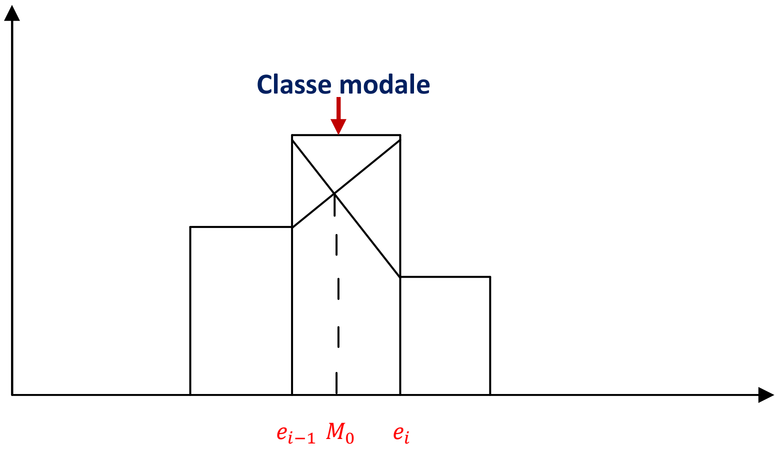
- **Cas discret :**
  - ✓ Sur le tableau statistique $(x_i, f_i)_{i=1,k}$ ou $(x_i, n_i)_{i=1,k}$, c'est le x_i pour lequel la fréquence est la plus élevée (effectif plus élevé).
  - ✓ Sur le diagramme en bâtons c'est la valeur x_i correspondant au bâton le plus haut.



Exemple : Nombre d'absences

$M_0 = 2$ qui correspond au plus grand effectif observé $n_3 = 11$.

- **Cas continu :** On suppose que les données sont regroupées dans des classes de même amplitude a.
  - ✓ **Classe modale** : C'est la classe du tableau ou de l'histogramme correspondant à la fréquence (effectif) maximum. C'est la classe qui contient le mode.
  - ✓ On peut déterminer graphiquement la valeur du mode (à l'intérieur de la classe modale) par la méthode des diagonales



- ✓ On peut déterminer analytiquement la valeur du mode comme suit :

Si $[e_{i-1}, e_i[$ est la classe modale alors

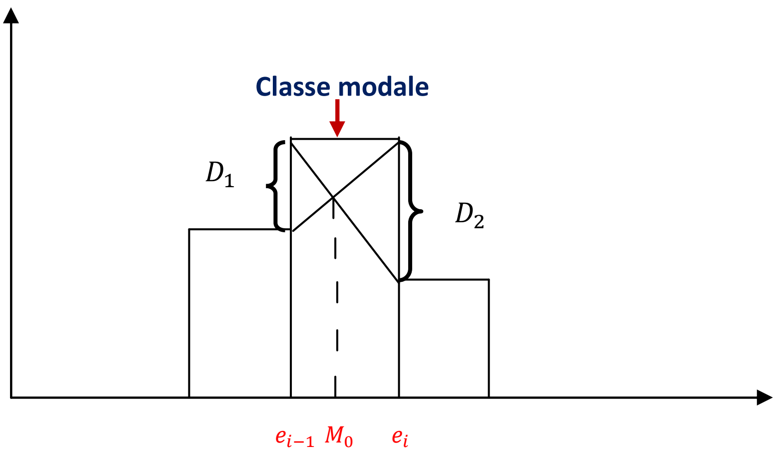
$$M_0 = e_{i-1} + \frac{D_1}{D_1 + D_2}(e_i - e_{i-1})$$

où

$D_1 = n_i - n_{i-1}$ est le surplus de la classe modale par rapport à la classe précédente,

$D_2 = n_i - n_{i+1}$ est le surplus de la classe modale par rapport à la classe suivante.

Notons qu'on peut remplacer les effectifs par les fréquences.



En effet, d'après le théorème de Thales, on a

$$\frac{e_i - M_0}{D_2} = \frac{M_0 - e_{i-1}}{D_1}$$

$$(e_i - M_0)D_1 = (M_0 - e_{i-1})D_2$$

$$e_i D_1 + e_{i-1} D_2 = M_0(D_1 + D_2).$$

Ce qui donne

$$M_0 = \frac{e_i D_1 + e_{i-1}(D_2 + D_1 - D_1)}{D_2 + D_1}$$

$$= e_i + \frac{D_1}{D_2 + D_1}(e_i - e_{i-1})$$

$$= e_{i-1} + \frac{D_1}{D_2 + D_1} a.$$

Exemple : Age des abonnés

La classe modale est $[30{,}35[$ (classe 4) qui correspond au plus grand effectif observé $n_4 = 272$. Les surplus de la classe modale par rapport aux classes précédente et suivante sont respectivement

$$D_1 = n_4 - n_3 = 272 - 168 = 104 \text{ et } D_2 = n_4 - n_5 = 272 - 176 = 96.$$

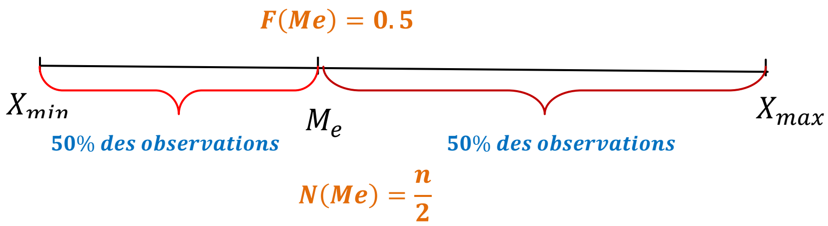
Par suite $M_0 = 30 + \frac{104}{104+96} 5 = 30 + \frac{104}{200} 5 = 32.6.$

Remarque : On peut prendre comme mode le centre de la classe modale.

Remarque : On dit qu'une série est unimodale lorsqu'elle possède un seul mode (cas discret) ou une seule classe modale (cas continu). Dans le cas où la série posséderait plusieurs modes ou plusieurs classes modales, on dira que la série est plurimodale (bimodale pour 2 modes, trimodale pour 3 modes et etc.).

2. La médiane (notation Me)

La médiane est le nombre qui partage la série des valeurs observées en deux séries de même taille.



Pour déterminer la médiane il faut d'abord ranger la série par ordre croissant (ou décroissant). Soit $X_1, \dots, X_n$ la série observée et $X_{(1)}, \dots, X_{(n)}$ la série ordonnée par ordre croissant.

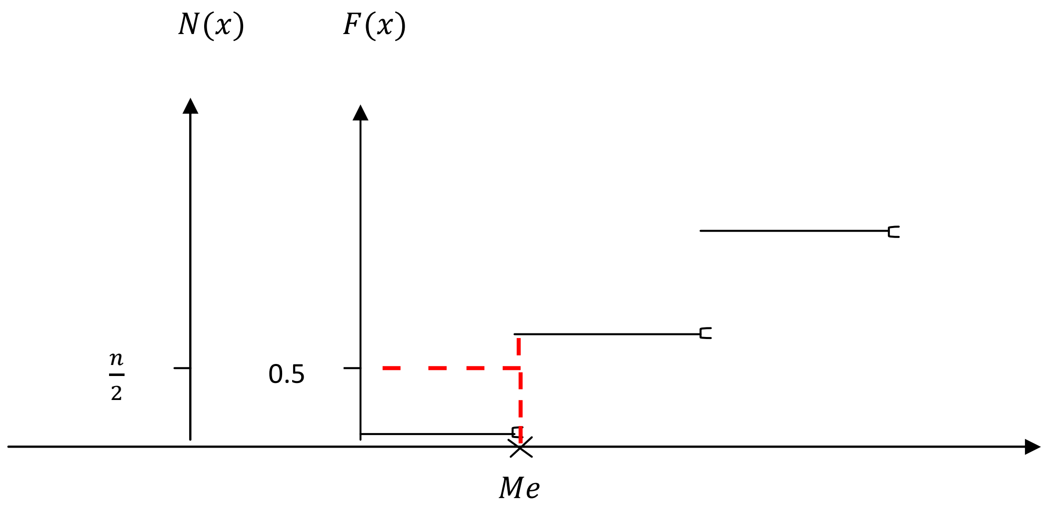
- **Cas discret** :

$$Me = \begin{cases} \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & si \ \dfrac{n}{2} \in \mathbb{N}^* \ (n \ pair) \\ X_{\left(\left[\frac{n}{2}\right]+1\right)} = X_{\left(\frac{n+1}{2}\right)} & si \ \dfrac{n}{2} \notin \mathbb{N}^* \ (\text{n impair}) \end{cases}$$

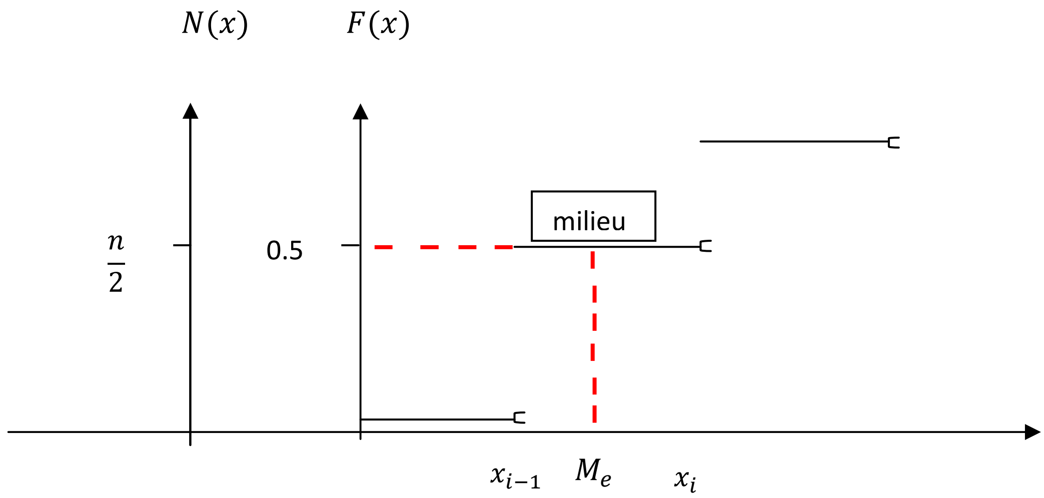
Dans le cas $n \ pair$, $[X_{\left(\frac{n}{2}\right)}, X_{\left(\frac{n}{2}+1\right)}]$ est appelé intervalle médian.

On peut déterminer graphiquement la médiane en utilisant la courbe cumulative. On distingue alors deux cas :

1er cas :



2ème cas :



$$Me = \frac{x_{i-1} + x_i}{2}$$

- **Cas continu :**
  - **Classe médiane** : C'est la classe qui contient la médiane.
    $[e_{i-1}, e_i[$ est la classe médiane $Me \in [e_{i-1}, e_i[)$ si $F_{i-1} < 0.5 \leq F_i$ (ou $N_{i-1} < \frac{n}{2} \leq N_i$).
  - La médiane est donnée par la formule

$$Me = e_{i-1} + \frac{0.5 - F_{i-1}}{f_i}(e_i - e_{i-1}) = e_{i-1} + a\frac{0.5 - F_{i-1}}{f_i}$$

$$Me = e_{i-1} + a\,\frac{\frac{n}{2} - N_{i-1}}{n_i}.$$

a **amplitude de la classe médiane,**

$\boldsymbol{f_i}$ **fréquence de la classe médiane,**

$\boldsymbol{n_i}$ **effectif de la classe médiane,**

$\boldsymbol{F_i}$ **fréquence cumulée de la classe médiane,**

$\boldsymbol{F_{i-1}}$ **fréquence cumulée de la classe qui précède la classe médiane,**

$\boldsymbol{N_{i-1}}$ **effectif cumulée de la classe qui précède la classe médiane.**

Sur $[e_{i-1}, e_i[$, on a

$$F(x) = F_{i-1} + \frac{x - e_{i-1}}{e_i - e_{i-1}} f_i.$$

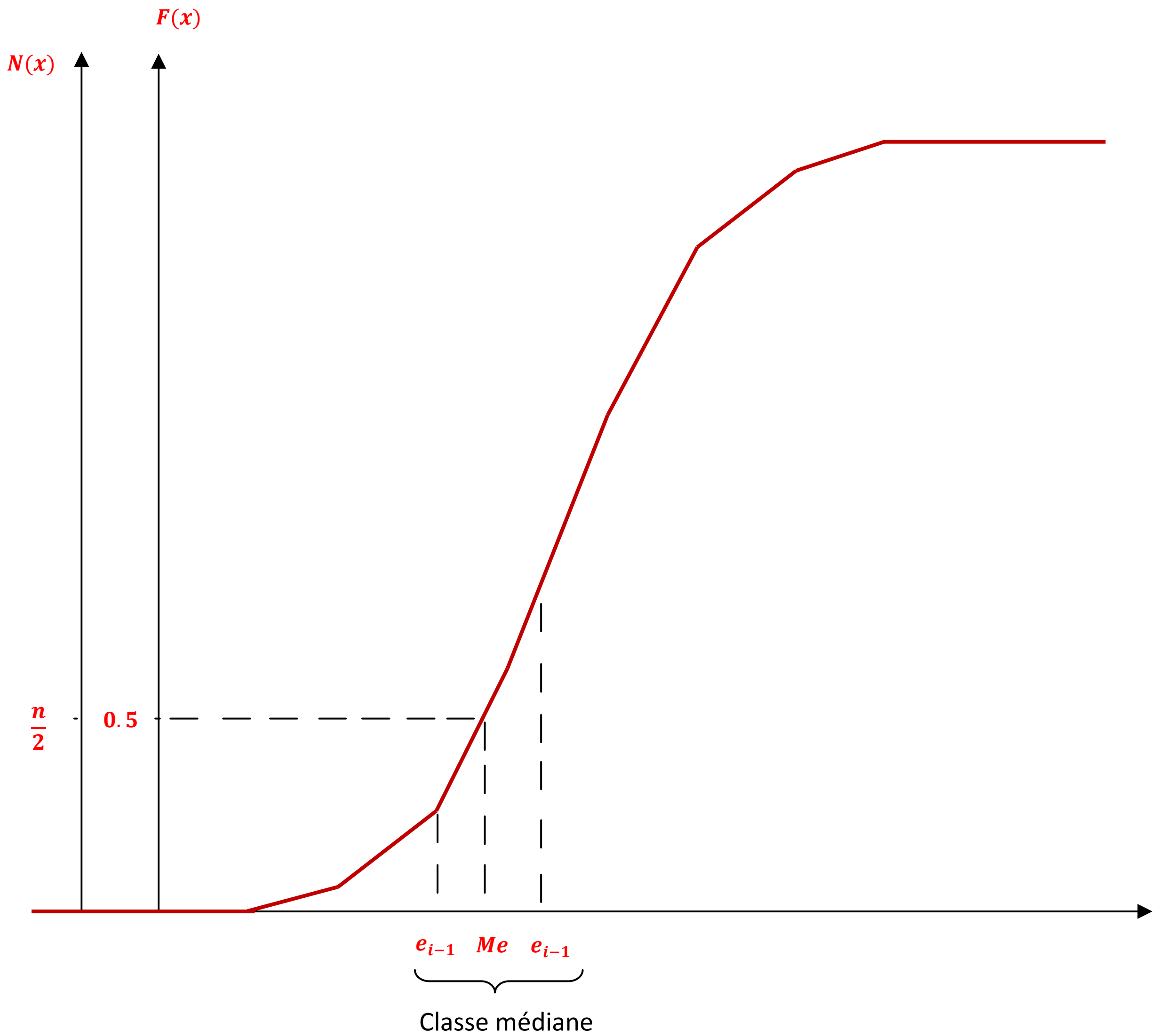
Pour $x = Me \in [e_{i-1}, e_i[$,

$$F(Me) = F_{i-1} + \frac{Me - e_{i-1}}{e_i - e_{i-1}} f_i = 0.5.$$

Par suite

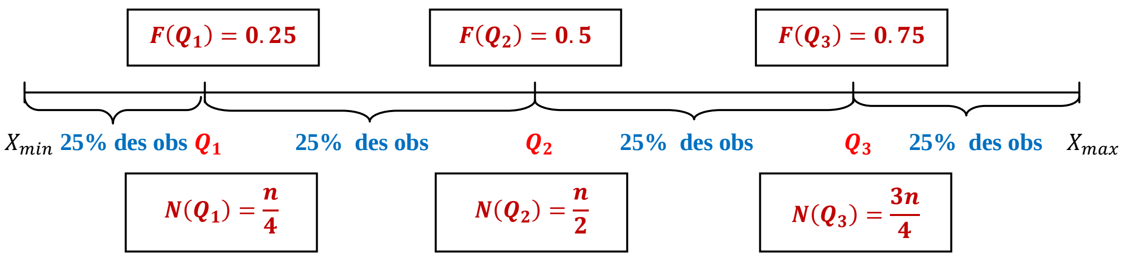
$$Me = e_{i-1} + \frac{0.5 - F_{i-1}}{f_i}(e_i - e_{i-1}).$$

On peut déterminer graphiquement la médiane en utilisant la courbe cumulative:



3. Les quartiles

Les quartiles sont 3 valeurs qui partagent la série en 4 groupes égaux : $Q_1,\ Q_2,\ Q_3$



L'intervalle $[Q_1, Q_3]$ s'appelle intervalle interquartile, il contient 50% des observations. Souvent l'intervalle interquartile est défini par la différence $Q_3 - Q_1$.

Les calculs de détermination des quartiles sont analogues à ceux de la médiane.

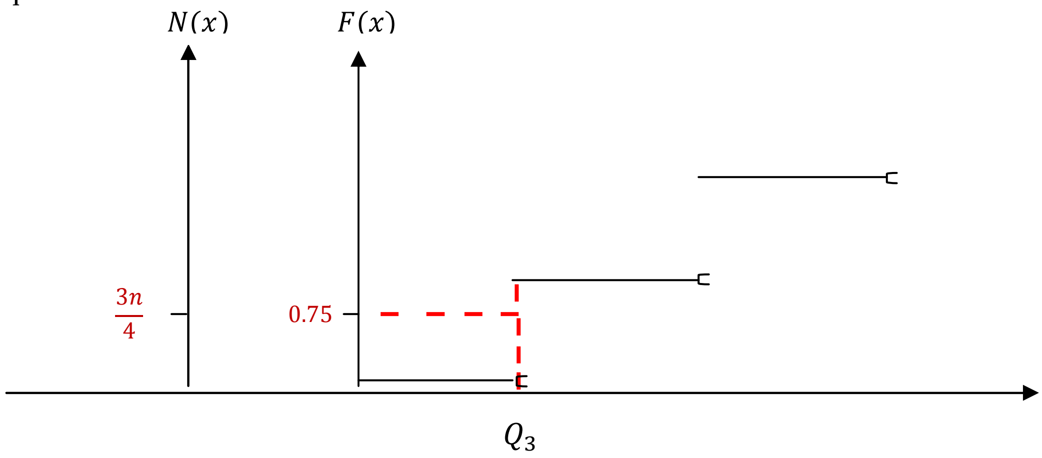
- **Cas discret :**

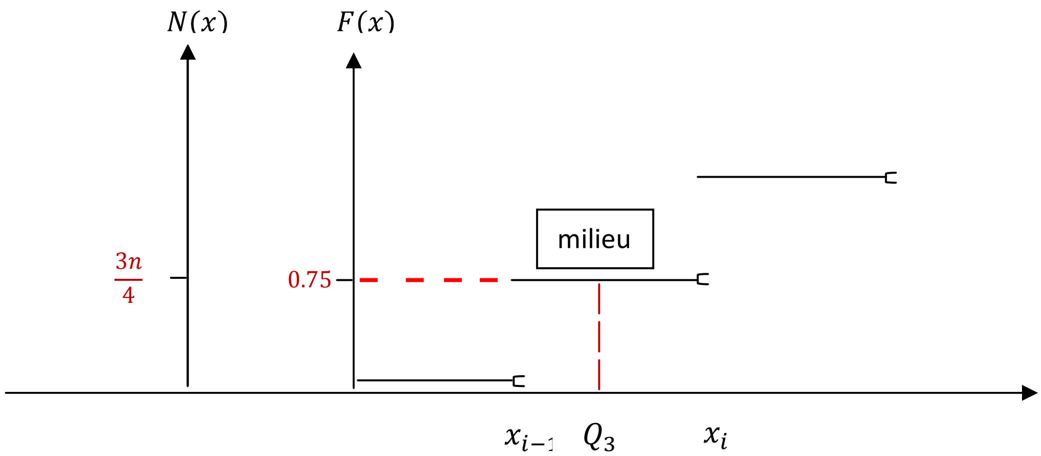
$$Q_1 = \begin{cases} \dfrac{X_{\left(\frac{n}{4}\right)} + X_{\left(\frac{n}{4}+1\right)}}{2} & si\ \dfrac{n}{4} \in \mathbb{N}^* \\ X_{\left(\left[\frac{n}{4}\right]+1\right)} & si\ \dfrac{n}{4} \notin \mathbb{N}^* \end{cases}$$

$$Q_2 = Me$$

$$Q_3 = \begin{cases} \dfrac{X_{\left(\frac{3n}{4}\right)} + X_{\left(\frac{3n}{4}+1\right)}}{2} & si\ \dfrac{3n}{4} \in \mathbb{N}^* \\ X_{\left(\left[\frac{3n}{4}\right]+1\right)} & si\ \dfrac{3n}{4} \notin \mathbb{N}^* \end{cases}$$

Graphiquement :





$$Q_3 = \frac{x_{i-1} + x_i}{2}$$

- **Cas continu**
  - ✓ $Q_1 \in [e_{i-1}, e_i[$ si $F_{i-1} < 0.25 \leq F_i$ (ou $N_{i-1} < \frac{n}{4} \leq N_i$)

$$Q_1 = e_{i-1} + \frac{0.25 - F_{i-1}}{f_i}(e_i - e_{i-1}) = e_{i-1} + \frac{\frac{n}{4} - N_{i-1}}{n_i}(e_i - e_{i-1}).$$

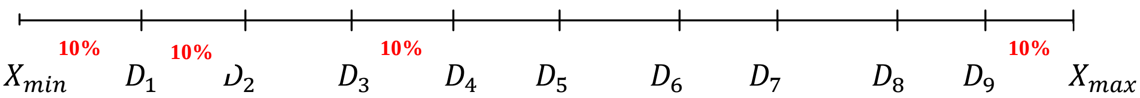
  - ✓ $Q_3 \in [e_{i-1}, e_i[$ si $F_{i-1} < 0.75 \leq F_i$ (ou $N_{i-1} < \frac{3n}{4} \leq N_i$)

$$Q_3 = e_{i-1} + \frac{0.75 - F_{i-1}}{f_i}(e_i - e_{i-1}) = e_{i-1} + \frac{\frac{3n}{4} - N_{i-1}}{n_i}(e_i - e_{i-1}).$$

Graphiquement : On procède de la même manière que la médiane. On remplace 0.5 ($\frac{n}{2}$) par 0.25 ($\frac{n}{4}$) pour le 1er quartile et par 0.75 ($\frac{3n}{4}$) pour le 3ème quartile.

4. Les déciles

Les déciles sont 9 valeurs $D_1, \ldots, D_9$ qui partagent les valeurs observées en 10 sous-ensembles égaux. Les intervalles qu'ils définissent contiennent chacun 10% des observations. L'intervalle $[D_1, D_9]$ $(D_9 - D_1)$ s'appelle intervalle interdécile, il contient 80% des observations.



Par exemple le premier décile D_1 est caractérisé par 10% des observations avant et 90% des observations après. Le cinquième décile D_5 n'est autre que la médiane ($D_5 = Me = Q_2$). Le troisième décile D_3 laisse 30% des observations avant et 70% après.

$$F(D_1) = 0.1 \qquad N(D_1) = \frac{n}{10}$$

$$F(D_2) = 0.2 \quad ou \quad N(D_2) = \frac{2n}{10}$$

$$\vdots \qquad\qquad \vdots$$

$$F(D_9) = 0.9 \qquad N(D_9) = \frac{9n}{10}$$

5. Les centiles

Les centiles sont 99 valeurs $C_1, \ldots, C_{99}$ qui partagent la série en 100 groupes égaux. Les intervalles qu'ils définissent contiennent chacun 1% des observations. L'intervalle intercentile est $[C_1, C_{99}]$ $(C_{99} - C_1)$, il contient 98% des observations. Le centile C_{50} n'est autre que la médiane.

$$F(C_1) = 0.01 \qquad\qquad N(C_1) = \frac{n}{100}$$

$$F(C_2) = 0.02 \qquad ou \qquad N(C_2) = \frac{2n}{100}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$F(C_{99}) = 0.99 \qquad\qquad N(C_{99}) = \frac{99n}{10}$$

6. Les quantiles : Soit $\alpha \in]0,1[$

On appelle quantile d'ordre α la valeur notée q_α telle que $n\alpha$ valeurs de la série sont $\leq q_\alpha$ et $n(1-\alpha)$ valeurs sont $> q_\alpha$.

$F(q_\alpha) = \alpha$

$100\,\alpha\,\%$ des obs q_α **$100(1-\alpha)$ des obs**

$N(q_\alpha) = n\alpha$

➢ **Cas discret :**

$$q_\alpha = \begin{cases} \dfrac{X_{(n\alpha)} + X_{(n\alpha+1)}}{2} & si \quad n\alpha \in \mathbb{N}^*, \\ X_{([n\alpha]+1)} & si \quad n\alpha \notin \mathbb{N}^*. \end{cases}$$

➢ **Cas continu :**

On détermine la classe $[e_{i-1}, e_i[$ qui contient le quantile q_α. C'est la 1ère classe dont l'effectif cumulé est $\geq n\alpha$ (ou fréquence cumulée$\geq \alpha$). La valeur de q_α est donnée alors par la formule

$$q_\alpha = e_{i-1} + \frac{\alpha - F_{i-1}}{f_i}(e_i - e_{i-1})$$

$$q_\alpha = e_{i-1} + \frac{n\alpha - N_{i-1}}{n_i}(e_i - e_{i-1}).$$

Graphiquement, à partir de la courbe cumulative, q_α est l'abscisse du point d'ordonnée α $(ou\ n\alpha)$.

Cas particuliers

Si $\alpha = 0.25$ $alors$ $q_{0.25} = Q_1 = C_{25}$,
Si $\alpha = 0.50$ $alors$ $q_{0.5} = Me = D_5 = C_{50}$,
Si $\alpha = 0.75$ $alors$ $q_{0.75} = Q_3 = C_{75}$.

- ✓ Le $i^{\text{ème}}$ quartile Q_i est le quantile d'ordre $\frac{i}{4}$, $i = 1,2,3$.
- ✓ Le $i^{\text{ème}}$ décile D_i est le quantile d'ordre $\frac{i}{10}$, $i = 1, \dots, 9$.
- ✓ Le $i^{\text{ème}}$ centile C_i est le quantile d'ordre $\frac{i}{100}$, $i = 1, \dots, 99$.

Exemple : Nombre d'absences

- Médiane : On a $n = 32$ est pair, donc
$$Me = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} = \frac{X_{(16)} + X_{(17)}}{2} = \frac{2+2}{2} = 2.$$
- 1^{er} quartile : On a $\frac{n}{4} = 8 \in \mathbb{N}^*$, donc
$$\mathrm{Q}_1 = \frac{X_{\left(\frac{n}{4}\right)} + X_{\left(\frac{n}{4}+1\right)}}{2} = \frac{X_{(8)} + X_{(9)}}{2} = \frac{1 + 1}{2} = 1.$$
- $3^{\text{ème}}$ quartile : On a $\frac{3n}{4} = 24 \in \mathbb{N}^*$, donc
$$\mathrm{Q}_3 = \frac{X_{\left(\frac{3n}{4}\right)} + X_{\left(\frac{3n}{4}+1\right)}}{2} = \frac{X_{(24)} + X_{(25)}}{2} = \frac{2 + 2}{2} = 2.$$
- $4^{\text{ème}}$ décile : $0.4n = 12.8 \notin \mathbb{N}^*$ donc
$$\mathrm{D}_4 = \mathrm{q}_{0.4} = \mathrm{C}_{40} = X_{([0.4n]+1)} = X_{(13)} = 1.$$
- $90^{\text{ème}}$ centile : $0.9n = 28.8 \notin \mathbb{N}^*$ donc
$$\mathrm{C}_{90} = \mathrm{q}_{0.9} = X_{([0.9n]+1)} = X_{(29)} = 3.$$

Exemple : Age des abonnés

- Médiane : $Me \in [30,35[$ car $N_3(= 264) < \frac{n}{2}(= 400) < N_4(= 536)$.
$$Me = 30 + \frac{400 - 264}{272} 5 = 32.5.$$
- 1^{er} quartile : $\mathrm{Q}_1 \in [25,30[$ car $N_2(= 96) < \frac{n}{4}(= 200) < N_3(= 264)$.
$$\mathrm{Q}_1 = 25 + \frac{200 - 96}{168} 5 = 28.09.$$
- $3^{\text{ème}}$ quartile : $\mathrm{Q}_3 \in [35,40[$ car $N_4(= 536) < \frac{3n}{4}(= 600) < N_5(= 712)$.
$$\mathrm{Q}_3 = 35 + \frac{600 - 536}{176} 5 = 36.81.$$
- $\mathrm{C}_2 \in [15,20[$ car $0.02n(= 16) < N_1(24)$
$$\mathrm{C}_2 = 15 + \frac{16 - 0}{24} 5 = 18.33.$$

7. Les moyennes

7.1. Moyenne arithmétique (notation $\bar{x}$)

- La moyenne arithmétique $\bar{x}$ d'une série statistique brute $X_1, \dots, X_n$ est donnée par
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
Cette moyenne est dite simple si les observations X_i sont distinctes.

- Soit $(x_i, n_i)_{i=\overline{1,k}}$ une distribution de fréquences dans le cas d'une variable discrète. On appelle moyenne arithmétique, le nombre $\bar{x}$ défini par
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i\, x_i \;\; avec \;\; n = \sum_{i=1}^{k} n_i \; et \; f_i = \frac{n_i}{n} \; .$$
- Soit $([e_{i-1}, e_i[, n_i)_{i=\overline{1,k}}$ une distribution de fréquences dans le cas d'une variable continue. On appelle moyenne arithmétique le nombre $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i\, x_i = \sum_{i=1}^{k} f_i\, x_i$ où x_i est le centre de la $i$ème classe.

Dans les deux derniers cas, la moyenne arithmétique est dite pondérée.

Exemple

Un automobiliste roule pendant une heure à la vitesse constante de 90 km/h, puis pendant encore une heure à la vitesse constante de 120 km/h. Déterminer à quelle vitesse constante il aurait dû rouler pendant la durée totale du trajet pour effectuer le même nombre de kilomètres.

L'automobiliste roule $d_1 = v_1 t_1 = 90\ km$ pendant la première heure et $d_2 = v_2 t_2 = 120\ km$ pendant la deuxième heure. La distance parcourue pendant les deux heures est donc $d = d_1 + d_2 = 210\ km$. La vitesse moyenne v vérifie
$$v = \frac{d}{t} = \frac{v_1 t_1 + v_2 t_2}{t_1 + t_2} = \frac{90 + 120}{2} = 105.$$
Exemple

Pendant 4 années, une personne achète du fuel à 16, 18, 21 et 25 DA par litre. Quel est le prix moyen du fuel au cours des 4 années s'il utilise la même quantité du fuel chaque année.

Soient x la quantité de fuel utilisée chaque année, p_i la somme d'argent dépensée la $i$ème année, $i = 1,2,3,4$. On a

$$\begin{cases} p_1 = 16x, \\ p_2 = 18x, \\ p_3 = 21x, \\ p_4 = 25x. \end{cases}$$

Le prix moyen du fuel p_{moy} au cours des 4 années est donné par

$$p_{moy} = \frac{somme\ totale\ d'argent\ dépensée}{quantité\ totale\ de\ fuel\ utilisée} = \frac{p_1 + p_2 + p_3 + p_4}{4x} = \frac{16x + 18x + 21x + 25x}{4x}$$

$$p_{moy} = \frac{16 + 18 + 21 + 25}{4} = 20.$$

Propriétés de la moyenne arithmétique

1. $\sum_{i=1}^{k} n_i(x_i - \bar{x}) = 0.$
2. $\sum_{i=1}^{k} n_i(x_i - a)^2$ est minimale pour $a = \bar{x}$.
3. Si $y_i = ax_i + b, i = \overline{1,k}$ alors $\bar{y} = a\bar{x} + b$ (propriété de linéarité).

Preuve

1. $\sum_{i=1}^{k} n_i(x_i - \bar{x}) = \sum_{i=1}^{k}(n_i x_i - n_i\bar{x}) = \sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i\bar{x} = \sum_{i=1}^{k} n_i x_i - \bar{x}\sum_{i=1}^{k} n_i$
   Par définition, $\sum_{i=1}^{k} n_i = n$. De plus, $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$ ou de manière équivalente $\sum_{i=1}^{k} n_i x_i = n\bar{x}$.
   Il vient donc
   $$\sum_{i=1}^{k} n_i(x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0.$$
2. TD.
3. $\bar{y} = \frac{1}{n}\sum_{i=1}^{k} n_i y_i = \frac{1}{n}\sum_{i=1}^{k} n_i(ax_i + b) = \frac{1}{n}\sum_{i=1}^{k} a n_i x_i + \frac{1}{n}\sum_{i=1}^{k} n_i b = a\frac{1}{n}\sum_{i=1}^{k} n_i x_i + b\frac{1}{n}\sum_{i=1}^{k} n_i$
   $$\bar{y} = a\bar{x} + b.$$

7.2. Moyenne géométrique (notation G, $m_g, \bar{x}_g$)

La moyenne géométrique est utilisée dans le calcul de taux d'accroissement moyens ou de moyennes de coefficients multiplicateur.

- Soient $X_1, \dots, X_n$ n nombres réels strictement positifs. La moyenne géométrique (simple) $\bar{\boldsymbol{x}}_{\boldsymbol{g}}$ de ces nombres est donnée par
  $$\bar{\boldsymbol{x}}_{\boldsymbol{g}} = \sqrt[n]{X_1 \dots X_n} = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}$$
  ou de manière équivalente
  $$log(\bar{x}_g) = \frac{1}{n}\sum_{i=1}^{n} log(X_i).$$
  $log(\bar{x}_g)$ est la moyenne arithmétique des $log(X_i)$.

- Soit $(x_i, n_i)_{i=\overline{1,k}}$ une distribution de fréquences dans le cas d'une variable discrète strictement positive. Dans ce cas la moyenne géométrique (dite pondérée) est donnée par
  $$\bar{\boldsymbol{x}}_{\boldsymbol{g}} = \sqrt[n]{{x_1}^{n_1} \dots {x_k}^{n_k}} = \left(\prod_{i=1}^{k} {x_i}^{n_i}\right)^{\frac{1}{n}} = \prod_{i=1}^{k} {x_i}^{f_i}$$

  ou de manière équivalente

$$log(\bar{x}_g) = \frac{1}{n}\sum_{i=1}^{k} n_i log(x_i) = \sum_{i=1}^{k} f_i log(x_i)\,.$$

- Soit $([e_{i-1}, e_i[, n_i)_{i=\overline{1,k}}$ une distribution de fréquences dans le cas d'une variable continue strictement positive. On appelle moyenne géométrique le nombre $\bar{x}_g$ défini par

$$\boldsymbol{\bar{x}_g} = \sqrt[n]{{x_1}^{n_1} \dots {x_k}^{n_k}} = \left(\prod_{i=1}^{k} {x_i}^{n_i}\right)^{\frac{1}{n}} = \prod_{i=1}^{k} {x_i}^{f_i}$$

où x_i est le centre de la $i$ème classe.

Exemple

Dans un certain pays, au mois de janvier d'une certaine année, les prix ont augmenté de 0,9 %, puis en février de 1,2 %. Déterminer l'augmentation mensuelle constante qu'il y aurait dû avoir pendant les deux mois pour obtenir le même résultat à l'issue des deux mois.

Soient P_0 le prix initial, P_1 le prix au mois de janvier et P_2 le prix au mois de février ;

$$P_1 = P_0\left(1 + \frac{0.9}{100}\right) = P_0(1.009), P_2 = P_1\left(1 + \frac{1.2}{100}\right) = P_0\left(1 + \frac{0.9}{100}\right)\left(1 + \frac{1.2}{100}\right) = P_0(1.009)(1.012).$$

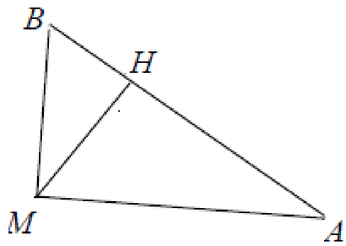
On cherche t (le pourcentage d'augmentation constant sur les deux mois) tel que

$$P_1 = P_0\left(1 + \frac{t}{100}\right) \; et \; P_2 = P_1\left(1 + \frac{t}{100}\right) = P_0\left(1 + \frac{t}{100}\right)^2.$$

Le nombre t satisfait l'équation $\left(1 + \frac{t}{100}\right)^2 = (1.009)(1.012)$, c.à.d. $t = \sqrt{(1.009)(1.012)} - 1 = 0.01049$.
Donc l'augmentation moyenne sur les deux mois est d'environ 1.49 %.

Exemple

Dans un triangle *AMB* rectangle en *M*, *H* est le pied de la hauteur issue de *M*. On pose $AH = a$ et $BH = b$.
Exprimer ***MH en fonction de a et b.***



D'après le théorème de Pythagore, on a

$$\begin{cases} AB^2 = AM^2 + MB^2 \\ MB^2 = BH^2 + MH^2 = b^2 + MH^2 \\ AM^2 = AH^2 + MH^2 = a^2 + MH^2 \end{cases}$$

Ce qui donne $AB^2 = a^2 + b^2 + 2MH^2$.
Comme $AB = AH + BH = a + b$ alors $MH^2 = ab$ $(MH = \sqrt{ab}\,)$.

Propriétés de la moyenne géométrique

Soient $X_1, \dots, X_n$ et $Y_1, \dots, Y_n$ deux séries de moyennes géométriques respectives G_X et G_Y.

1. La moyenne géométrique de la série $Z_1, \dots, Z_n$, définie par $Z_i = X_i Y_i$, est donnée par

$$G_Z = G_X G_Y.$$

2. La moyenne géométrique de la série $T_1, \dots, T_n$, définie par $T_i = \frac{X_i}{Y_i}$, est donnée par

$$G_T = \frac{G_X}{G_Y}.$$

7.3. Moyenne Harmonique (notation H, $m_h, \bar{x}_h$)

La moyenne harmonique est employée dans le calcul des moyennes des pourcentages ou de rapports, et notamment dans celui des durées moyennes et des vitesses moyennes. L'inverse de la moyenne harmonique est la moyenne arithmétique de l'inverse des valeurs de la variable (non nulle) considérée.

- La moyenne harmonique $\bar{\boldsymbol{x}}_{\boldsymbol{h}}$ d'une série statistique brute $X_1, \dots, X_n$ est donnée par

$$\bar{x}_h = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{X_i}}$$

ou de manière équivalente

$$\frac{\mathbf{1}}{\bar{\boldsymbol{x}}_{\boldsymbol{h}}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}.$$

- Moyenne harmonique pondérée

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

ou

$$\frac{\mathbf{1}}{\bar{\boldsymbol{x}}_{\boldsymbol{h}}} = \frac{1}{n}\sum_{i=1}^{k}\frac{n_i}{x_i} = \sum_{i=1}^{k}\frac{f_i}{x_i}\ avec\ n = \sum_{i=1}^{k} n_i\ et\ f_i = \frac{n_i}{n}\ .$$

Exemple

Un automobiliste roule 100 km à la vitesse constante de 90 km/h, puis encore 100 km à la vitesse constante de 120 km/h. Déterminer à quelle vitesse constante il aurait dû rouler sur la distance totale pour que la durée du voyage soit identique.
La durée de la première partie du voyage est de $t_1 = \frac{100}{90}\ h$ et celle de la deuxième partie est de $t_2 = \frac{100}{120}\ h$.
Comme la distance parcourue est $d = 200\ km$ alors la vitesse moyenne v vérifie l'équation

$$v = \frac{d}{t} = \frac{200}{\frac{100}{90} + \frac{100}{120}} = \frac{2}{\frac{1}{90} + \frac{1}{120}} = 102.85\ \ km/h.$$

Exemple

Pendant 4 années, une personne achète du fuel à 16, 18, 21 et 25 DA par litre. Quel est le prix moyen du fuel au cours des 4 années s'il dépense la même somme d'argent chaque année.

Soient p la somme d'argent dépensée chaque année, x_i la quantité de fuel utilisée la $i$ème année, $i = 1,2,3,4$. On a

$$\begin{cases} p = 16x_1, \\ p = 18x_2, \\ p = 21x_3, \\ p = 25x_4. \end{cases}$$

Le prix moyen du fuel p_{moy} au cours des 4 années est donné par

$$p_{moy} = \frac{somme\ totale\ d'argent\ dépensée}{quantité\ totale\ de\ fuel\ utilisée} = \frac{4p}{x_1 + x_2 + x_3 + x_4} = \frac{4p}{\frac{p}{16} + \frac{p}{18} + \frac{p}{21} + \frac{p}{25}}$$

$$= \frac{4}{\frac{1}{16} + \frac{1}{18} + \frac{1}{21} + \frac{1}{25}} = 19.44.$$

7.4. Moyenne quadratique (notation Q, $m_q, \bar{x}_q$)

La moyenne quadratique s'emploie surtout pour des moyennes d'écarts à une valeur centrale. Elle permet de ne pas avoir à manipuler des écarts négatifs.

- Moyenne quadratique simple :

$$\bar{\boldsymbol{x}}_{\boldsymbol{q}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{\frac{1}{2}}$$

ou de manière équivalente

$$\bar{\boldsymbol{x}}_{\boldsymbol{q}}^{\boldsymbol{2}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

- Moyenne quadratique pondérée :

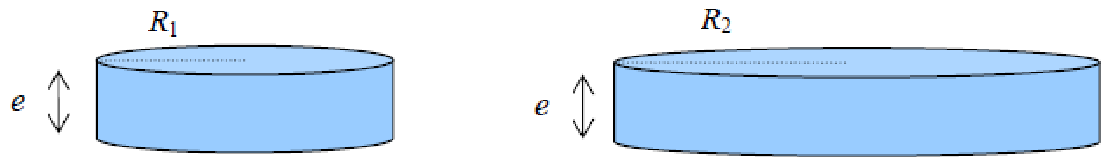
$$\bar{\boldsymbol{x}}_{\boldsymbol{q}} = \left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{k} f_i x_i^2\right)^{\frac{1}{2}}$$

ou

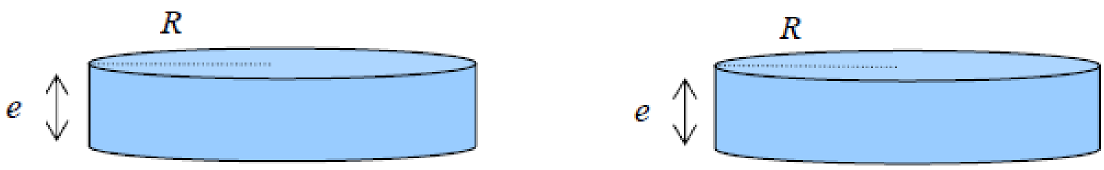
$$\bar{\boldsymbol{x}}_{\boldsymbol{q}}^{\boldsymbol{2}} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 \quad avec\ n = \sum_{i=1}^{k} n_i\ et\ f_i = \frac{n_i}{n}.$$

Exemple

On possède deux plaques métalliques de formes cylindriques, de même épaisseur e, mais de rayons différents qu'on note R_1, R_2 .



Nous souhaitons fondre ces deux plaques pour en fabriquer deux autres, de même épaisseur e, et de même rayon R :



Le volume d'un cylindre est donné par $V = \pi R^2 e$. Le rayon R est solution de l'équation

$$2\pi R^2 e = \pi {R_1}^2 e + \pi {R_2}^2 e.$$

Ce qui donne $R^2 = {R_1}^2 + {R_2}^2$ ou $R = \sqrt{{R_1}^2 + {R_2}^2}$.
Pour $R_1 = 18\ cm$ et $R_2 = 24\ cm$, on a $R = \sqrt{900} = 30\ cm$.

7.5. Moyenne d'ordre r (notation $\mathcal{M}_r$)

- Moyenne simple d'ordre r :

$$\mathcal{M}_r = \sqrt[r]{\frac{1}{n}\sum_{i=1}^{n} {X_i}^r} = \left(\frac{1}{n}\sum_{i=1}^{n} {X_i}^r\right)^{\frac{1}{r}}$$

- Moyenne pondérée d'ordre r :

$$\mathcal{M}_r = \left(\frac{1}{n}\sum_{i=1}^{k} n_i {x_i}^r\right)^{\frac{1}{r}} = \left(\sum_{i=1}^{k} f_i {x_i}^r\right)^{\frac{1}{r}} \quad avec\ n = \sum_{i=1}^{k} n_i\ et\ f_i = \frac{n_i}{n}\ .$$

Selon la valeur de r, on retrouve les formules des différentes moyennes $\bar{x}$, $\bar{x}_g$, $\bar{x}_h$, et $\bar{x}_q$:

$$\begin{cases} Si\ \boldsymbol{r = 1\ alors\ \mathcal{M}_1} = \bar{x}\ , \\ Si\ \boldsymbol{r = 2\ alors\ \mathcal{M}_2} = \bar{x}_q, \\ Si\ \boldsymbol{r = -1\ alors\ \mathcal{M}_{-1}} = \bar{x}_h, \\ \lim\limits_{r \to 0} \boldsymbol{\mathcal{M}_r} = \bar{x}_g. \end{cases}$$

Comparaison des moyennes

Pour $X_1, \dots, X_n$ strictement positifs, les différentes moyennes définies précédemment sont ordonnées comme suit :

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_q.$$

(Il y a égalité entre les moyennes lorsque les nombres X_i sont égaux)

Plus généralement, $\boldsymbol{\mathcal{M}_{r_1} \leq \mathcal{M}_{r_2}}$ **si** $\boldsymbol{r_1 < r_2}$ **(avec égalité dans le cas d'observations égales).**

7.6. La φ −moyenne

Soit $\varphi: \mathbb{R} \to \mathbb{R}$ une fonction. On appelle φ −moyenne de la série $X_1, \dots, X_n$ la moyenne arithmétique de la série $\varphi(X_1), \dots, \varphi(X_n)$: $\frac{1}{n}\sum_{i=1}^{n} \varphi(X_i)$.

Pour des données groupées, la φ −moyenne est donnée par $\frac{1}{n}\sum_{i=1}^{k} n_i \varphi(x_i)$.

- Si $\varphi(x) = x$, on retrouve la moyenne arithmétique,
- Si $\varphi(x) = x^2$, on retrouve la moyenne quadratique,
- Si $\varphi(x) = lnx$, on retrouve la moyenne géométrique,
- Si $\varphi(x) = \frac{1}{x}$, on retrouve la moyenne harmonique,
- Si $\varphi(x) = x^r$, on retrouve la moyenne d'ordre r.