Cas des variables continues :

Classe de valeurs :

En combien de classes partageons-nous les valeurs ? la réponse n'est pas unique. Soit N l'effectif total. Nous pouvons considérer dans ce cours trois réponses à titre d'exemple.

1. Un réponse $\sqrt{N}$, $[\sqrt{N}]$ (la partie entière) ou encore $[\sqrt{N}]+1$. Donc le nombre de classes $k \simeq \sqrt{N}$
2. Une réponse la formule de **STURGES**
$$k = \frac{X_{max} - X_{min}}{1 + 3.3\log_{10} N}.$$
3. Une réponse la formule de **YULE**
$$k = 2,5\sqrt[4]{N}$$

Remarques :

1. Pour le nombre de classes que comptera notre distribution de fréquences, y a 3 cas possibles :
   (i) Il peut être choisi aribtrairement.
   (ii) Il peut être imposé.
   (iii) Il peut être fixé par une méthode mathématique (règle de **STURGES**) .
2. Si les classes sont d'amplitudes inégales, il faut d'abord réctifier l'effectif avant de tracer l'histogramme, pour celà on doit multiplier chaque classe par la rapport : $\dfrac{\text{amplitude la plus petite}}{\text{amplitude de la classe}}$.

Valeurs centrales des variables centrées :

1. Moment centré d'ordre k noté μ_k est donné par
$$\mu_k(x) = \frac{1}{N}\sum_{i=1}^{n} n_i\big(x_i - \overline{x}\big)^k.$$
2. Moment d'ordre k noté m_k est donné par
$$m_k(x) = \frac{1}{N}\sum_{i=1}^{n} n_i x_i^k.$$

On remrque que :
$$\mu_2(x) = Var(x) = \frac{1}{N}\sum_{i=1}^{n} n_i\big(x_i - \overline{x}\big)^2 = \frac{1}{N}\sum_{i=1}^{n} n_i x_i^2 - \overline{x}^2 = m_2 - m_1^2$$
$$\mu_3(x) = m_3 - 3m_1m_2 + 2m_1^3$$
$$\mu_4(x) = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

Une mesure de la dispersion relative : On l'appelle aussi coefficient de dispersion relative ou le coefficient de variation et on le note par CV, il permet d'apprécier la représentativité de la moyenne par rapport à l'ensemble des observations et de comparer la dispersion dans différentes série. Il donne une bonne idée du degré d'homogénéité d'une série. Il faut qu'il soit le plus faible possible ($< 15\%$ en pratique).
$$CV = \frac{\sigma}{\overline{x}} = \frac{\sqrt{Var(x)}}{\overline{x}}$$

Le coefficient interquartile De par la définition des quartiles, l'intervalle interquartile $[Q1, Q3]$ contient 50% des observations. Sa longueur, notée $e_Q = Q_3 - Q_1$ (Étendue InterQuartile), est un indicateur de dispersion, à partir de là on défini le coefficient interQuartile noté C_Q donné par
$$C_Q = \frac{Q_3 - Q_1}{Me} \times 100$$

D'autres moyennes

1-**La moyenne géométrique :** C'est la moyenne applicable à des mesures de grandeurs dont la croissance est géométrique ou exponentielle. La moyenne géométrique conserve le produit des x_i : si on modifie les valeurs de deux observations tout en conservant leur produit, la moyenne géométrique sera inchangée.

La moyenne géométrique G de la série de valeurs $x_1, \ldots, x_i, \ldots, x_k$ supposées toutes positives (strictement), est définie ainsi lorsque chaque valeur x_i n'intervient qu'une seule fois :

$$G = \sqrt[N]{\prod_{i=1}^{k} x_i} = \sqrt[N]{x_1 \times x_2 \times \ldots \times x_k} \Longrightarrow \ln G = \frac{1}{N} \sum_{i=1}^{k} \ln x_i$$

Lorsque la distribution de la variable statistique est donnée par les k couples (x_i, n_i) (moyenne géométrique pndérée), les x_i étant tous positifs ; la moyenne géométrique a pour expression :

$$G = \sqrt[N]{\prod_{i=1}^{k} x_i^{n_i}} = \prod_{i=1}^{k} x_i^{f_i} = \sqrt[N]{x_1^{n_1} \times x_2^{n_2} \times \ldots \times x_k^{n_k}} \Longrightarrow \ln G = \frac{1}{N} \sum_{i=1}^{k} n_i \ln x_i = \sum_{i=1}^{k} f_i \ln x_i$$

Exemple (i)- Supposons que pendant une décennie, les salaires aient été multipliés par 2 et que pendant la décennie suivante, ils aient été multipliés par 4 ; le coefficient multiplicateur moyen par décennie est égal à :

$$\sqrt{2.4} = \sqrt{8} \simeq 2,83$$

La moyenne arithmétique ($\overline{x} = 3$) n'est pas égale au coefficient demandé.

Prenons, par exemple, un salaire de $300DA$ au début de la première décennie, il sera de $300.2.4 = 2400DA$ au bout des vingt ans, ce qui équivaut à $300.(2,83)^2$, soit un coefficient multiplicateur moyen de $2,83$ par décennie.

(ii)- Au cours des 4 dernières années, le taux de croissance annuels de la production intérieur brut (PIB) ont été les suivants :

1^{re} Année $\longrightarrow +7,2\%$
2^{me} Année $\longrightarrow +6,3\%$
3^{me} Année $\longrightarrow +7,0\%$
4^{me} Année $\longrightarrow +4,8\%$

- Quel est le taux moyen de croissance de la PIB au cours de ces 4 années ?

$G = \sqrt[4]{107,2 \times 106,3 \times 107 \times 104,8} \Longrightarrow \ln G = \frac{1}{4} (\ln 107,2 + \ln 106,3 + \ln 107 + \ln 104,8) = 2,027$ donc $G = 10^{2,027} = 106,3\%$.

Au cours de cette période, le taux moyen de croissance est de $6,3\%$.

(iii)- 3 équipes se sont succédées à la direction d'une entreprise pendant la 1-ère période qui a durée 4 ans les bénifices réalisés ont augmenté de 50% par an, pendant la seconde période de 3 ans de 17% par an et pendant la dernière période de 2 ans, les bénifices ont enregistré une baisse de 30% par an.

- Quel est le taux de croissance annuel moyen des bénifices réalisé au cours de ces 9 années ?

$G = \sqrt[9]{150^4 \times 117^3 \times 70^2} \Longrightarrow \ln G = \frac{1}{9} (4 \ln 150 + 3 \ln 117 + 2 \ln 70) = 2,06$ donc $G = 10^{2,06} = 114,8\%$.

L'augmentation moyenne annuelle est donc de $14,8\%$.

2- **La moyenne harmonique :** La moyenne harmonique est l'inverse de la moyenne arithmétique des inverses des valeurs. L'inverse de la moyenne harmonique conserve ainsi la somme des inverses des x_i : si on modifie les valeurs de deux observations tout en conservant la somme de leurs inverses, la moyenne harmonique sera inchangée.

$$H = \frac{N}{\sum_{i=1}^{k} \frac{1}{x_i}} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_k}}$$

La moyenne harmonique pondérée est donnée par

$$H = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}} = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots \frac{n_k}{x_k}}$$

La moyenne harmonique peut être utilisée lorqu'il est possible d'attribuer un sens réel aux inverses des données en particulier pour les taux de change, les taux d'équipement, le pouvoir d'achat, les vitesses. Elle est notamment utilisée dans les calculs d'indices .

Exemple (i)- On achète des dollars une première fois pour 100 DA au cours de 1,23 DA le dollar, une seconde fois pour 100 DA au cours de 0,97 DA le dollar. Le cours moyen du dollar pour l'ensemble de ces deux opérations est égal à :

$$\frac{200}{\frac{100}{1.23} + \frac{100}{0,97}} \simeq 1,085DA$$

La moyenne arithmétique (= 1,1) ne représente pas le cours moyen du dollar.

(ii)- Une entreprise de transport possède 3 camions qui effectues des rotations entre Alger-Oran. Au cours d'une celles-ci, le trajet Alger-Oran a été couvert aux vitesses moyennes suivantes : 40km/h, 60km/h et 80km/h.

- Quelle est la vitesse moyenne ?

$H = \dfrac{3}{\frac{1}{40} + \frac{1}{60} + \frac{1}{80}} = 56,6km/h.$

(iii)- Une entreprise de transport possède 10 camions qui font des rotations entre 2 villes. Au cours d'une celles-ci, le trajet a été couvert par ces véhicules aux vitesses moyennes présentées comme suit :

Vitesse moyenne km/h	40	60	70
nbre de camions	4	4	2

- Quelle est la vitesse moyenne globale ?

$H = \dfrac{10}{\frac{4}{40} + \frac{4}{60} + \frac{2}{70}} = 51,5km/h.$

Comparaison des 3 moyennes étudiées On montre que si les x_i sont tous positifs :

$$\min_{1\le i\le n} x_i \le H \le G \le \overline{x} \le \max_{1\le i\le n} x_i$$

L'égalité de deux de ces moyennes entre elles entraîne leur égalité dans leur ensemble, et dans ce cas, toutes les valeurs x_i sont égales.

Exemple Un étudiant à obtenu les notes suivantes : Math=14 (coefficient 3), Stat =8 (coefficient 2) et Physique=6 (coefficient 1).

- Calculer les moyennes pondérées.

1. $\overline{x} = \dfrac{7 \times 3 + 4 \times 2 + 3 \times 1}{6} = \dfrac{32}{6} = 5,33.$
2. $G = \sqrt[6]{7^3 \times 4^2 \times 3^1} = (343 \times 16 \times 3)^{1/6} = 16464^{1/6} = 5,01.$
   $\ln G = \dfrac{1}{6}(3\ln 7 + 2\ln 4 + \ln 3) = 0,7 \Longrightarrow G = 10^{0,7} = 5,01.$
3. $H = \dfrac{6}{\frac{3}{7} + \frac{2}{4} + \frac{1}{3}} = \dfrac{6}{1,26} = 4,76.$

Dans ce cas la moyenne arithmétique est la plus indiquée.

Caractéristiques de position

- Déciles divisent la population en 10 parties égales $(D_1, D_2, \dots, D_{10})$, les déterminations approchées sont du même type que celles indiquées pour les quartiles. On peut remarquer que $D_5 = Q_2 = Me$.

- Percentiles divisent la population en 100 parties égales $(P_1, P_2, \ldots, P_{100})$, les déterminations approchées sont du même type que celles indiquées pour les quartiles. On peut remarquer que $P_{50} = D_5 = Q_2 = Me$.

Remarque :

- L'intervalle interdéciles $D = D_9 - D_1$ contient 80% des observations en laissant 10% des observations à droite et à gauche. Autrement dit on élimine 10% des valeurs se trouvant aux extrémités des distributions.

Caractéristiques de forme

Les caractéristiques de forme permettent de préciser l'allure de la courbe des effectifs ou des fréquences sans besoin de les tracer par rapport à une courbe idéale dite **courbe normale**. Nous définissons 2 sortes de caractéristiques.

- Caractéristiques d'asymétrie.
- Caractéristiques d'aplatissement.

– **Asymétrie** Une distribution est symétrique si les 3 caractéristiques de tendance sont confondues (Figure 1)

(i) $Mo = Me = \overline{x}$.

(ii) Q_1 et Q_3 équidistant de la médiane (situés à égale distance).

(iii) D_1 et D_9 équidistant de la médiane (situés à égale distance).

La position respective de ces élélements va permettre le calcul de coefficient d'asymétrie, qui sont nul en cas de symétrie et s'éloignent plus de la valeur 0 que la distribution est plus asymétrique.

Les principaux coefficient d'asymétrie ne sont valable que si les séries statistiques sont unimodales (l'existence d'un mode unique).

1. Coefficient de Pearson noté β_1 et $\beta_2 = F_1^2$ est nul pour une distribution symétrique, négatif pour une distribution unimodale étalée vers la gauche, positif pour une distribution unimodale étalée vers la droite (Figure 1), donné par :

$$\beta_1 = \frac{\overline{x} - Mo}{\sigma}$$

(i) Si $\beta_1 = 0$ on a une distribution symétrique.

(ii) Si $\beta_1 > 0$ on a un étalement à droite, la distribution est plus accentué à gauche, autrement dit Mo est à gauche de Me (courbe oblique à gauche).

(iii) Si $\beta_1 < 0$ on a un étalement à gauche, la distribution est plus accentué à droite, autrement dit Mo est à droite de Me (courbe oblique à droite).

$$F_1^2 = \beta_2 = \frac{\mu_3^2}{(Var(x))^3} = \frac{\mu_3^2}{(\mu_2^3}$$

(i) Si $\beta_2 = 0$ on a une distribution symétrique.

(ii) Si $\beta_1 \neq 0$ on a une courbe asymétrique

$$\begin{cases} \mu_3 > 0, & \text{étalement à droite (Mo est à gauche de Me)}\,; \\ \mu_3 < 0, & \text{étalement à gauche (Mo est à droite de Me).} \end{cases}$$

2. Coefficient de Fisher noté F_1 ou γ_1 il est nul pour une distribution symétrique, négatif pour une distribution unimodale étalée vers la gauche, positif pour une distribution unimodale étalée vers la droite (Figure 2), donné par :

$$\gamma_1 = F_1 = \frac{\mu_3}{\sigma^3}$$

(i) Si $\gamma_1 = F_1 = 0$ on a une distribution symétrique.

(ii) Si $\gamma_1 = F_1 > 0$ on a un étalement à droite, la distribution est plus accentué à gauche, autrement dit Mo est à gauche de Me (courbe oblique à gauche).

(iii) Si $\gamma_1 = F_1 < 0$ on a un étalement à gauche, la distribution est plus accentué à droite, autrement dit Mo est à droite de Me (courbe oblique à droite).

3. Coefficient d'asymétrie de Yule : Le coefficient d'asymétrie de Yule est basé sur les positions des 3 quartiles (1er quartile, médiane et troisième quartile), et est normalisé par la distance interquartile :

$$C_y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_2}.$$

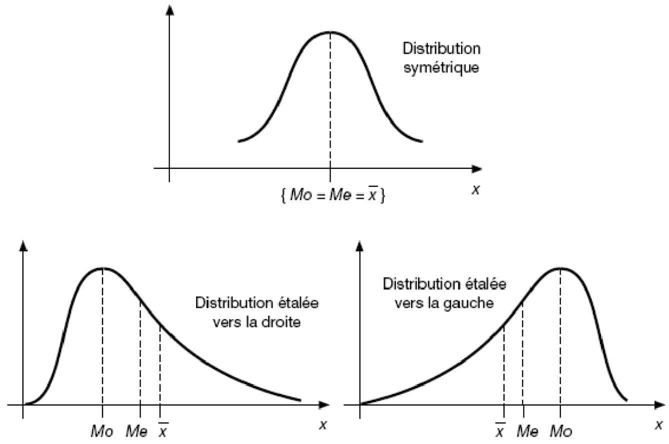


Figure 1 – Positions respective du mode, de la médiane et de la moyenne

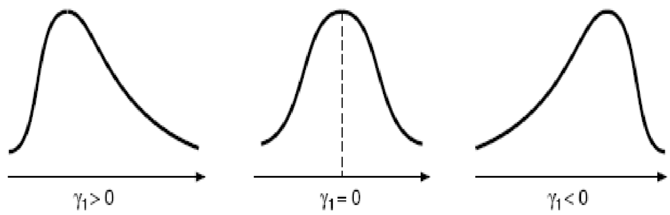


Figure 2 – Signe de coefficient d'asymétrie

- **Aplatissement**
  Les coefficients d'aplatissement de Kurtosis mesurent l'aplatissement d'une distribution ou l'importance des **"Queues"** d'une distribution, ces coefficients nous renseignent sur l'aplatissement relatif d'une distribution comparée à la distribution de la loi normale ; leurs formules sont indiquées ci-dessous :

L'aplatissement est mesuré par le coefficient d'aplatissement de Pearson

$$\beta_3 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\big(Var(x)\big)^2}$$

en général on a toujours $\mu_4 > Var(x)$.

Ou le coefficient d'aplatissement de Fisher, pour une distribution normale, ce coefficient est nul ; une valeur positive indique une distribution plus pointue que la loi normale ; une valeur négative indique à l'inverse une distribution plus aplatie.

$$F_2 = \beta_3 - 3 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mu_4}{\big(Var(x)\big)^2} - 3$$

C'est un coefficient sous dimension, invariant par changement de variable et nul pour les distributions symétriques. Ce coefficient est nul pour une distribution normale, positif ou négatif selon que la distribution est plus ou moins aplatie que la distribution normale de même moyenne et de même écart-type.

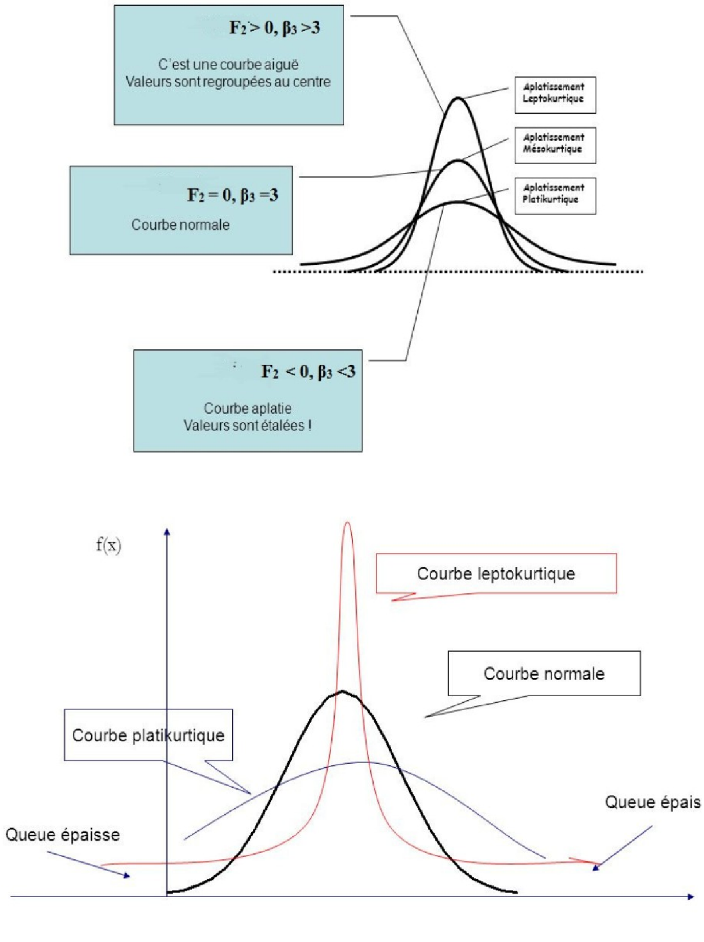


Figure 3 – Courbe Normale, Leptokurtique et Platikurtique

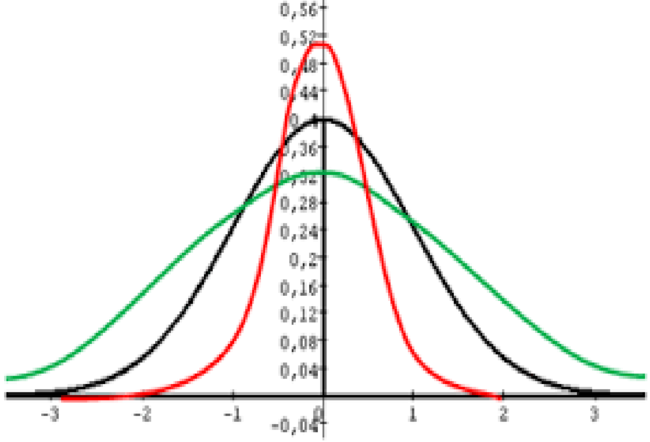
Donc

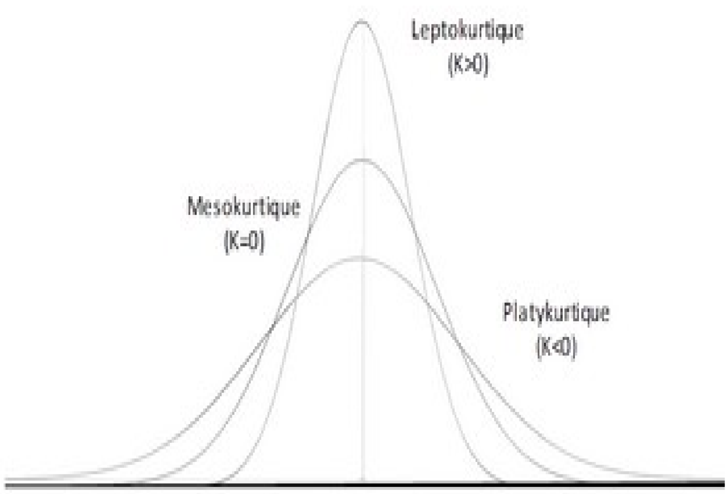
(i) Si $\beta_3 = 3$ ou $F_2 = 0$ on a une courbe normale ou Mésokurtique.

(ii) Si $\beta_3 > 3$ ou $F_2 > 0$ on a une courbe Leptokurtique ou moins aplatie. Elle est plus pointue et possède des queues plus longues. Autrement dit la concentration des valeurs de la série autour de la moyenne est forte.

(iii) Si $\beta_3 < 3$ ou $F_2 < 0$ on a une courbe Platikurtique ou plus aplatie. Elle est plus arrondie et possède des queues plus courtes. Autrement dit la concentration des valeurs de la série autour de la moyenne est faible.

Remarques : Les coefficients d'asymétrie et d'aplatissement sont invariants par changement d'origine et d'échelle, mais ils sont sensibles aux fluctuations d'échantillonnage puisqu'ils font intervenir des moments d'ordre élevé.





Exemple La répartition de 200 assurés en fonctions du nombre X d'accident d'automobiles signalés à leurs compagnie d'assurance en 5 ans est donnée par le tableau suivant :

x_i (nbre d'accidents)	0	1	2	3	4	5	6	Total
n_i (nbre d'assurés)	747	718	376	118	31	6	4	200
$n_i x_i$	0	718	752	354	124	30	24	2002
$x_i - \overline{x}$	-1	0	1	2	3	4	5	
$n_i x_i^2$	0	718	1504	1062	496	150	144	4074
$n_i(x_i - \overline{x})^4$	747	0	376	1888	2511	1536	2500	9558
$N_i \nearrow$	747	1465	1841	1959	1990	1996	2000	
$N_i \searrow$	2000	1253	535	159	41	10	4	

$\overline{x} = \dfrac{1}{N}\sum_{i=1}^{7} n_i x_i = \dfrac{2002}{2000} = 1,001 \simeq 1$ accident/assuré.

$Var(x)\dfrac{1}{N}\sum_{i=1}^{7} n_i x_i^2 - \overline{x}^2 = \dfrac{40742}{2000} - (1,001)^2 = 2,037 - 1,002 = 1,035$

$\sigma = \sqrt{Var(x)} = \sqrt{1,035} = 1,02 \simeq 1$ accident.

$\mu_4 = \dfrac{1}{N}\sum_{i=1}^{7} n_i (x_i - \overline{x})^4 = \dfrac{9558}{2000} = 4,779.$

$\dfrac{1}{4}N = 500,\ \dfrac{1}{2}N = 1000,\ \dfrac{3}{4}N = 1500,\ Q_1$ se situe à la 500 observation $\Longrightarrow Q_1 = 0,\ Q_2 = Me = 1$ et $Q_3 = 2.$

Calcule des coefficients d'asymétrie et d'aplatissement

• Coefficient de Yule

$$C_y = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{2 - 2(1) + 0}{2 - 0} = \frac{0}{2} = 0$$

On remarque que le mode $Mo = 0$.

• Coefficient d'asymétrie

$$\beta_1 = \frac{\overline{x} - Mo}{\sigma} = \frac{1,001}{1,02} = 0,98 \simeq 1 > 0$$

Donc on a une forte asymétrie (étalement à droite).

• Coefficient d'aplatissement

$$\beta_3 = \frac{\mu_4}{Var(x)^2} = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{Var(x)^2} = \frac{4,779}{(1,02)^2} = \frac{4,779}{1,082} = 4,42 > 3.$$

Donc la distribution est moins aplatie (courbe Leptokurtique).