

Chapitre 5 : Les paramètres de forme

Introduction

Les paramètres de forme permettent de préciser l'allure de la courbe des effectifs (ou des fréquences) sans avoir besoin de la tracer.

Par rapport à une courbe « idéale » dite **courbe normale**, nous définissons deux sortes de caractéristiques :

- Les caractéristiques d'asymétrie
- Les caractéristiques d'aplatissement.

Ces caractéristiques sont définies à partir des moments de la série statistique que nous allons définir ci-après.

5.1. Les moments

5.1.1. Moments simples (Moments par rapport à l'origine)

Le moment simple (la variable est centrée sur zéro) d'ordre k (k entier positif quelconque) d'une variable statistique X est la quantité définie par :

$$m_k(X) = \frac{\sum_{i=1}^n n_i x_i^k}{N}$$

Nous reconnaitrons les expressions :

$$m_1(X) = \frac{\sum_{i=1}^n n_i x_i}{N} = \bar{X} : \text{Moyenne arithmétique, moment du 1}^{\text{er}} \text{ ordre.}$$

$$m_2(X) = \frac{\sum_{i=1}^n n_i x_i^2}{N} : \text{Carré de la moyenne quadratique, moment du 2}^{\text{ème}} \text{ ordre.}$$

5.1.2. Moments centrés

Nous parlons de moment centré (la variable est centrée sur la moyenne), d'ordre k si son expression est donnée par :

$$\mu_k = \frac{\sum_{i=1}^n n_i (x_i - \bar{X})^k}{N}$$

Ainsi peut-on établir :

$$\mu_1(X) = 0 \text{ (Somme des variables centrées nulle)}$$

$$\mu_2(X) = V(X) \text{ (Variance).}$$

Le moment centré d'ordre 2 est évidemment le plus important parmi les moments centrés.

Exemple : le tableau ci-après donne la répartition de 100 ménages selon le nombre téléviseurs

Tableau 24 : La répartition de 100 ménages selon le nombre de téléviseurs

X_i	n_i	$n_i \times x_i$	$n_i \times x_i^2$	$n_i \times x_i^3$	$n_i(x_i - \bar{X})^1$	$n_i(x_i - \bar{X})^2$
1	25	25	25	25	-25	25
2	50	100	200	400	0	0
3	25	75	225	675	25	25
Total	100	200	450	1100	0	50

$$m_1 = \frac{\sum n_i x_i}{N} = \frac{200}{100} = 1$$

$$m_2 = \frac{\sum n_i x_i^2}{N} = \frac{450}{100} = 4,5$$

$$m_3 = \frac{\sum n_i x_i^3}{N} = \frac{1100}{100} = 11$$

$$\mu_1(X) = \frac{\sum n_i(x_i - \bar{X})^1}{N} = \frac{0}{100} = 0$$

$$\mu_2(X) = \frac{\sum n_i(x_i - \bar{X})^2}{N} = \frac{50}{100} = 0,5$$

Remarque :

Les moments par rapport à l'origine (simples), faciles à calculer, sont utilisés pour le calcul des moments centrés :

$$\mu_2(X) = V(X) = \frac{\sum_{i=1}^n n_i x_i^2}{N} - \bar{X}^2 = m_2 - m_1^2$$

$$\mu_3(X) = m_3 - 3m_1 m_2 + 2m_1^3$$

$$\mu_4(X) = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4$$

$\mu_3(X)$ et $\mu_4(X)$ sont aussi importants que le moment centré d'ordre 2 (variance), puisqu'ils sont utilisés dans le calcul des caractéristiques de forme.

5.2. Asymétrie

Nous savons qu'une distribution est **symétrique** si les trois caractéristiques de tendance centrale sont confondues :

- Moyenne = Médiane = Mode.
- Q_1 et Q_3 équidistants de la médiane
- D_1 et D_9 équidistants de la médiane

Dans ce cas, elle est représentée par un histogramme dit « en cloche ». La position respective de ces éléments va permettre le calcul de coefficients d'asymétrie, qui sont nuls en cas de symétrie et s'éloignent d'autant plus de la valeur 0 que la distribution est plus asymétrique.

5.2.1. Principaux coefficients d'asymétrie.

Nous en retiendrons trois, du nom de leurs auteurs qui en sont les initiateurs :

Fisher, Yule et Pearson.

Conditions d'utilisations : les coefficients d'asymétrie ne sont valables que si on possède une population nombreuse et une distribution unimodale.

a. Coefficient de Fisher

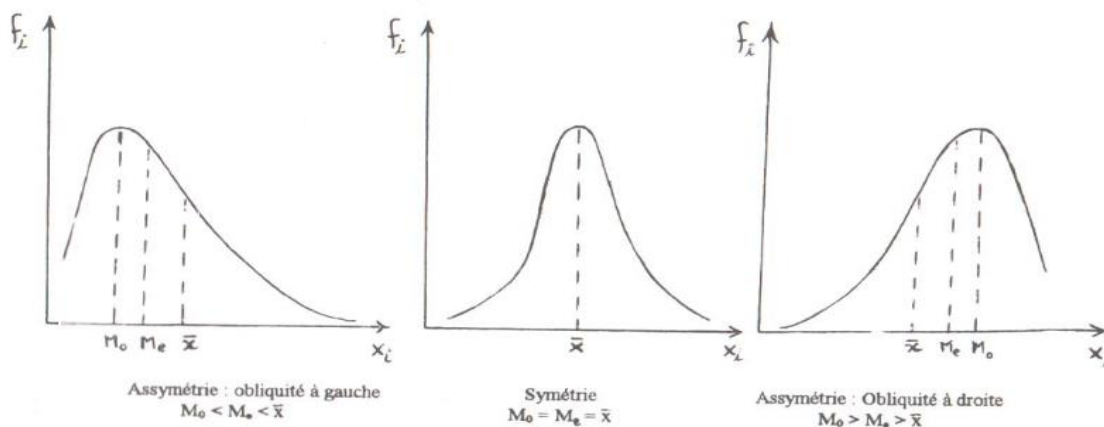
$$F_1 = \frac{\mu_3(X)}{\sigma^3(X)}$$

Nous constatons que le coefficient de Fisher n'est autre que la racine carrée du coefficient β_2 de Pearson (voir ultérieurement). Ce coefficient est sans dimension, invariant par changement d'origine et nul pour les distributions symétriques

* $F_1 = 0 \rightarrow$ Courbe symétrique (les observations sont dispersées de manière égale de part et d'autre d'une valeur centrale)

* $F_1 > 0 \rightarrow$ Étalement à droite (la courbe est oblique à gauche : M_0 est à gauche de la médiane).

* $F_1 < 0 \rightarrow$ Étalement à gauche (la courbe est oblique à droite : M_0 est à droite de la médiane).



b. Coefficient de Yule :

Ce coefficient est simple se calcule à partir des quartiles.

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

- $C_y = 0 \rightarrow$ Courbe symétrique
- $C_y > 0 \rightarrow$ La courbe est oblique à gauche (distribution étalée à droite : $M_0 < M_e < \bar{X}$)
- $C_y < 0 \rightarrow$ La courbe est oblique à droite (distribution étalée à gauche : $\bar{X} < M_e < M_0$)

Remarque : la nullité de ce coefficient ($C_y = 0$) n'est pas un indice de symétrie fiable, si l'asymétrie porte essentiellement sur des « queues » de distribution. Celle-ci est alors révélée par le coefficient de Pearson.

c. Coefficient de Pearson

Pearson propose deux coefficients pour mesurer l'asymétrie :

- **Un premier assez simple :**

$$\beta_1 = \frac{\bar{X} - M_0}{\sigma(X)}$$

- $\beta_1 = 0 \rightarrow$ Courbe symétrique
- $\beta_1 > 0 \rightarrow$ La courbe est oblique à gauche (distribution étalée à droite : $M_0 < M_e < \bar{X}$)
- $\beta_1 < 0 \rightarrow$ La courbe est oblique à droite (distribution étalée à gauche : $\bar{X} < M_e < M_0$)

- **Un second plus élaboré :**

$$\beta_2 = \frac{\mu_3^2(X)}{\mu_2^3(X)}$$

Il est calculé à partir des moments centrés de la série statistique.

- Si $\beta_2 = 0 \rightarrow$ courbe symétrique
- Si $\beta_2 \neq 0 \rightarrow$ courbe asymétrique ; seulement, **le sens de l'étalement est donné par le signe de $\mu_3(X)$.**

En effet :

- $\mu_3(x) > 0 \rightarrow$ La courbe est oblique à gauche (distribution étalée à droite : $M_0 < M_e < \bar{X}$)
- $\mu_3(x) < 0 \rightarrow$ La courbe est oblique à droite (distribution étalée à gauche : $\bar{X} < M_e < M_0$)

5.2. 2 Utilisation des coefficients d'asymétrie

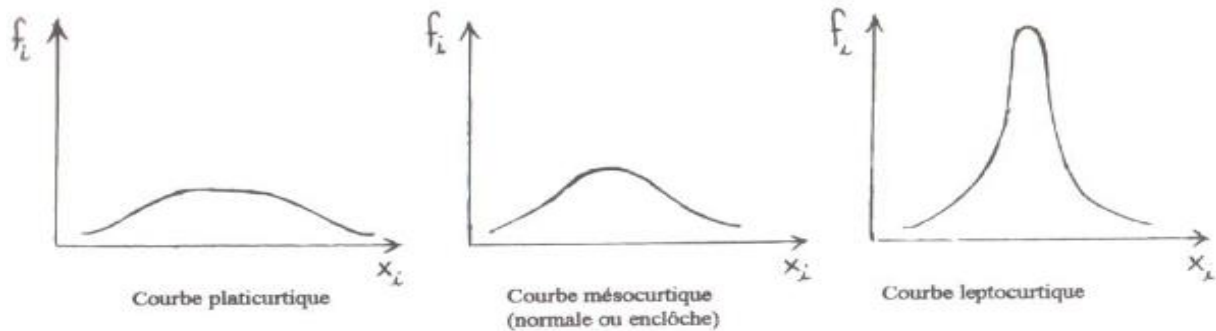
- On utilise de préférence le coefficient de Fisher F_1 lorsqu'on connaît les cas extrêmes.
- Il est préférable d'utiliser le coefficient de Yule, si on ne connaît pas les cas extrêmes (classes ouvertes), ou si on veut écarter ou on doute des cas extrêmes. Ce coefficient permet d'écarter les cas extrêmes.
- Le coefficient de Pearson β_1 est rarement utilisé, son intérêt est de donner une première idée sur la dissymétrie de la distribution.

5.3. Aplatissement

Comme pour l'asymétrie, on prend comme référence de la courbe « **normale** ». Une distribution est plus ou moins aplatie selon que les fréquences des valeurs voisines des valeurs centrales sont plus ou moins élevées par rapport aux autres.

Une courbe plus aplatie que la courbe normale est dite « **platicurtique** »

Une courbe moins aplatie que la courbe normale est dite « **leptocurtique** »



5.3.1. Principaux coefficients d'aplatissement

Les principaux coefficients exprimant le degré d'aplatissement que nous retenons sont les suivants :

a. Coefficient de Pearson

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)}$$

En général, nous savons toujours l'intégralité : $\mu_4 > \mu_2^2$

C'est pourquoi le coefficient β_3 est toujours supérieur à 1.

- Si $\beta_3 = 3 \rightarrow$ courbe normale
- Si $\beta_3 > 3 \rightarrow$ courbe leptocurtique (moins aplatie que la courbe normale).
- Si $\beta_3 < 3 \rightarrow$ courbe platicurtique (plus aplatie que la courbe normale).

Sa valeur dépasse rarement 5.

b. Coefficient de Fisher :

$$F_2 = \beta_3 - 3 = \frac{\mu_4(X)}{\sigma^4(X)} - 3$$

C'est un coefficient sans dimension, invariant par changement de variable et nul pour les distributions symétriques.

- Si $F_2 = 0 \rightarrow$ courbe normale (mésocurtique).
- Si $F_2 > 0 \rightarrow$ courbe leptocurtique (moins aplatie que la courbe normale).
- Si $F_2 < 0 \rightarrow 3 \rightarrow$ courbe platicurtique (plus aplatie que la courbe normale).

Le coefficient de Fisher F_2 est toujours supérieur à -2 du fait de l'intégralité signalée précédemment. $\mu_4(X) > \sigma^4(X)$

Remarque :

La constante 3 est choisie de telle façon que le coefficient soit nul lorsque la distribution est normale. En effet, lorsque la distribution est normale, $\mu_4(X)=3$ et $\mu_2(X)=1$ donc $\sigma(X) = 1$.

5.4. Exercices corrigés

Exercice 01 :

Répartition de 2000 assurés en fonction du nombre X d'accidents d'automobiles signalés à leur compagnie d'assurance en 5 ans. X peut varier théoriquement par valeurs entières de 0 à l'infini.

Nombre d'accident Xi	0	1	2	3	4	5	6 et +
Nombre d'assurés ni	747	718	376	118	31	6	4

- 1) Calculer les moments m_1, m_2, μ_2 et μ_4
- 2) Déterminer le mode et calculer la moyenne et l'écart type.
- 3) Calculer le coefficient d'asymétrie β_1
- 4) Calculer le coefficient d'aplatissement β_3

Corrigé exercice 01

Xi : nombre d'accident	n_i	$n_i \times x_i$	$n_i \times x_i^2$	$n_i(X_i - \bar{X})^4$
0	747	0	0	747
1	718	718	718	0
2	376	752	1504	376
3	118	354	1062	1888
4	31	124	496	2511
5	6	30	150	1536
6	4	24	144	2500
Total	2000	2002	4074	9558

1) Calcul des moments

- $m_1 = \frac{\sum n_i x_i}{N} = \frac{2002}{2000} = 1,001 = \bar{X}$
- $m_2 = \frac{\sum n_i x_i^2}{N} = \frac{4074}{2000} = 2,037$
- $\mu_2(X) = m_2 - m_1^2 = 2,037 - (1,001)^2 = 1,035$
- $\mu_4(X) = \frac{\sum n_i (X_i - \bar{X})^4}{N} = \frac{9558}{2000} = 4,779$

2) Détermination du mode et calcul de la moyenne et de l'écart type $\sigma(X)$

- **Le mode** : $Mo = 0$ accident
- **La moyenne** : $\bar{X} = m_1 = 1,001 \approx 1$ accident
- **La variance** : $V(X) = \mu_2(X) = 1,035$
- **L'écart type** : $\sigma(X) = \sqrt{V(X)} = \sqrt{1,035} = 1,017 \approx 1$ accident

3) Calcul du coefficient d'asymétrie de Pearson β_1

$$\beta_1 = \frac{\bar{X} - Mo}{\sigma(X)} = \frac{1,001 - 0}{1,02} = 0,98$$

Le coefficient β_1 est très proche de la valeur 1, ce qui prouve une forte asymétrie.

- La représentation graphique de la courbe de fréquences montre bien que le **Mo** est situé à gauche de médiane puisque l'étalement est à droite ($\beta_1 > 0$).

4) Calcul du coefficient d'aplatissement de Pearson β_3

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)} \quad \sigma^4(X) = \mu_2^2(X) = (1,035)^2 = 1,072$$

$$\text{D'où : } \beta_3 = \frac{4,779}{1,072} = 4,45.$$

Comme ($\beta_3 > 3$), donc la distribution est moins aplatie que la courbe normale, c'est donc une courbe leptocurtique.

Exercice 02 :

Soit la série suivante concernant la distribution de 100 salariés selon le salaire horaire en euros

Salaire horaire (€)	n_i
[20-25[	9
[25-30[	29
[30-35[	18
[35-40[	6
[40-45[	26
[45-50[	12
Total	100

- 1) Calculer le mode
- 2) Calculer la moyenne arithmétique
- 3) Calculer la variance et l'écart type
- 4) Calculer les principaux paramètres de position
- 5) Calculer les paramètres d'asymétrie
- 6) Calculer les paramètres d'aplatissement.

Corrigé exercice 02

Salaire horaire	n_i	$n_i \uparrow$	c_i	$n_i c_i$	$n_i c_i^2$	$n_i c_i^3$	$n_i c_i^4$	$n_i (C_i - \bar{X})^4$
[20-25[	9	9	22,5	202,5	4556,25	102515,63	2306601,56	209368,017
[25-30[	29	38	27,5	797,5	21931,25	603109,38	16585507,8	84634,4847
[30-35[	18	56	32,5	585	19012,5	617906,25	20081953,1	548,964113
[35-40[	6	62	37,5	225	8437,5	316406,25	11865234,4	295,893037
[40-45[	26	88	42,5	1105	46962,5	1995906,3	84826015,6	89046,9582
[45-50[	12	12	47,5	570	27075	1286062,5	61087968,8	307286,406
Total	100	-	-	3485	127975	4921906,3	196753281	691180,723

1) Calcul du Mode

On est dans le cas où les amplitudes des classes sont égales donc :

$$Mo \in [25 - 30[$$

$$M_o = X_{min} + \frac{\Delta 1}{\Delta 1 + \Delta 2} \times ai \rightarrow Mo = 25 + \frac{(29 - 9)}{(29 - 9) + (29 - 18)} \times 5 = 28,225 \text{ €}$$

2) Calcul de la moyenne :

$$\bar{X} = \frac{\sum n_i c_i}{N} = \frac{3485}{100} = 34,85 \text{ €}$$

3) Calcul de la variance :

$$V(X) = \frac{\sum n_i \times c_i^2}{N} - \bar{X}^2 = \frac{127975}{100} - (34,85)^2 = 65,227$$

Calcul de l'écart type :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{65,227} = 8,076 \text{ €}$$

4) Calcul des principaux paramètres de position

- $Q_1 \in [25 - 30[$

$$Q_1 = X_{min} + \frac{\frac{N}{4} - n_i \uparrow_{Q_1-1}}{n_{i_{Q_1}}} \times a_i$$

$$Q_1 = 25 + \frac{25 - 9}{29} \times 5 = 27,758 \text{ €}$$

- $Q_2 \in [30 - 35[$

$$Q_2 = X_{min} + \frac{\frac{N}{2} - n_i \uparrow_{Q_2-1}}{n_{i_{Q_2}}} \times ai$$

$$Q_2 = 30 + \frac{50 - 38}{18} \times 5 = 33,333 \text{ €}$$

- $Q_3 \in [40 - 45[$

$$Q_3 = X_{\min} + \frac{N \cdot \left(\frac{3}{4}\right) - n_i^{\uparrow Q_{3-1}}}{ni_{Q_3}} \times ai$$

$$Q_3 = 40 + \frac{75 - 62}{26} \times 5 = 42,5 \text{ £}$$

- $D_1 \in [25 - 30[$

$$D_1 = X_{\min} + \frac{\frac{N}{10} - ni^{\uparrow D_{1-1}}}{ni_{D_1}} \times ai$$

$$D_1 = 25 + \frac{10 - 9}{29} \times 5 = 25,172 \text{ £}$$

- $D_9 \in [45 - 50[$

$$D_9 = X_{\min} + \frac{N \cdot \left(\frac{9}{10}\right) - n_i^{\uparrow D_{9-1}}}{ni_{D_9}} \times ai$$

$$D_9 = 45 + \frac{90 - 88}{12} \times 5 = 45,833 \text{ £}$$

5) Calculer des paramètres d'asymétrie et d'aplatissement

5.1. Coefficients d'asymétrie

❖ Coefficient de Yule :

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

$$C_y = \frac{42,5 - 2(33,333) + 27,758}{(42,5 - 27,758)} = \frac{3,592}{14,742} = 0,243$$

$C_y > 0$ Étalement à droite (M_0 est à gauche de la médiane).

❖ Coefficient de Pearson β_1

$$\beta_1 = \frac{\bar{X} - Mo}{\sigma(X)} = \frac{34,85 - 28,225}{8,076} = 0,82$$

$\beta_1 > 0$ Donc l'étalement à droite (M_0 est à gauche de M_e)

5.2. Coefficients d'aplatissement

❖ Coefficient de Pearson β_3

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)}$$

$$\mu_4(X) = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

$$\mu_4(X) = \frac{\sum n_i c_i^4}{N} - 4 \left(\frac{\sum n_i c_i}{N} \times \frac{\sum n_i c_i^3}{N} \right) + 6 \left(\left(\frac{\sum n_i c_i}{N} \right)^2 \times \frac{\sum n_i c_i^2}{N} \right) - 3 \left(\frac{\sum n_i c_i}{N} \right)^4$$

$$\mu_4(X) = 1967532,81 - 4(34,85 \times 49219,063) + 6((34,85)^2 \times 1279,75) - 3(34,85)^4 = 6911,73$$

$$\text{Ou bien : } \mu_4(X) = \frac{\sum n_i (c_i - \bar{X})^4}{N} = \frac{691180,723}{100} = 6911,807$$

$$\sigma^4(X) = 4253,88$$

$$\beta_3 = \frac{\mu_4(X)}{\sigma^4(X)} = \frac{6911,73}{4253,88} = 1,62$$

$\beta_3 < 3$ Donc la courbe platicurtique (plus aplatie que la courbe normale).