

Statistique descriptive

Chapitre 8 : Les indicateurs de forme

Jaouad Madkour

12 octobre 2026

Plan du chapitre

Au-delà du centre	2
Ce qui manque encore	2
Les quatre premiers moments	2
Les moments, à la main	2
L'asymétrie	3
Trois formes	3
Le coefficient de Fisher	3
Fisher, sur deux exemples	4
Fisher sur données groupées	4
Les coefficients sur données groupées	4
Les autres coefficients	5
Trois coefficients, trois valeurs	5
Pourquoi ils diffèrent autant	5
Le logarithme redresse les revenus	6
L'aplatissement	6
Le coefficient d'aplatissement	6
L'aplatissement, en chiffres	6
Ce que l'aplatissement dit vraiment	7
En pratique	7
La fiche d'identité d'une distribution	7
Une mise en garde finale	8
S'entraîner	8
Exercice 1 : l'asymétrie des délais de paiement	8
Exercice 1 : solution	8
Exercice 2 : le logarithme et la symétrie	9
Exercice 2 : solution	9
Ce qu'il faut retenir	9
Le résumé du chapitre	9

Au-delà du centre

Ce qui manque encore

Deux distributions, même moyenne, même écart-type

Et pourtant deux réalités différentes : l'une symétrique, l'autre étalée d'un côté ; l'une aux extrêmes rares, l'autre aux extrêmes fréquents.

Il faut donc deux indicateurs supplémentaires, qui répondent à deux questions distinctes :

- l'**asymétrie** : la distribution penche-t-elle d'un côté ?
- l'**aplatissement** : les valeurs extrêmes sont-elles fréquentes ?

Moment centré d'ordre r

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

Les quatre premiers moments

Ce que mesure chaque ordre

Ordre	Valeur	Ce qu'il mesure
$r = 1$	$m_1 = 0$	rien — c'est la propriété d'équilibre
$r = 2$	$m_2 = V(x)$	la dispersion
$r = 3$	m_3	l' asymétrie
$r = 4$	m_4	l' aplatissement

Pourquoi les ordres impairs mesurent l'asymétrie

Une puissance **impaire** conserve le signe : un écart de -3 contribue pour -27 , un écart de $+3$ pour $+27$. Ils se compensent si la distribution est symétrique, et ne se compensent pas sinon. Une puissance **paire** détruit le signe : $(-3)^4$ et $(+3)^4$ valent tous deux 81.

L'ordre trois détecte le déséquilibre, l'ordre quatre l'épaisseur des queues.

Les moments, à la main

Cinq valeurs : 2, 3, 4, 8, 13 ; moyenne 6

x_i	$e_i = x_i - 6$	e_i^2	e_i^3	e_i^4
2	-4	16	-64	256
3	-3	9	-27	81
4	-2	4	-8	16
8	2	4	8	16
13	7	49	343	2401
Total	0	82	252	2770

La lecture de la colonne des cubes

Les quatre premiers cubes totalisent -91 ; le seul 343 de la valeur 13 fait basculer la somme.
Le moment d'ordre trois est l'affaire de la queue.

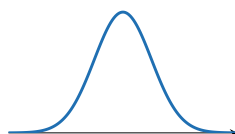
L'asymétrie

Trois formes

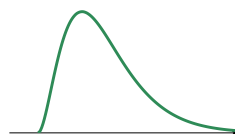
asymétrie **négative**
 $\gamma_1 < 0$: étalée à gauche



symétrique
 $\gamma_1 = 0$: moyenne = médiane

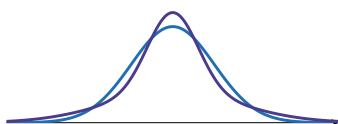


asymétrie **positive**
 $\gamma_1 > 0$: étalée à droite



l'aplatissement

— normale, $\gamma_2 = 3$
— leptokurtique, $\gamma_2 > 3$



Deux questions distinctes, deux moments.

L'*asymétrie* demande : la distribution penche-t-elle d'un côté ?
Elle se mesure par le moment centré d'ordre **trois**, réduit : $\gamma_1 = m_3/\sigma^3$.

L'*aplatissement* demande : les valeurs extrêmes sont-elles plus fréquentes que sous une loi normale ? Moment d'ordre **quatre** : $\gamma_2 = m_4/\sigma^4$.

Sur nos salaires : $\gamma_1 = 1,16$ et $\gamma_2 = 4,84$. Étendue à droite, et des extrêmes plus fréquents que sous la normale.

Le coefficient de Fisher

Coefficient d'asymétrie

$$\gamma_1 = \frac{m_3}{\sigma^3}$$

On divise par σ^3 pour obtenir un nombre **sans unité**, comparable d'une variable à l'autre.

La lecture

Valeur	Forme	Position des indicateurs
$\gamma_1 = 0$	symétrique	mode = médiane = moyenne
$\gamma_1 > 0$	étalée à droite	mode < médiane < moyenne
$\gamma_1 < 0$	étalée à gauche	mode > médiane > moyenne

Fisher, sur deux exemples

Les cinq valeurs

$m_2 = 82/5 = 16,4$, donc $\sigma = 4,05$ et $\sigma^3 = 66,4$; $m_3 = 252/5 = 50,4$:

$$\gamma_1 = \frac{50,4}{66,4} = 0,76$$

Nos 500 salaires

$$\gamma_1 = 1,16$$

Nettement positif : distribution étalée à droite, ce que confirment mode 6 188 < médiane 6 770 < moyenne 7 158.

Repères usuels : **à peu près symétrique** si $|\gamma_1| < 0,5$, **modérément asymétrique** entre 0,5 et 1, **fortement asymétrique** au-delà — la règle pour les revenus.

Fisher sur données groupées

Les moments à partir des six classes, en milliers de dirhams

c_i	n_i	$e_i = c_i - 7,21$	$n_i e_i^3$	$n_i e_i^4$
3	31	-4,21	-2 313	9 738
5	157	-2,21	-1 695	3 745
7	166	-0,21	-2	0
9	79	1,79	453	811
12	57	4,79	6 264	30 007
17	10	9,79	9 383	91 860
Total	500		12 091	136 162

Les coefficients sur données groupées

Asymétrie et aplatissement, à partir du tableau

$$m_3 = \frac{12\,091}{500} = 24,18 \quad \sigma^3 = 2,772^3 = 21,31 \quad \gamma_1 = \frac{24,18}{21,31} = 1,13$$

$$m_4 = \frac{136\,162}{500} = 272,3 \quad \sigma^4 = 7,686^2 = 59,07 \quad \gamma_2 = \frac{272,3}{59,07} = 4,61$$

Contre 1,16 et 4,84 sur les données exactes.

Tout se joue dans la queue

Les deux dernières classes apportent 15 647 à la somme des cubes, qui ne vaut que 12 091 : sans elles, elle serait négative. Elles font aussi près de 90 % de la somme des puissances quatrièmes.

Les autres coefficients

Deux mesures alternatives

Coefficient de Pearson — fondé sur l'écart entre moyenne et mode :

$$A_P = \frac{\bar{x} - \text{mode}}{\sigma}$$

Coefficient de Yule — fondé sur les quartiles :

$$A_Y = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1}$$

Trois façons de voir l'asymétrie

Fisher utilise tous les écarts, Pearson l'écart entre la moyenne et le mode, Yule la position de la médiane dans la boîte.

Trois coefficients, trois valeurs

Les trois, sur nos salaires

Coefficient	Calcul	Valeur
Fisher γ_1	m_3/σ^3	1,16
Pearson A_P	$(7\,158 - 6\,188)/2\,598$	0,37
Yule A_Y	$(8\,330 + 5\,318 - 2 \times 6\,770)/3\,012$	0,036

Un écart qui demande une explication

De 0,036 à 1,16 : les trois coefficients s'accordent sur le **sens** de l'asymétrie, pas sur son **ampleur**. La diapositive suivante dit pourquoi.

Pourquoi ils diffèrent autant

Ils ne mesurent pas la même chose

Yule ne regarde que la boîte : la médiane est presque au centre de l'intervalle interquartile, d'où une valeur quasi nulle. Le cœur de la distribution est presque symétrique.

Fisher utilise toutes les valeurs, en donnant un poids au cube aux plus éloignées : il capte pleinement la queue droite.

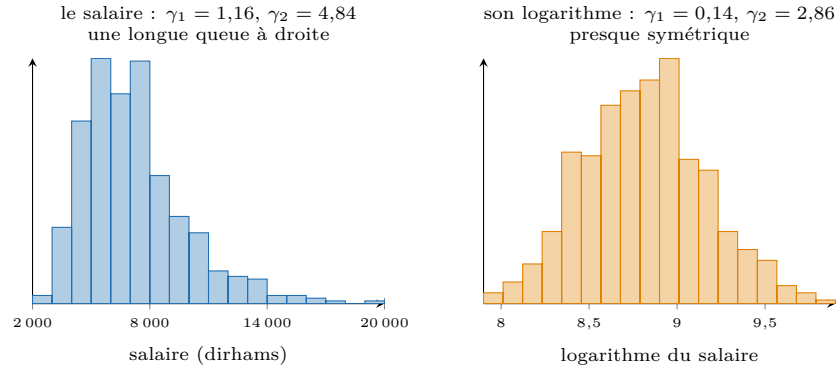
La comparaison des deux est un diagnostic : $A_Y \simeq 0$ et $\gamma_1 = 1,16$ signifient que l'asymétrie ne vient pas du centre mais **de la queue**. La masse des salariés est répartie symétriquement ; quelques hauts salaires créent l'étalement.

Ce que cela vaut pour une négociation

« Les salaires sont-ils inégalement répartis ? » Pour la majorité du personnel, la répartition est équilibrée ; ce sont les rémunérations les plus élevées qui créent l'asymétrie. Un seul

indicateur aurait manqué cette nuance : **on en calcule toujours deux.**

Le logarithme redresse les revenus



Une propriété des revenus

Le logarithme comprime les grandes valeurs bien plus que les petites : la queue droite se replie, et γ_1 tombe de 1,16 à 0,14. D'où l'usage du **logarithme du salaire** en économie — et la moyenne géométrique (6 738 dirhams) si proche de la médiane.

L'aplatissement

Le coefficient d'aplatissement

Coefficient d'aplatissement

$$\gamma_2 = \frac{m_4}{\sigma^4}$$

La référence est la **loi normale**, pour laquelle $\gamma_2 = 3$.

La lecture

Valeur	Nom	Interprétation
$\gamma_2 = 3$	mésokurtique	comme la loi normale
$\gamma_2 > 3$	leptokurtique	pic plus aigu, queues épaisses
$\gamma_2 < 3$	platykurtique	distribution plus plate

L'aplatissement, en chiffres

Sur nos salaires

$$\gamma_2 = 4,84 \quad \text{soit un excès de } 1,84$$

Distribution **leptokurtique** : les salaires très éloignés de la moyenne sont plus fréquents que sous une loi normale de même écart-type.

La notation à surveiller

Certains logiciels affichent l'**excès d'aplatissement**, $\gamma_2 - 3$, dont la référence est 0. R (kurtosis de moments) donne γ_2 ; Python (`scipy.stats.kurtosis`) donne l'excès par défaut, sauf avec `fisher=False`. **Vérifiez toujours la convention avant d'interpréter.**

Ce que l'aplatissement dit vraiment

Une idée fausse répandue

On enseigne souvent que γ_2 mesure « la hauteur du pic ». C'est inexact. Le moment d'ordre quatre donne un poids en $(x - \bar{x})^4$: les valeurs proches de la moyenne contribuent **presque rien**. Dans le tableau groupé, les deux dernières classes font 121 867 des 136 162 — près de 90 %.

L'aplatissement mesure l'épaisseur des queues, pas la forme du sommet.

L'indicateur central de la finance

Les rentabilités boursières quotidiennes ont souvent un γ_2 de 5 à 10, parfois davantage : les fortes variations sont bien plus fréquentes que ne le prédit une loi normale. Les modèles de risque qui supposent la normalité **sous-estiment** les pertes extrêmes — un reproche souvent adressé à la Value at Risk gaussienne après la crise de 2008.

En pratique

La fiche d'identité d'une distribution

Les six nombres à donner

Question	Indicateur	Valeur
Où ?	moyenne	7 158
Où ? (robuste)	médiane	6 770
Combien dispersée ?	écart-type	2 598
Dispersion relative	CV	0,363
Penche-t-elle ?	γ_1	1,16
Queues épaisses ?	γ_2	4,84

Ce que ces six nombres racontent

Des salaires centrés autour de 6 800 dirhams, assez dispersés, **nettement étalés vers les hauts salaires**, avec des valeurs élevées plus fréquentes que sous une loi normale. Le diagnostic tient en une phrase et se recalcule par n'importe qui sur les mêmes données.

Une mise en garde finale

Les moments sont fragiles

γ_1 élève les écarts au cube, γ_2 à la puissance quatre. Une **seule** valeur aberrante peut donc les bouleverser.

Sur nos données, remplacer le plus haut salaire par 190 050 dirhams ferait passer γ_1 de 1,16 à 19,4 et γ_2 de 4,8 à 414 — une seule ligne modifiée sur cinq cents.

Toujours regarder l’histogramme et la boîte à moustaches avant d’interpréter un coefficient de forme. Une valeur extrême de γ_2 signale bien plus souvent une erreur de saisie qu’une propriété de la population.

S’entraîner

Exercice 1 : l’asymétrie des délais de paiement

Les délais de paiement de 100 clients

Délai (jours)	$[0 ; 30[$	$[30 ; 60[$	$[60 ; 90[$	$[90 ; 150[$
Clients n_i	40	30	20	10

1. Calculer les centres c_i , puis $u_i = (c_i - 45)/15$.
2. Calculer $\bar{u}$ et $V(u)$, puis $\bar{x}$ et σ_x .
3. Calculer le moment centré d’ordre trois de u , puis γ_1 . Pourquoi peut-on le calculer directement sur u ?

Exercice 1 : solution

Moyenne et dispersion

1. Centres 15, 45, 75, 120 ; d’où $u = -2, 0, 2, 5$.
2. $\sum n_i u_i = -80 + 0 + 40 + 50 = 10$, donc $\bar{u} = 0,1$; $\sum n_i u_i^2 = 160 + 0 + 80 + 250 = 490$, donc $V(u) = 4,9 - 0,01 = 4,89$. Alors $\bar{x} = 45 + 15 \times 0,1 = 46,5$ jours et $\sigma_x = 15 \times \sqrt{4,89} = 33,2$ jours.

Le moment d’ordre trois

3. Écarts $u_i - 0,1$: $-2,1$; $-0,1$; $1,9$; $4,9$. Leurs cubes pondérés : $-370,4$; $-0,03$; $137,2$; $1\,176,5$, de somme 943,2. Donc $m_3(u) = 9,43$ et

$$\gamma_1 = \frac{9,43}{4,89^{3/2}} = \frac{9,43}{10,81} = 0,87$$

γ_1 est **sans unité** : changer d’origine et d’unité ne le modifie pas. Distribution modérément étalée à droite.

Exercice 2 : le logarithme et la symétrie

Cinq revenus qui doublent

Cinq ménages déclarent des revenus mensuels de 2, 4, 8, 16 et 32 milliers de dirhams.

1. Calculer la moyenne et la médiane. Que dit leur position ?
2. Prendre le logarithme de base 2 de chaque revenu. Quelle forme a la nouvelle série ? Que vaut son γ_1 ?
3. Revenir à l'échelle des dirhams : à quoi correspond la moyenne des logarithmes ?

Exercice 2 : solution

Une asymétrie qui disparaît

1. Moyenne $62/5 = 12,4$, médiane 8 : la moyenne dépasse nettement la médiane, la série est **étalée à droite** ($\gamma_1 = 0,89$).
2. Les logarithmes valent 1, 2, 3, 4, 5 : une série **parfaitement symétrique**, de moyenne 3 et de $\gamma_1 = 0$.
3. $2^3 = 8$ milliers : c'est la **moyenne géométrique** des revenus, $\sqrt[5]{2 \times 4 \times 8 \times 16 \times 32} = 8$. Elle coïncide ici avec la médiane.

La leçon

Des revenus qui progressent **par multiplication** sont asymétriques en niveau et symétriques en logarithme. C'est l'origine de l'asymétrie à droite de presque toutes les distributions de revenus.

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. **Moment centré d'ordre r** : $m_r = \frac{1}{n} \sum (x_i - \bar{x})^r$. Ordre 2 : dispersion. Ordre 3 : asymétrie. Ordre 4 : aplatissement.
2. Ordres **impairs** pour le sens, ordres **pairs** pour l'ampleur.
3. $\gamma_1 = m_3/\sigma^3$: positif, la distribution est étalée **à droite**. Yule et Pearson donnent d'autres regards.
4. $\gamma_2 = m_4/\sigma^4$: la référence est 3. Au-delà, **queues épaisses**.
5. Les coefficients sont **sans unité** : on peut les calculer sur une variable transformée u .
6. Le **logarithme** redresse les distributions de revenus.
7. Ces coefficients sont **très sensibles aux valeurs extrêmes** : jamais sans le graphique.

La suite

Il reste une question que ni la dispersion ni la forme ne traitent : **comment la masse totale se répartit-elle entre les individus** ? C'est celle de la **concentration**, et c'est le chapitre 9.