

6. Forme des distributions

Statistique Descriptive

Anicet Ebou

Institut National Polytechnique Félix Houphouët-Boigny

2026-2027

1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Deux séries peuvent avoir la **même moyenne** et le **même écart-type** tout en ayant des formes très différentes.

Exemple : rendements (t/ha)

Série A : symétrique

$$\bar{x} = 4,0 \quad s = 0,8$$

Série B : étalée à droite

$$\bar{x} = 4,0 \quad s = 0,8$$

La **forme** révèle :

- ▶ La présence de **valeurs extrêmes** (queues)
- ▶ Si la distribution est **symétrique** ou non
- ▶ Si les données viennent d'une seule population ou de **deux groupes mélangés**
- ▶ Quelle mesure de centre est la plus **fiable**

En agronomie : un rendement rarement très élevé ne mérite pas le même traitement qu'un rendement rarement très bas.

Asymétrie
(skewness)
direction du déséquilibre

+

Aplatissement
(kurtosis)
poids des queues

Ensemble, ils décrivent complètement la forme
d'une distribution univariée.

Asymétrie répond à :

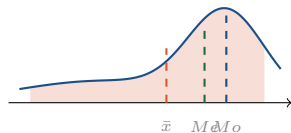
- ▶ La queue est-elle à gauche ou à droite ?
- ▶ Moyenne $\geq$ médiane ?

Aplatissement répond à :

- ▶ La distribution est-elle **pointue** ou **plate** ?
- ▶ Y a-t-il beaucoup de valeurs dans les queues ?

1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Les trois cas d'asymétrie

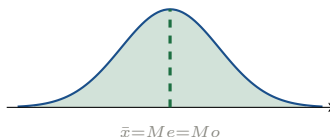


Asymétrie négative

(étalée à **gauche**)

queue longue à gauche

$$\bar{x} < Me < Mo$$

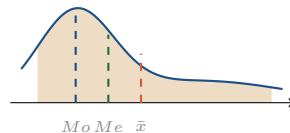


Symétrique

(asymétrie nulle)

queue égale des deux côtés

$$\bar{x} = Me = Mo$$



Asymétrie positive

(étalée à **droite**)

queue longue à droite

$$Mo < Me < \bar{x}$$

Mnémotechnique : la queue indique la **direction** de l'asymétrie, et tire la **moyenne** vers elle.

Exemples agronomiques typiques

- ▶ **Revenus agricoles** : souvent *étalés à droite*, la plupart des exploitations ont des revenus modestes, quelques grandes exploitations tirent la moyenne vers le haut
- ▶ **Temps de germination** : souvent *étalés à droite*, on germe rarement en dessous du minimum biologique
- ▶ **Taux de mortalité des plants** : souvent *étalés à gauche* dans une plantation bien gérée, la plupart survivent, quelques parcelles sont perdues
- ▶ **Poids des fruits mûrs** : souvent *symétrique* si la variété est homogène

Conséquence pratique

En présence d'asymétrie positive :

- ▶ La **moyenne** est gonflée par les valeurs extrêmes
- ▶ Préférer la **médiane** pour le centre
- ▶ L'**EIQ** pour la dispersion
- ▶ Mentionner l'asymétrie dans tout rapport

Coefficient d'asymétrie γ_1

$$\gamma_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Moment centré réduit d'ordre 3.

Valeur de γ_1	Interprétation
$\gamma_1 < 0$	étalée à gauche
$\gamma_1 = 0$	symétrique
$\gamma_1 > 0$	étalée à droite

Pourquoi l'exposant 3 ?

- ▶ L'exposant **impair** conserve le signe des écarts
- ▶ Les grands écarts positifs **dominent** si la queue est à droite
- ▶ $|\gamma_1| < 0,5$: asymétrie faible
- ▶ $|\gamma_1| \in [0,5 ; 1]$: modérée
- ▶ $|\gamma_1| > 1$: forte

Approx. rapide de Pearson

$$\gamma_1 \approx \frac{3(\bar{x} - Me)}{s}$$

Utilise uniquement moyenne, médiane, écart-type.

```
# Revenus mensuels (*1000 FCFA), 12
  exploitations
revenus <- c(85, 90, 92, 95, 95, 98, 100,
            105, 108, 115, 140, 210)
# mean: 111.1, median: 99.0, sd: 34.5
mean(revenus); median(revenus)
sd(revenus)
# Approx. de Pearson
gamma_pearson <- 3 * (mean(revenus) -
                    median(revenus)) / sd(revenus)
gamma_pearson # 1.06 → forte asymetrie
               positive
# Package e1071 (moment exact)
# install.packages("e1071")
library(e1071)
skewness(revenus) # 1.74
```

Interprétation

$\gamma_1 \approx 1,7$: forte asymétrie positive.
La **moyenne** (111 000 FCFA) est gonflée par les deux grandes exploitations (140 et 210).
La **médiane** (99 000 FCFA) représente mieux le revenu "typique".

Règle : si $|\bar{x} - Me| > 0,1 \times s$,
l'asymétrie est notable, vérifier γ_1 .

1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Qu'est-ce que l'aplatissement ?

L'aplatissement mesure la **concentration des valeurs** autour du centre et le **poids des queues** par rapport à une distribution de référence.

Kurtosis (excès)

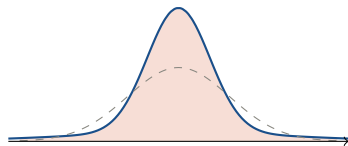
$$\gamma_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - 3$$

Le -3 centre la mesure sur la distribution normale (référence : $\gamma_2 = 0$).

- ▶ $\gamma_2 = 0$: **mésokurtique**
(profil normal, référence)
- ▶ $\gamma_2 > 0$: **leptokurtique**
pic plus pointu, queues plus lourdes
(*lepto* = mince)
- ▶ $\gamma_2 < 0$: **platykurtique**
distribution plus plate, queues légères
(*platy* = plat)

L'exposant 4 amplifie les valeurs éloignées du centre, le kurtosis est très sensible aux valeurs extrêmes.

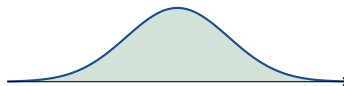
Visualiser les trois profils de kurtosis



Leptokurtique

$$\gamma_2 > 0$$

pic élevé, queues lourdes

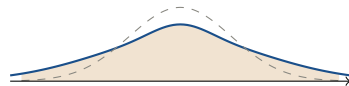


référence

Mésokurtique

$$\gamma_2 = 0$$

profil normal



Platykurtique

$$\gamma_2 < 0$$

distribution aplatie

La courbe en tirets gris est la référence normale (mésokurtique) superposée pour comparaison.

```
library(e1071)
# Poids de 200 noix de cajou (g)
set.seed(42)
poids_normal ← rnorm(200, mean = 6.8, sd
  = 1.1)
poids_lepto ← c(rnorm(180, 6.8, 0.4),
  rnorm(20, 6.8, 4.0))
kurtosis(poids_normal) # proche de 0
kurtosis(poids_lepto) # > 0
# Visualisation cote a cote
par(mfrow = c(1, 2))
hist(poids_normal, main = "Normal", col
  = "lightgreen")
hist(poids_lepto, main = "Lepto", col =
  "lightyellow")
par(mfrow = c(1, 1))
```

Signification agronomique

Un kurtosis **positif** sur une mesure de poids indique la présence de quelques valeurs très éloignées de la moyenne :

- ▶ défauts de calibre extrêmes
- ▶ lots de qualités très différentes mélangés
- ▶ erreurs de mesure ponctuelles

En pratique, le kurtosis est moins utilisé que l'asymétrie. Il importe surtout avant d'appliquer un test statistique qui suppose la normalité.

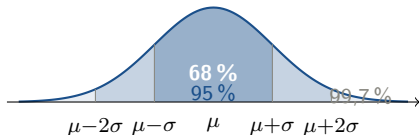
1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Distribution normale $\mathcal{N}(\mu, \sigma^2)$

Courbe en cloche, parfaitement symétrique, entièrement caractérisée par μ et σ .

$$\gamma_1 = 0 \quad \gamma_2 = 0$$

Règle empirique 68–95–99,7



Lecture pratique :

- ▶ 68 % des valeurs dans $[\mu - \sigma ; \mu + \sigma]$
- ▶ 95 % dans $[\mu - 2\sigma ; \mu + 2\sigma]$
- ▶ 99,7 % dans $[\mu - 3\sigma ; \mu + 3\sigma]$

Application

Hauteur des anacardiers :

$$\bar{x} = 4,2 \text{ m}, s = 0,72 \text{ m}$$

⇒ 95 % des arbres entre

$$4,2 - 1,44 = 2,76 \text{ m et}$$

$$4,2 + 1,44 = 5,64 \text{ m}$$

Distribution log-normale

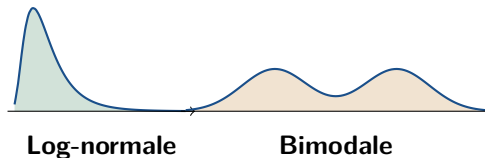
Si $\ln(X)$ suit une loi normale, alors X suit une **loi log-normale**.

- ▶ Toujours **positive** et étalée à droite
- ▶ Fréquente pour les : revenus, surfaces de parcelles, concentrations de nutriments, biomasses

Distribution bimodale

Deux pics distincts dans l'histogramme.

- ▶ Souvent signe de **deux groupes mélangés**
 - ▶ Exemples : poids de récolte sur deux variétés confondues, rendements avant/après irrigation
- ⇒ **Stratifier** avant d'analyser !



1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Pourquoi vérifier la normalité ?

De nombreuses méthodes statistiques supposent que les données suivent une **loi normale** :

- ▶ Tests t de comparaison de moyennes
- ▶ ANOVA
- ▶ Régression linéaire (résidus normaux)

Si cette hypothèse est violée, les résultats peuvent être **incorrects**.

Deux approches complémentaires :

1. **Visuels** : histogramme, QQ-plot
2. **Tests formels** : Shapiro-Wilk

Quand la normalité importe moins

- ▶ Grands échantillons ($n > 30$) : le **théorème central limite** assure la normalité de la moyenne
- ▶ Méthodes non paramétriques : ne supposent aucune distribution
- ▶ Statistiques descriptives pures : pas besoin de normalité pour calculer une moyenne ou un quartile

Le QQ-plot : outil visuel clé

```
# QQ-plot : quantiles observes vs.
#          normaux
hauteur <- c(3.2,4.1,2.8,5.0,3.7,4.4,
             3.1,6.2,4.8,3.5,4.0,2.6, 5.5,3.9,4.2,
             3.3,4.7,5.1,3.6,4.9)
qqnorm(hauteur, main = "QQ-plot -
             hauteur",
        pch = 16, col = "steelblue")
qqline(hauteur, col = "red", lwd = 2)

# Test de Shapiro-Wilk
shapiro.test(hauteur)
# p-value > 0.05 → pas de preuve contre
#          normalité
```

Lire un QQ-plot :

- ▶ Points **sur la diagonale rouge**
⇒ normal
- ▶ Points en **S** ⇒ asymétrie
- ▶ Queues **relevées** ⇒ queues lourdes (lepto)
- ▶ Queues **abaissées** ⇒ queues légères (platy)

Shapiro-Wilk :

- ▶ H_0 : les données sont normales
- ▶ $p > 0,05$: pas de raison de rejeter H_0
- ▶ $p \leq 0,05$: non-normalité détectée

1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Analyse complète d'une distribution en R

```
rendement ← c(8.2,12.5,6.1,15.0,9.8,11.3,7.5,18.4,14.2,9.0,10.5,5.8,16.7,
  10.2, 11.8, 8.7,13.5,15.3,9.4,14.8)
# --- Centre et dispersion ---
cat("Moyenne :", round(mean(rendement), 2), "\n")
cat("Mediane :", median(rendement), "\n")
cat("Ecart-type:", round(sd(rendement), 2), "\n")
cat("CV (%) :", round(sd(rendement)/mean(rendement)*100, 1), "\n")
# --- Forme ---
library(e1071)
cat("Asymetrie:", round(skewness(rendement), 3), "\n")
cat("Kurtosis :", round(kurtosis(rendement), 3), "\n")
# --- Normalite ---
shapiro.test(rendement)
# --- Graphique complet ---
hist(rendement, breaks = 6, col = "lightgreen", main = "Distribution du
  rendement", xlab = "Rendement (kg/arbre/an)")
abline(v = mean(rendement), col = "red", lwd = 2)
abline(v = median(rendement), col = "blue", lwd = 2)
```



Résultats attendus

Mesure	Valeur
Moyenne	11,9 kg
Médiane	11,6 kg
Écart-type	3,4 kg
CV	28,6 %
γ_1 (asymétrie)	+0,41
γ_2 (kurtosis)	-0,58
Shapiro-Wilk p	0,51

Lecture :

- Asymétrie faible ($|\gamma_1| < 0,5$) : distribution quasi-symétrique
 - Kurtosis légèrement négatif : distribution légèrement aplatie
 - Shapiro $p = 0,51 > 0,05$: normalité **non rejetée**
 - Moyenne $\approx$ médiane : confirmé par γ_1 faible
- ⇒ La **moyenne** est un résumé fiable ici, la distribution le permet.

1. Pourquoi s'intéresser à la forme ?
2. Asymétrie (skewness)
3. Aplatissement (kurtosis)
4. Distributions particulières
5. Tester la normalité
6. Tableau de bord complet
7. Récapitulatif

Les mesures de forme

- ▶ **Asymétrie** γ_1 :
 - ▶ > 0 : queue à droite, moyenne $>$ médiane
 - ▶ < 0 : queue à gauche, moyenne $<$ médiane
 - ▶ $|\gamma_1| > 1$: forte
- ▶ **Kurtosis** γ_2 :
 - ▶ > 0 : pic pointu, queues lourdes
 - ▶ < 0 : distribution plate
- ▶ **Shapiro-Wilk** : test formel de normalité ($n \leq 50$)

Les réflexes

1. Toujours **tracer** histogramme + QQ-plot
2. Comparer $\bar{x}$ et Me : un écart notable signale une asymétrie
3. Asymétrie forte $\Rightarrow$ utiliser **médiane + EIQ**
4. Avant tout test paramétrique : **vérifier la normalité**
5. Distribution bimodale $\Rightarrow$ **stratifier** et analyser séparément

Décrire une distribution = centre + dispersion + forme.
Sans la forme, le tableau de bord est incomplet.

Merci de votre attention

Statistiques Descriptives · Module 6

Equipe Bioinformatique et Biostatistique

Laboratoire de Microbiologie, Biotechnologie et Bioinformatique

Institut National Polytechnique Félix Houphouët-Boigny

