



L'Insee

Institut national de la statistique et des études économiques

Auteur
Fonction
Date





Le statut

- › **L'Institut national de la statistique et des études économiques est une direction générale du ministère de l'Economie, de l'Industrie et de l'Emploi.**
- › **L'Insee, ce sont :**
 - **une direction générale**
 - **24 directions régionales**
 - **6400 agents**



Les missions

- **Produire** des statistiques et des bases de données d'intérêt général
- **Analyser** l'information économique et sociale nationale et régionale
- **Diffuser** l'ensemble des résultats auprès d'un large public



Les 5 champs d'intervention

- › **Développement économique**
- › **Aménagement du territoire**
- › **Population - politiques sociales**
- › **Aménagement urbain - logements**
- › **Emploi - formation**



L'activité de la statistique publique est encadrée par la loi

➡ **la loi n° 51-711 du 7 juin 1951** modifiée sur l'obligation, la coordination et le secret en matière de statistique.

➡ **la loi n° 78-17 du 6 janvier 1978** modifiée, relative à l'informatique, aux fichiers et aux libertés.



Retrouvez facilement tous les produits de l'Insee sur

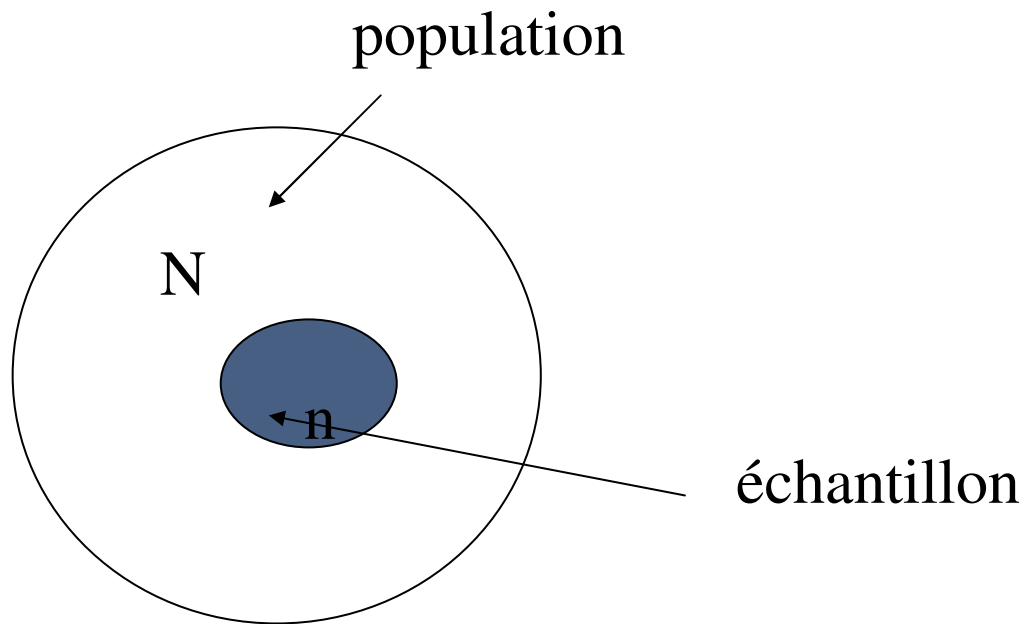
www.insee.fr





Pour faire de la statistique est se comprendre il faut un peu de vocabulaire :

- la population : ensemble d'unités réunies selon certaines caractéristiques ou variables
- Un sous-ensemble de cette population : un échantillon
- Un élément de cette population : une unité statistique



i ← Un individu i échappé



Encore du vocabulaire :

Les variables :

- quantitatives
- qualitatives
- ordinales

les variables prennent souvent des valeurs encadrées par un mini et un maxi :

Ex : la taille des personnes varie entre 0,45 m et 2,4 m

Les revenus entre 0 et....



Deux classes, dont on a relevé les notes des élèves, ont pour moyenne des notes, respectivement :

Classe A : 10,41

Classe B : 10,04

a priori la classe A est mieux notée que la classe B

Est-ce vrai ?



Basic Statistical Measures			
Location		Variability	
Mean	10.41837	Std Deviation	2.68357
Median	10.00000	Variance	7.20153
Mode	10.00000	Range	11.50000
		Interquartile Range	3.00000

Stats de base: moyenne, médiane, mode, variance, étendue



Basic Statistical Measures			
Location		Variability	
Mean	10.04082	Std Deviation	3.77607
Median	10.00000	Variance	14.25872
Mode	12.00000	Range	16.00000
		Interquartile Range	6.00000



La moyenne arithmétique :

$$\frac{x_1 + x_2 + \dots x_i + \dots x_n}{n} = \bar{x}$$

\bar{x} se dit "x barre"



Examinons deux séries de note

A	B
1	5
5	5
4	5
2	5
15	5
3	5
8	5
0	5
5	5
7	5

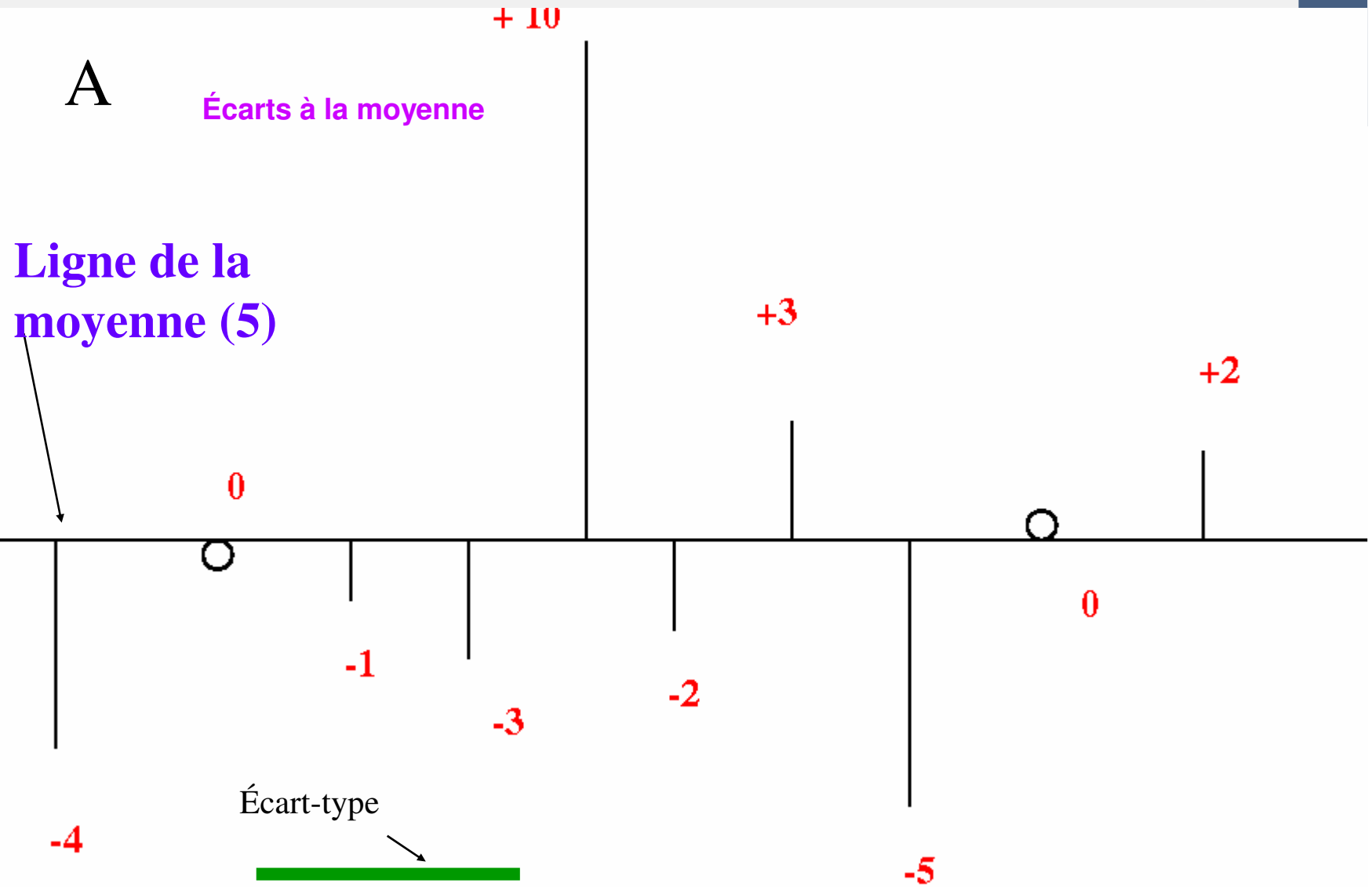
La moyenne est
égale à 5 dans
les deux séries



Comment les distinguer, alors que les 2 moyennes sont égales ?

Avec une autre moyenne à qui on donne un nom,

On l'appellera variance





B

Ligne de la
moyenne (5)



Écarts à la moyenne



Écart-type



La **moyenne** des **carrés** des **écarts** à la moyenne

1 : je calcule les écarts à la moyenne

2 : j'élève chacun d'entre eux au carré

3 : je fais la moyenne arithmétique de ces carrés



Comparer deux moyennes ne suffit pas,
Il faut associer les écarts-types à la
comparaison



On peut aussi, et on devrait le faire
avant tout calcul, comparer les formes
des distributions



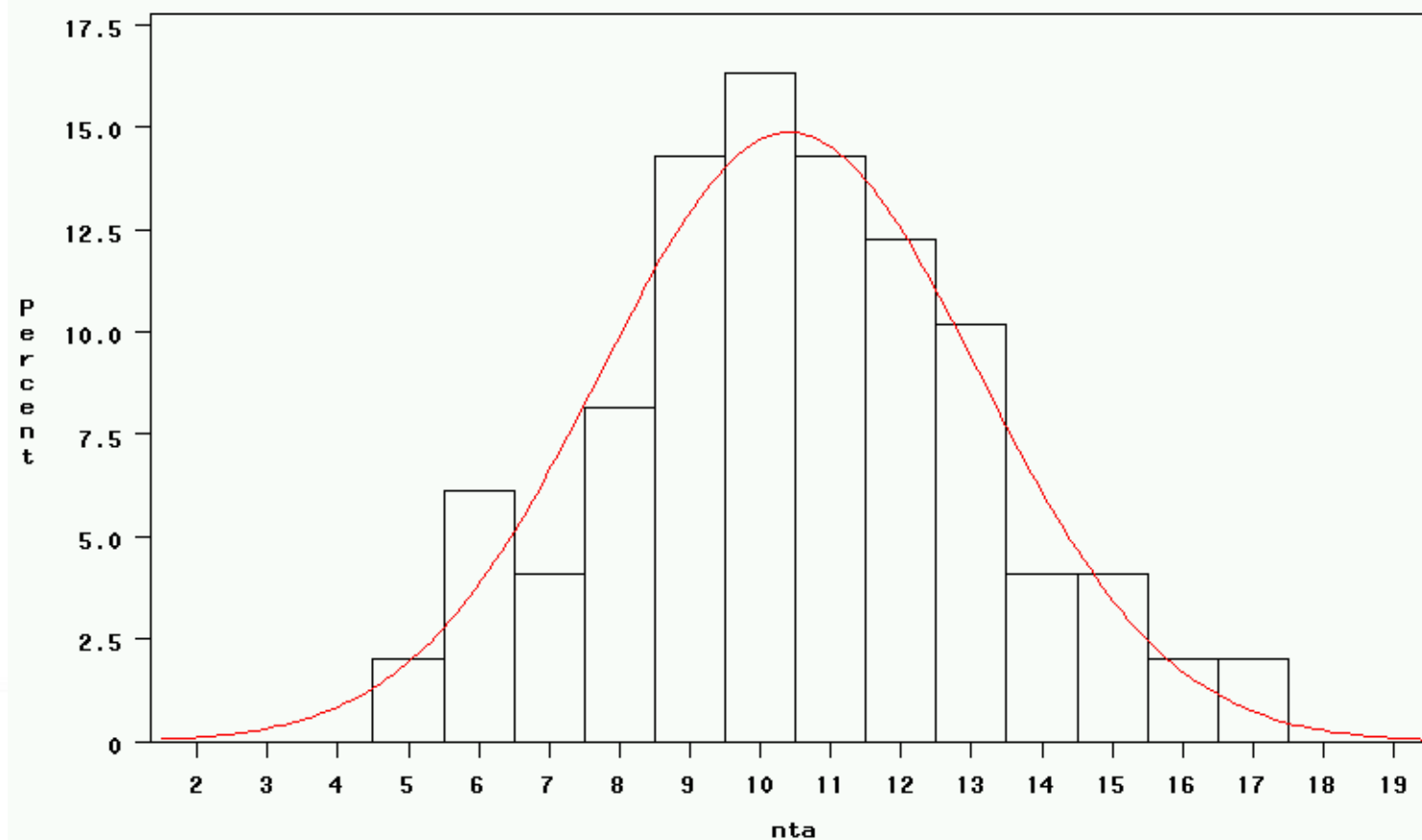
Stem Leaf	#	Boxplot
16 5	1	
16 0	1	
15		
15		
14 55	2	
14 00	2	
13		
13 00000	5	
12		
12 0000	4	+-----+
11 55	2	
11 000000	6	
10 5	1	
10 0000000	7	*---+---*
9 5	1	
9 00000	5	+-----+
8 55	2	
8 0000	4	
7		
7		
6 55	2	
6 0	1	
5 55	2	
5 0	1	
		-----+-----+-----+-----+



Stem Leaf	#	Boxplot
18 00	2	
17		
16 5	1	
15 005	3	
14 0005	4	
13 0005	4	+-----+
12 0000005	7	
11 0000	4	*-----*
10 0005	4	+
9 0005	4	
8 00005	5	+-----+
7 0000	4	
6 00	2	
5 0	1	
4 0	1	
3 55	2	
2 0	1	
		-----+

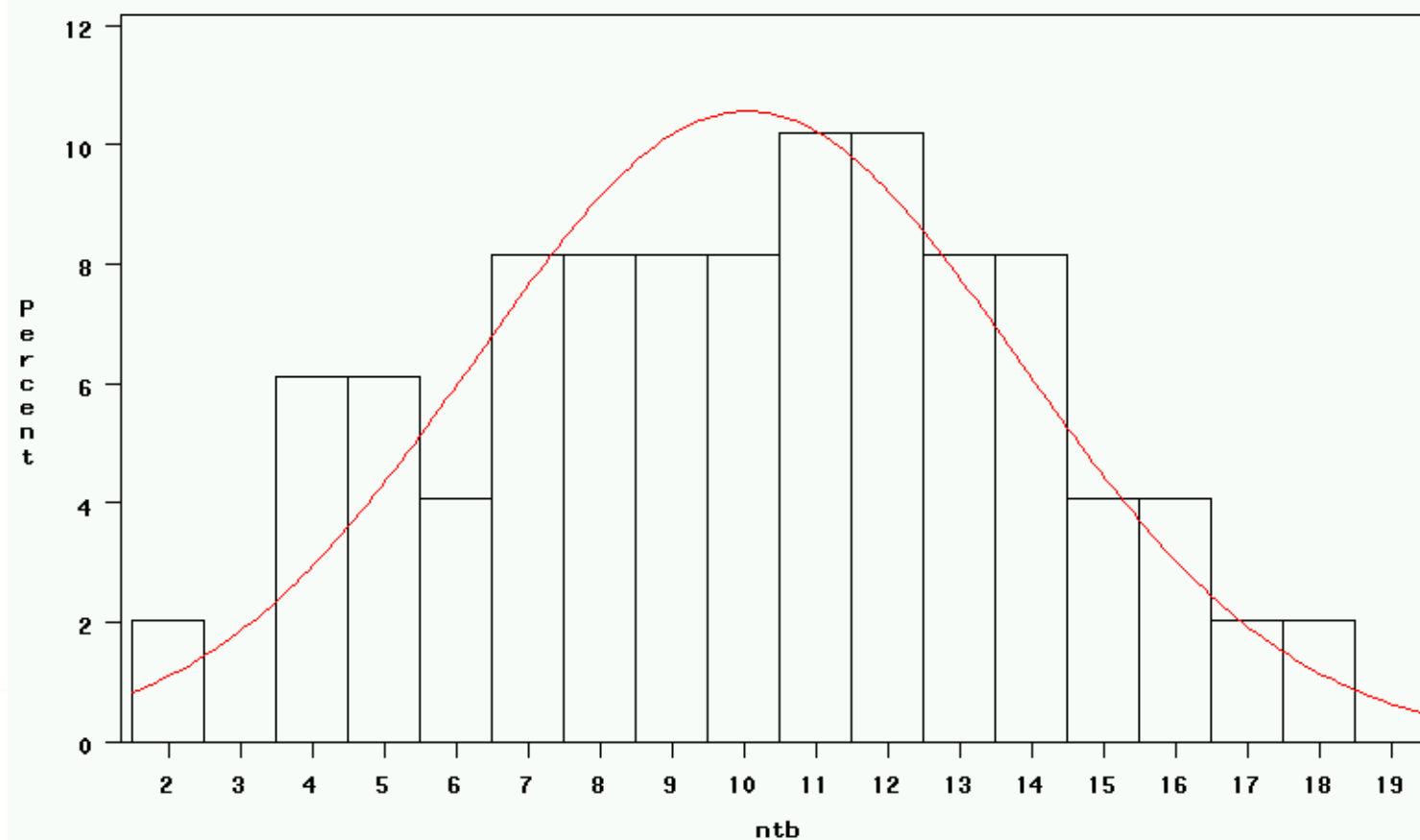


stat de base sur les notes de la classe A





stat de base sur les notes de la classe B





Une proportion est aussi une moyenne



Anglicistes	Germanistes
63	37
63%	37%

$$PA = \frac{63}{63+37}$$

Supposons que nous connaissions pour chacune des personnes la langue apprise

**Si la première de la liste apprend l'anglais, je tape 1
la seconde l'allemand, je tape 0, et ainsi de suite**



$$\overline{X_A} = \frac{1+0+0+1+1+0+\dots+1}{100} = 0,63$$

**Cela ressemble bien à une
moyenne**



On relève les salaires dans deux entreprises A et B

Le salaire moyen dans A est égal à 15

Le salaire moyen dans B est égal à 14

En conclusion A paye mieux que B

Mais nous savons autre chose



Effet de structure

	Hommes	Femmes	SALH	SALF
A	75	25	16	12
B	25	75	17	13
salaire moyen A			$((75*16) + (25*12)) / 100 = 15$	
salaire moyen B			$((25*17)+(75*13))/100 = 14$	

En moyenne, l'entreprise B paye moins bien

Pourtant les hommes et les femmes sont mieux payés en B
qu'en A



Si dans l'entreprise B, la répartition des sexes était celle de l'entreprise A, le salaire moyen y serait de 16

Le salaire réel étant de 14, l'effet de la répartition est de réduire la moyenne de $14/16$

Si dans l'entreprise A, la répartition des sexes était celle de l'entreprise B, le salaire moyen y serait de 13

Le salaire réel étant de 15, l'effet de la répartition augmente la moyenne de $15/13$



En conclusion :

La seule connaissance de la moyenne ne suffit pas pour

décider s'il vaut mieux être embauché en A qu'en B,

Du moins sur le seul critère de la rémunération



Des histoires de pharmaciens, de voitures, de chats et de chiens par habitants

Un tigre pour :

100 habitants

200

250

500

dans 4 régions également peuplées

C'EST FAUX !!!

Combien d'habitants par tigre, en moyenne sur tout le territoire ?

$$(100+200+250+500)/4 = 262,5$$



Le véritable calcul :

$$\frac{1}{4} (1/100+1/200+1/250+1/500) = 190$$

Dans la première région, en supposant qu'il y a 1000 habitants dans chacune d'entre elles, on compte :

$$1000/100 = 10 \text{ tigres}$$

Puis, $1000/200 = 5$ tigres, etc

En tout 21 tigres pour 4000 habitants, soient $4000/21 = 190$ habitants par tigre

C'est une moyenne harmonique



Des affaires de famille....

Dans une famille de n enfants, chaque enfant à $(n-1)$ frères et sœurs

Si la famille moyenne à n enfants, un enfant a-t-il $n-1$ frères et sœurs ?



Dans un échantillon de 100 familles, 50 ont 4 enfants, 50 aucun enfant

En moyenne, les familles ont deux enfants

$$(0,5 \times 4 + 0,5 \times 0) = 2$$

Chaque enfant a-t-il 1 frère ou une sœur en moyenne ?

Mais non, tous les enfants ont 3 frères ou soeurs



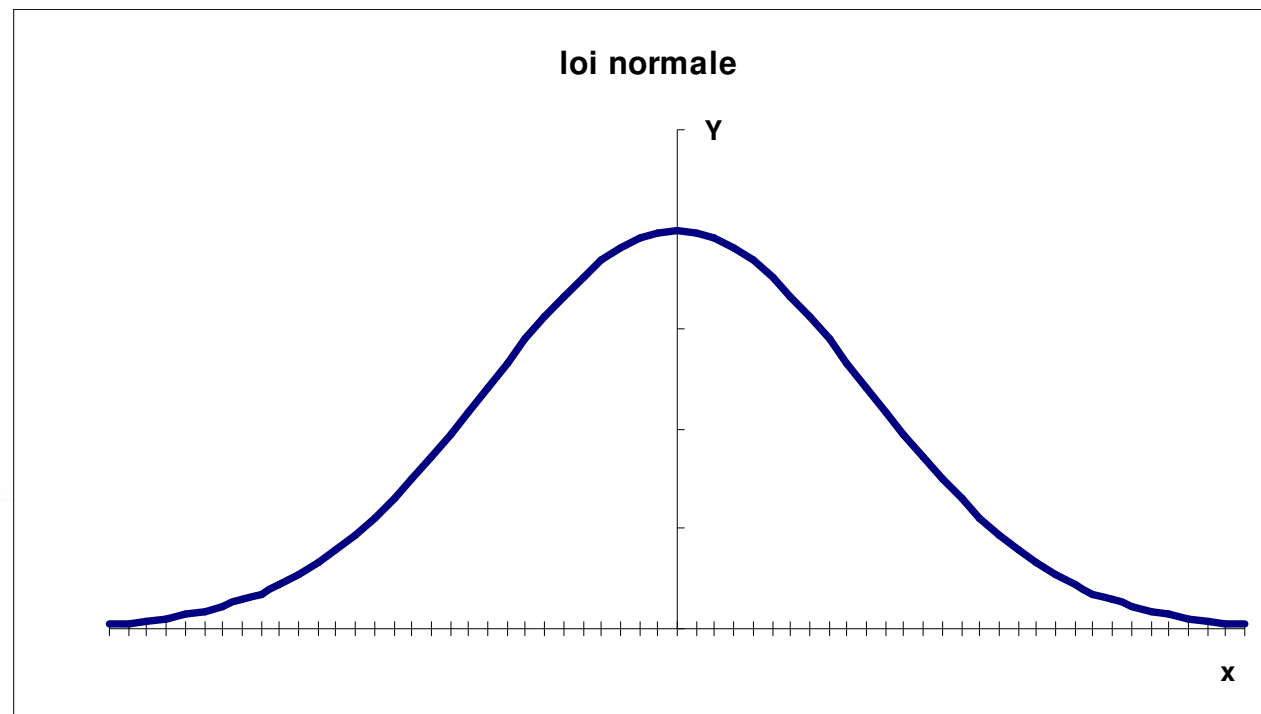
Petit détour par les sondages

Avant de venir, j'ai su quelle était la moyenne en math de toutes les classes de 3e de la région Rhône-Alpes , c'est 10,5 Puis, j'ai oublié, mais je sais par ailleurs que si on me donne la liste des moyennes de chaque classe, j'en aurai beaucoup autour de 10,5 et beaucoup moins autour de 5 ou de 15

Si je les représente sur un graphique, j'aurai quelque chose comme ça



On peut imaginer que la forme de la distribution sera celle-ci :





Vous vous souvenez de la moyenne de la classe A : 10,41
et celle de la classe B : 10,04

**Elles sont effectivement proches de la
moyenne générale de toute la région, et
leur différence n'est pas significative.**

Deux conclusions :

- **il est indifférent de choisir A ou B**
 - **« Il n'est pas nécessaire de manger le
bœuf entier pour savoir qu'il est coriace »**
- Samuel Jonhson**



Samuel Jonhson est cité par Wonnacott and Wonnacott

Les exemples sont tirés de Michel Louis Lévy :

« Comprendre les statistiques »