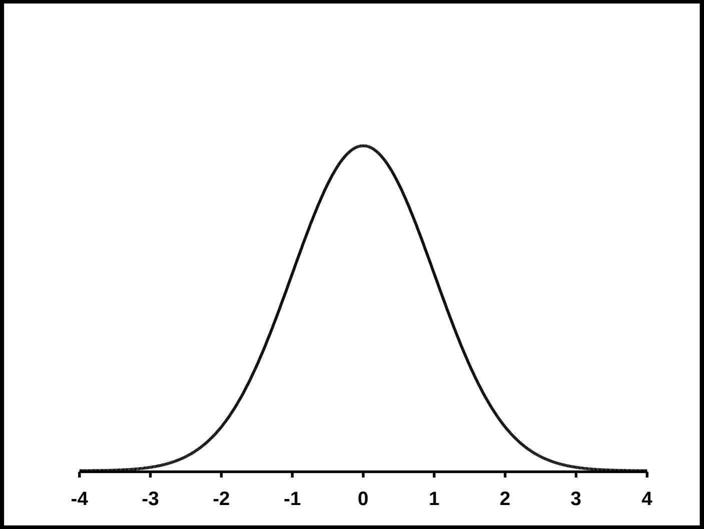
Chapitre 4

Préalables aux statistiques inférentielles



Le principe d'inférence

Etudes scientifiques ont souvent pour objectif de tirer des conclusions générales à propos de la relation entre variables

Ex: Etude sur le cannabis « les fumeurs de cannabis présentaient deux fois plus souvent des symptômes psychotiques que les non-fumeurs. »

Le principe d’inférence

Etudes scientifiques

=> extrapoler les conclusion à partir d’un échantillon vers une population

Statistiques inférentielles

= outils pour éviter les généralisations abusives

= les gardiens de l’inférence

Population - Echantillon

<u>Population</u> = ensemble des sujets sur lesquels on aimerait tirer une conclusion

Très souvent : impossible de mesurer toute la population

- Ressources limitées
- Population virtuelle

Population - Echantillon

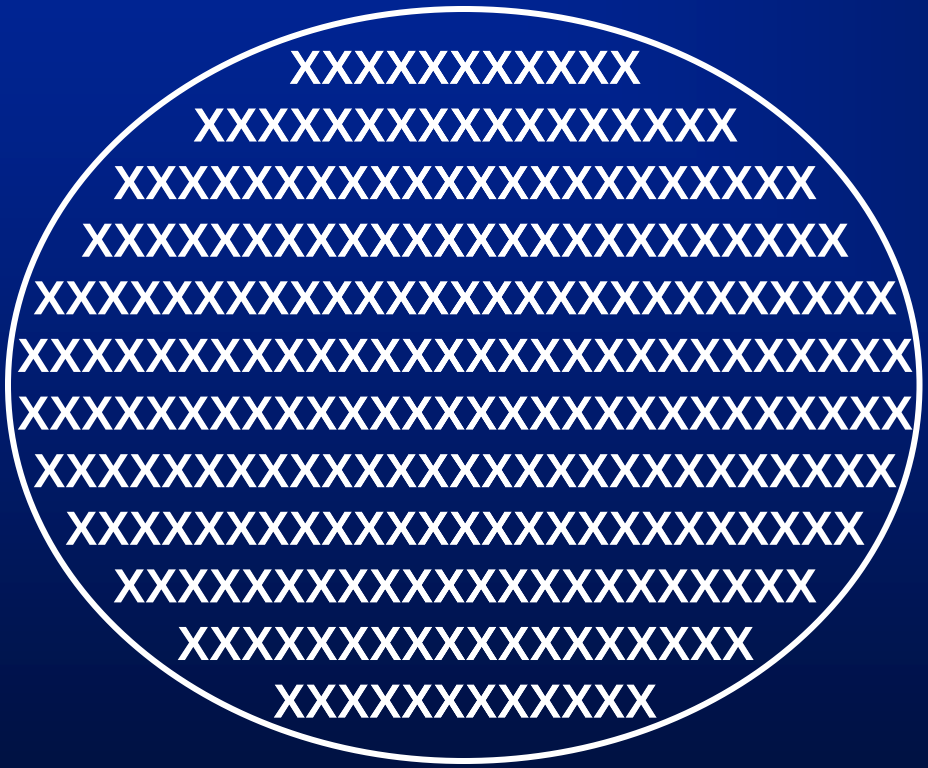
1. **Sélectionner un échantillon représentatif**

2. **Tirer une conclusion sur cet échantillon**

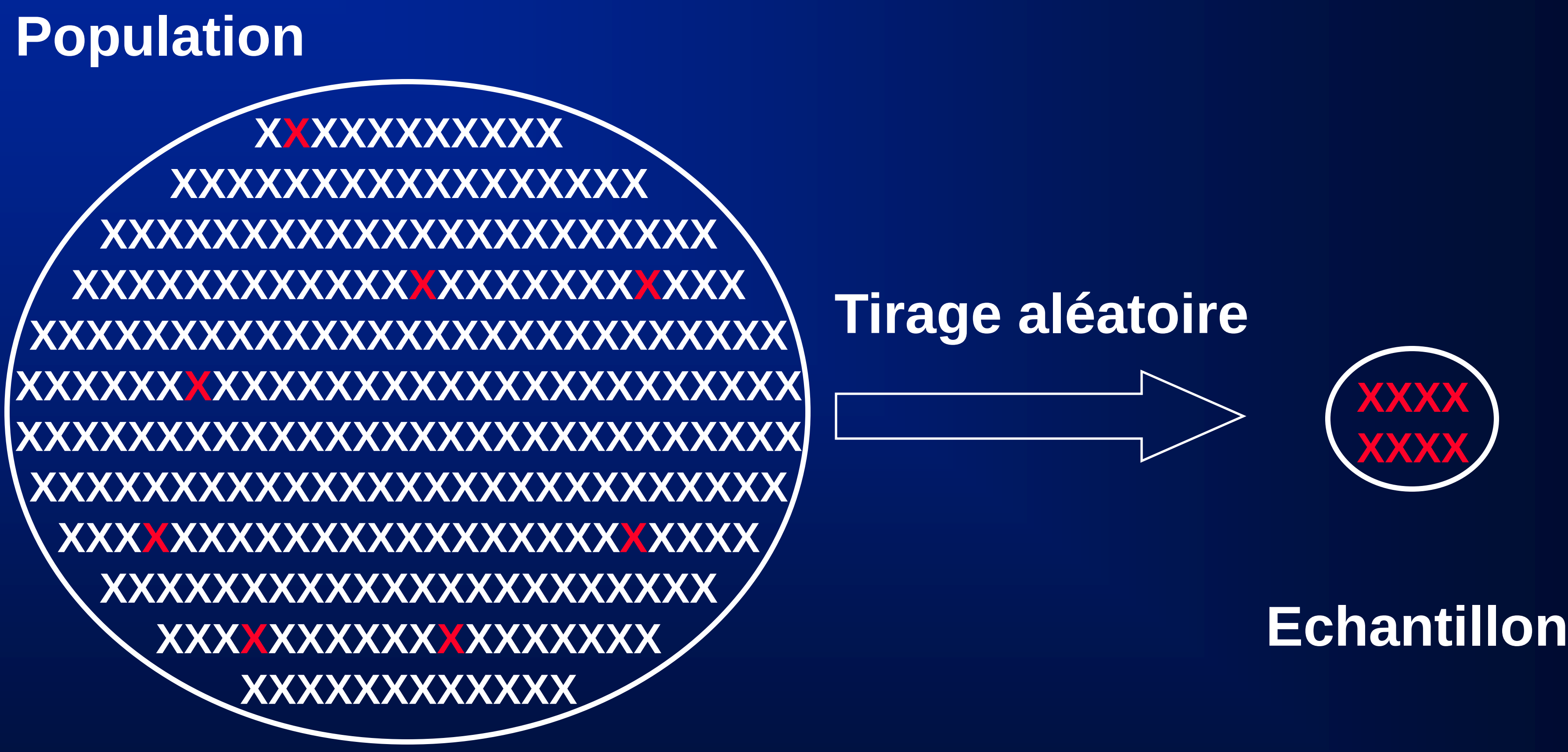
3. **Extrapoler la conclusion à la population**

Population - Echantillon

Population



Population - Echantillon



Population - Echantillon

<u>Idéalement</u>, un échantillon doit être rassemblé par une méthode de sélection aléatoire

= Chaque sujet à la même probabilité d'être sélectionné

Dans la pratique, très difficile (souvent : échantillon de convenance)

Terminologie

<u>Mesures effectuées sur un échantillon</u>

= statistiques d'échantillons

Exemple : $\overline{X}$ **S^2** **S**

<u>Mesures sur une population</u>

= paramètres de population

Exemple : μ σ^2 σ

Terminologie

Dans la majorité des études, les statistiques d'échantillon sont utilisées pour estimer les paramètres de population

=> Certaines particularités des formules

Formule de la variance

$$S^2 = \frac{\sum \left(X - \overline{X}\right)^2}{N - 1}$$

N – 1 sont les degrés de liberté

On perd un degré de liberté car il faut estimer la moyenne à partir du même échantillon

Estimation à partir de l’échantillon

La statistique d’échantillon sert à estimer le paramètre de population :

1. **<u>Estimation ponctuelle</u>**
   Précis mais grand risque d’erreur

2. **<u>Estimation par intervalle de confiance</u>**
   Moins précis mais moindre risque d’erreur

Population - Echantillon

<u>Statistique inférentielle</u> = outil mathématique permettant une extrapolation correcte

Eviter les extrapolations abusives liées à

“l’erreur d’échantillonnage”

L’erreur d’échantillonnage

= variabilité due au hasard d’échantillonnage

Ce qui est mesuré sur un échantillon peut être très différent de ce qui se passe dans la population

L’erreur d’échantillonnage

<u>Conséquences de l’erreur d’échantillonnage</u> :

1. Les mesures sur l’échantillon peuvent être peu représentatives

2. Plusieurs échantillons tirés dans une même population sont variables

L’erreur d’échantillonnage

L’erreur d’échantillonnage est fonction :

1. De la variabilité dans la population

2. De la taille de l’échantillon

L’erreur d’échantillonnage

Elle peut faire apparaitre des relations factices entre variables dans l’échantillon

Une apparente relation dans un échantillon ne traduit donc pas nécessairement une véritable relation entre ces variables

L’erreur d’échantillonnage

Population

OR = 1

	Dyslexique	Non dyslexique	Total
Droitier	7 200 (8%)	82 800 (92%)	90 000
Gaucher	800 (8%)	9 200 (92%)	10 000
Total	8 000	92 000	100 000

L'erreur d'échantillonnage

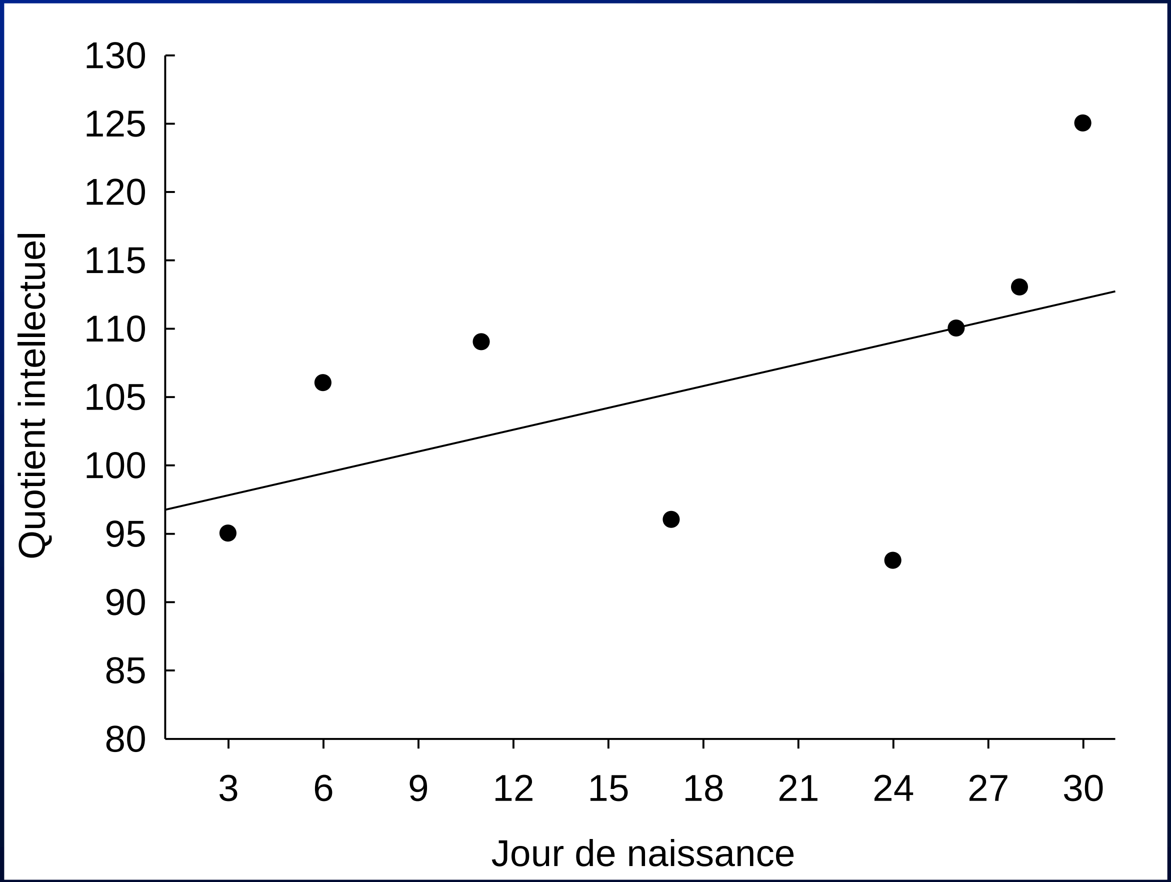
Echantillon

OR = 3,64

	Dyslexique	Non dyslexique	Total
Droitier	6 (7%)	80 (93%)	86
Gaucher	3 (21%)	11 (79%)	14
Total	9	91	100

L'erreur d'échantillonnage

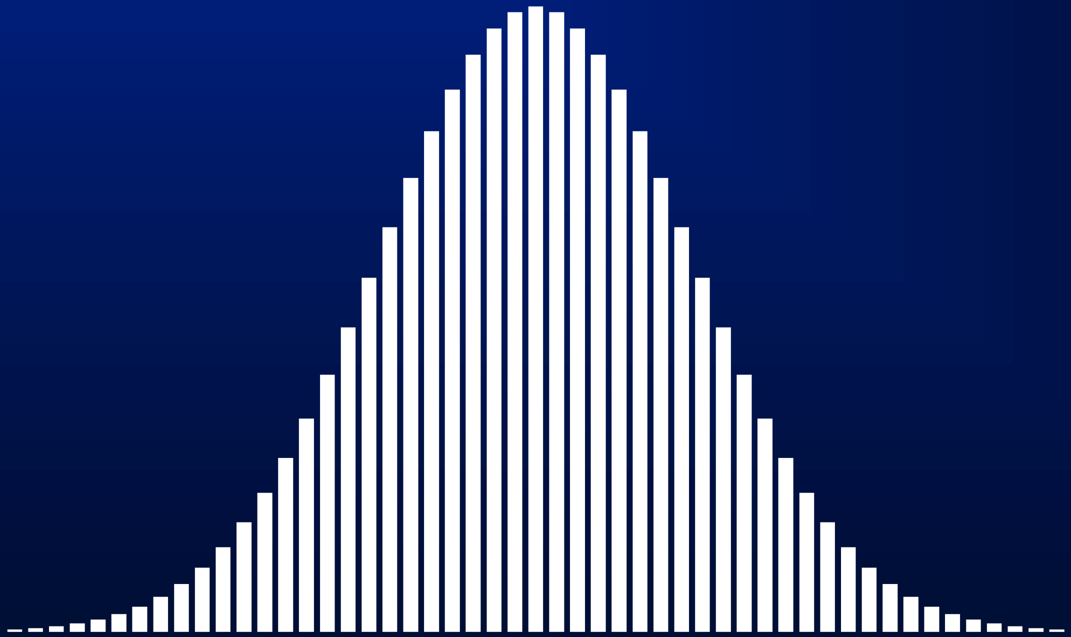
r = 0,51



La distribution normale

Distribution en forme de cloche

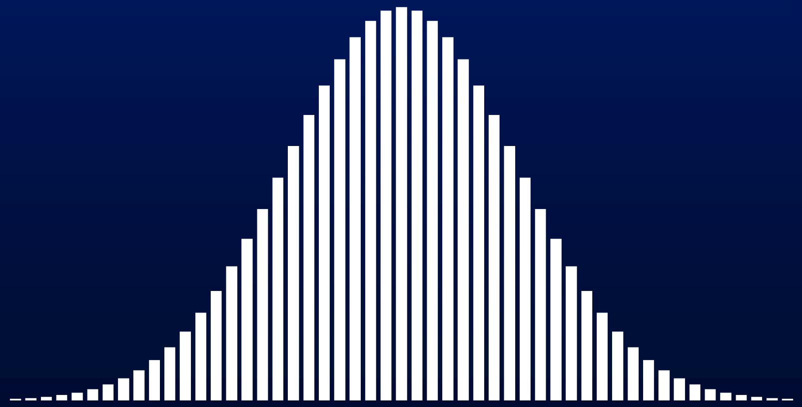
Distribution de Gauss



Importance de la distribution normale

Beaucoup de variables sont normalement distribuées

La plupart des procédures statistiques classiques impliquent la normalité

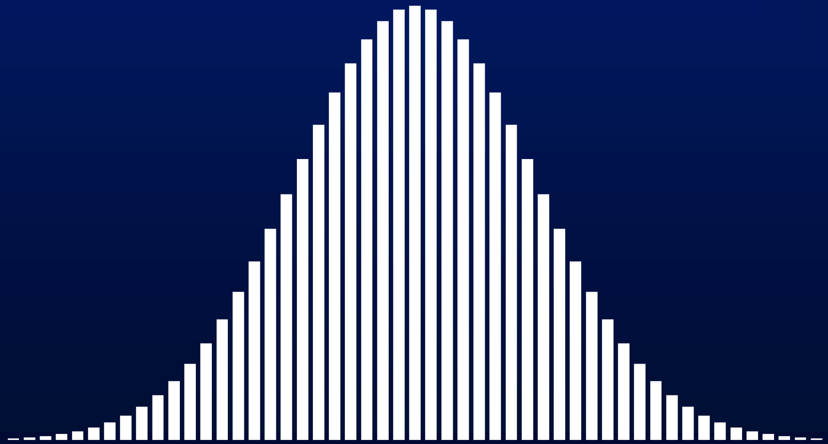


La distribution normale

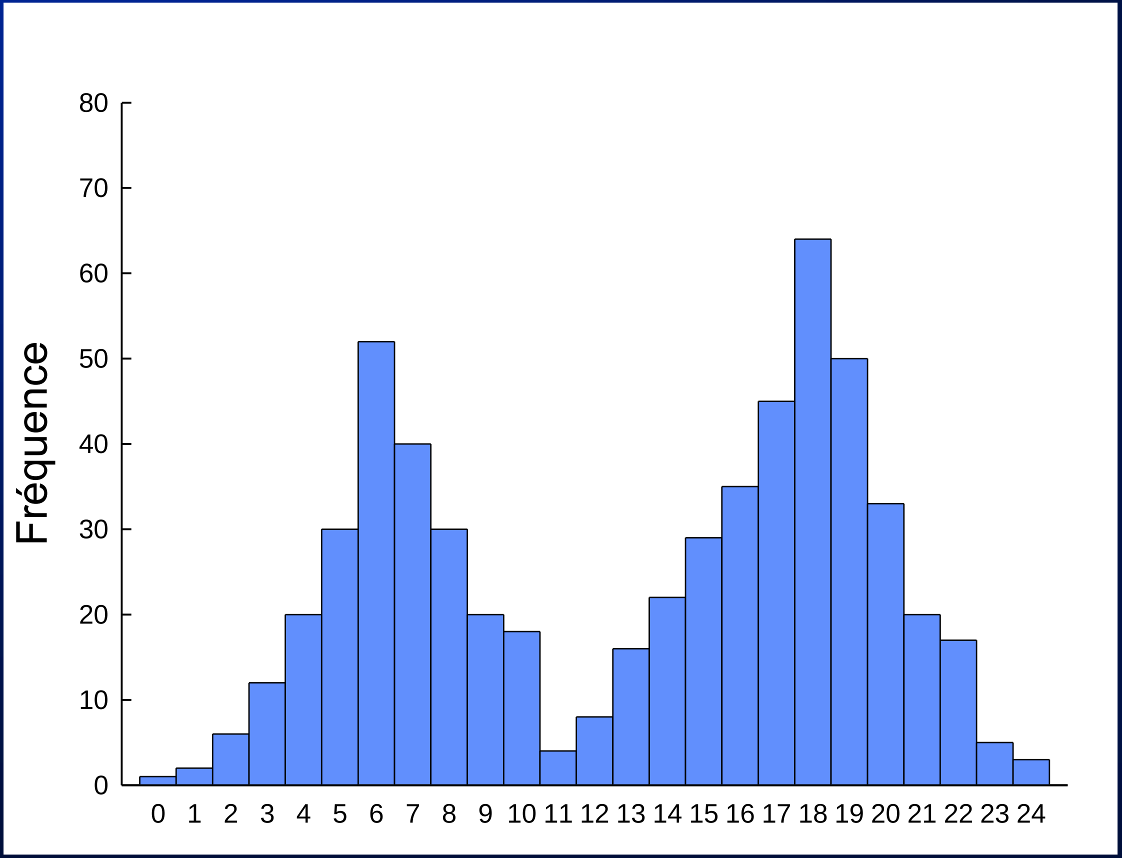
Distribution unimodale et symétrique

Voussure adéquate

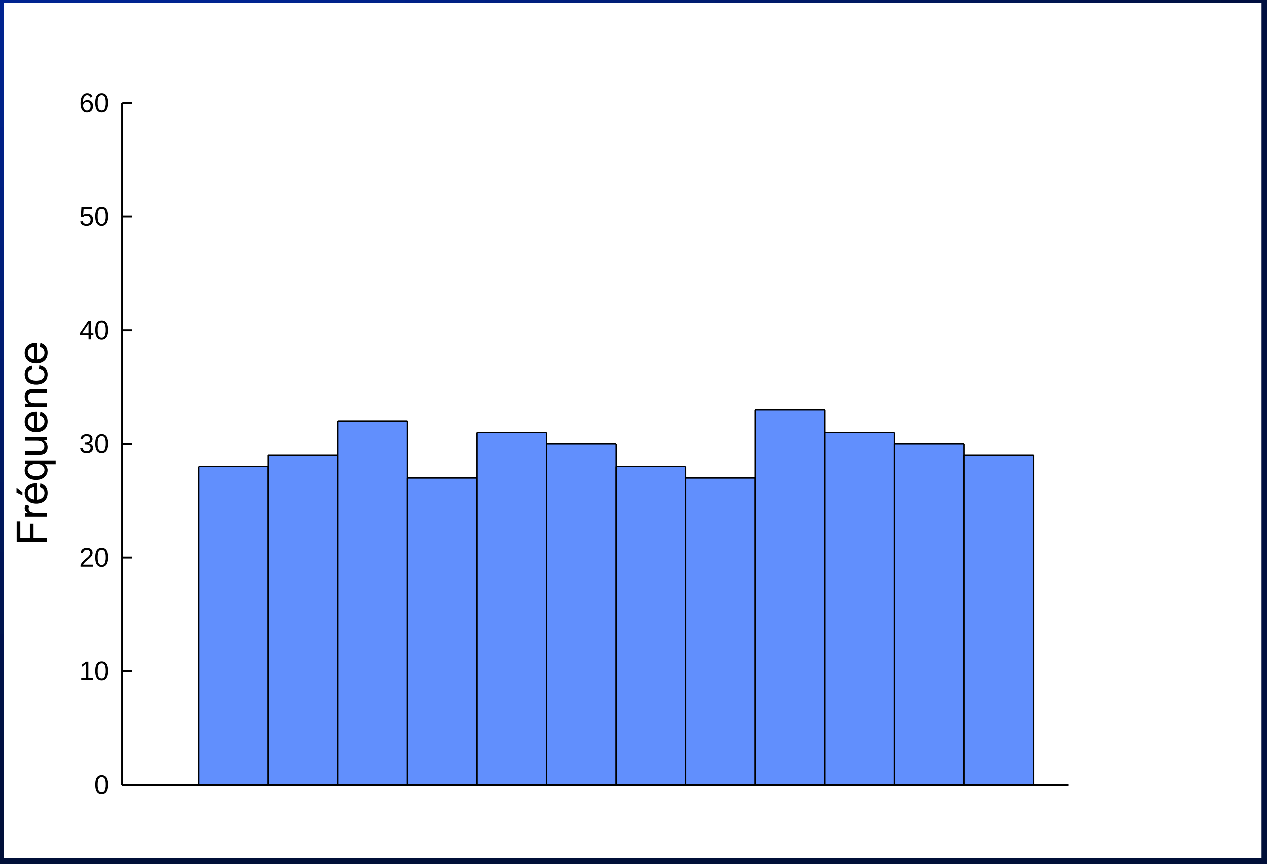
Mode = médiane = moyenne



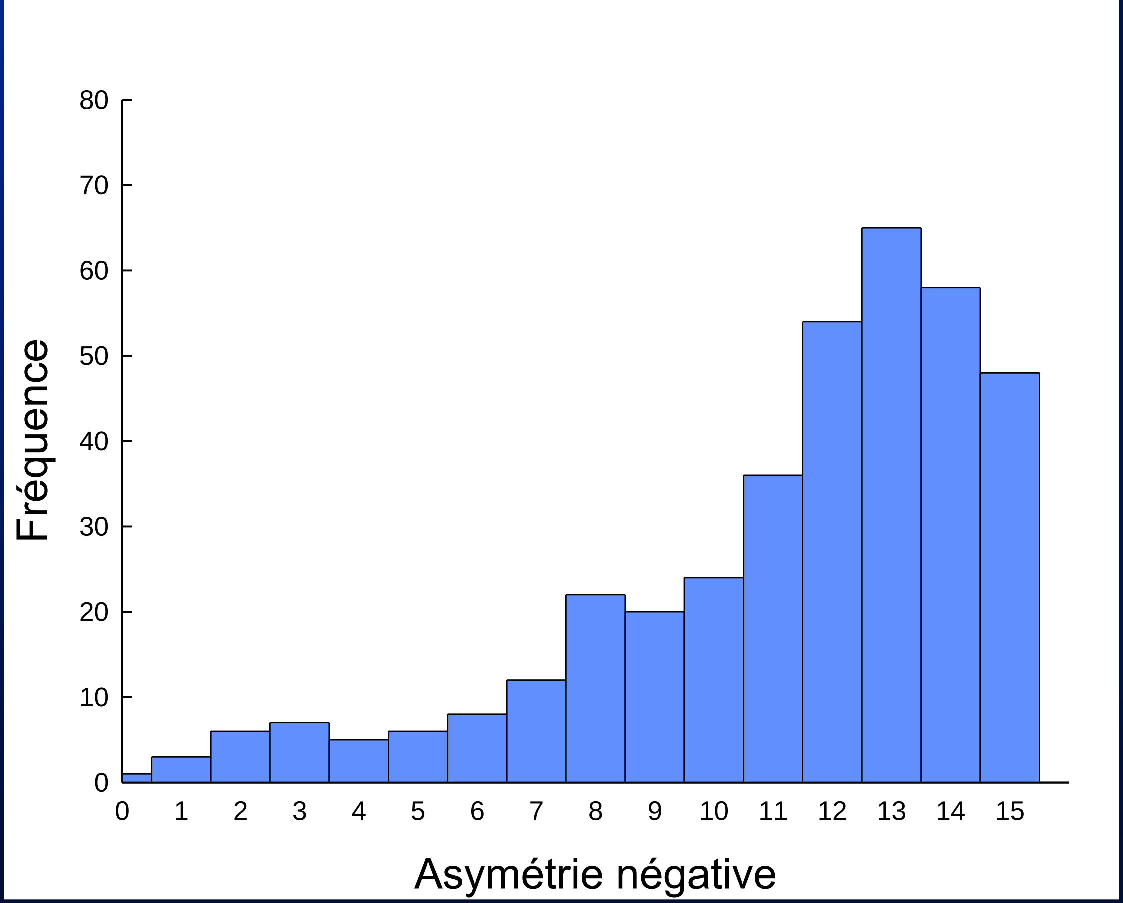
Distribution bimodale



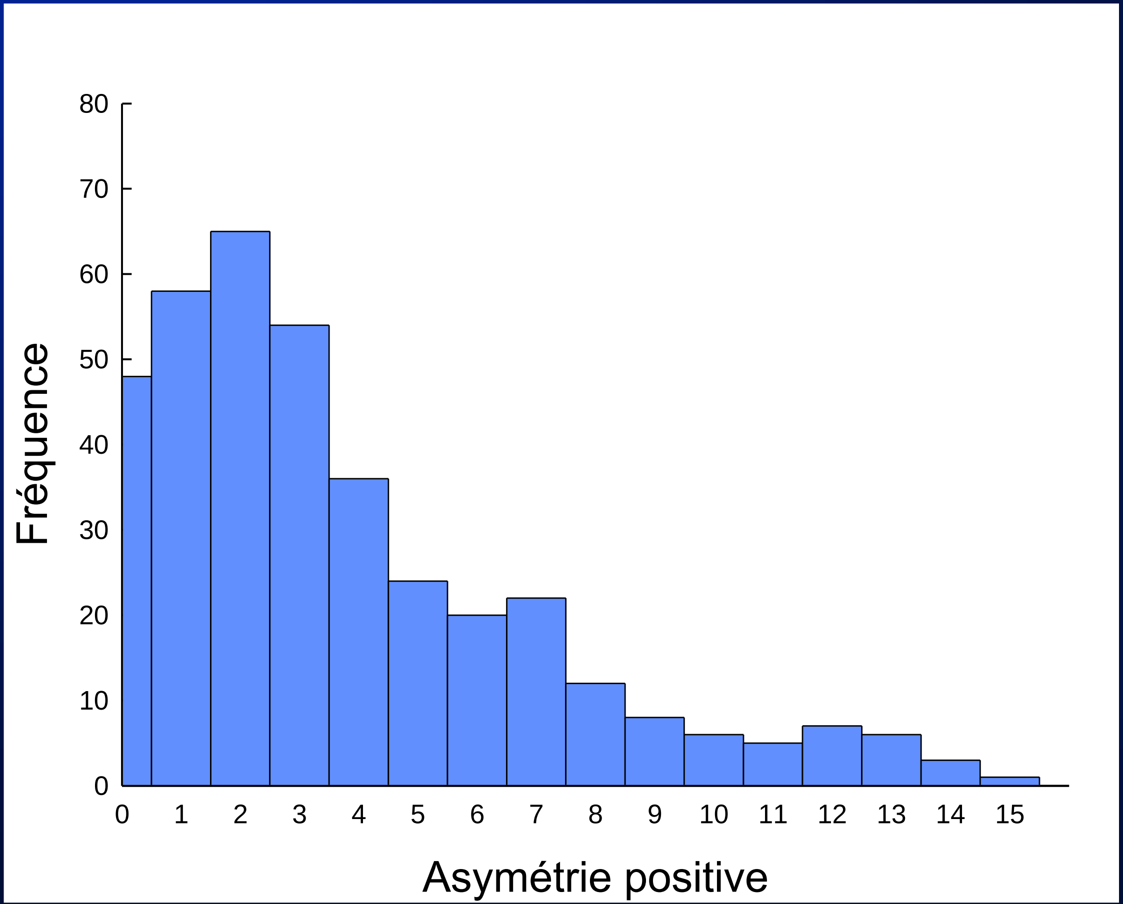
Distribution rectangulaire (uniforme)



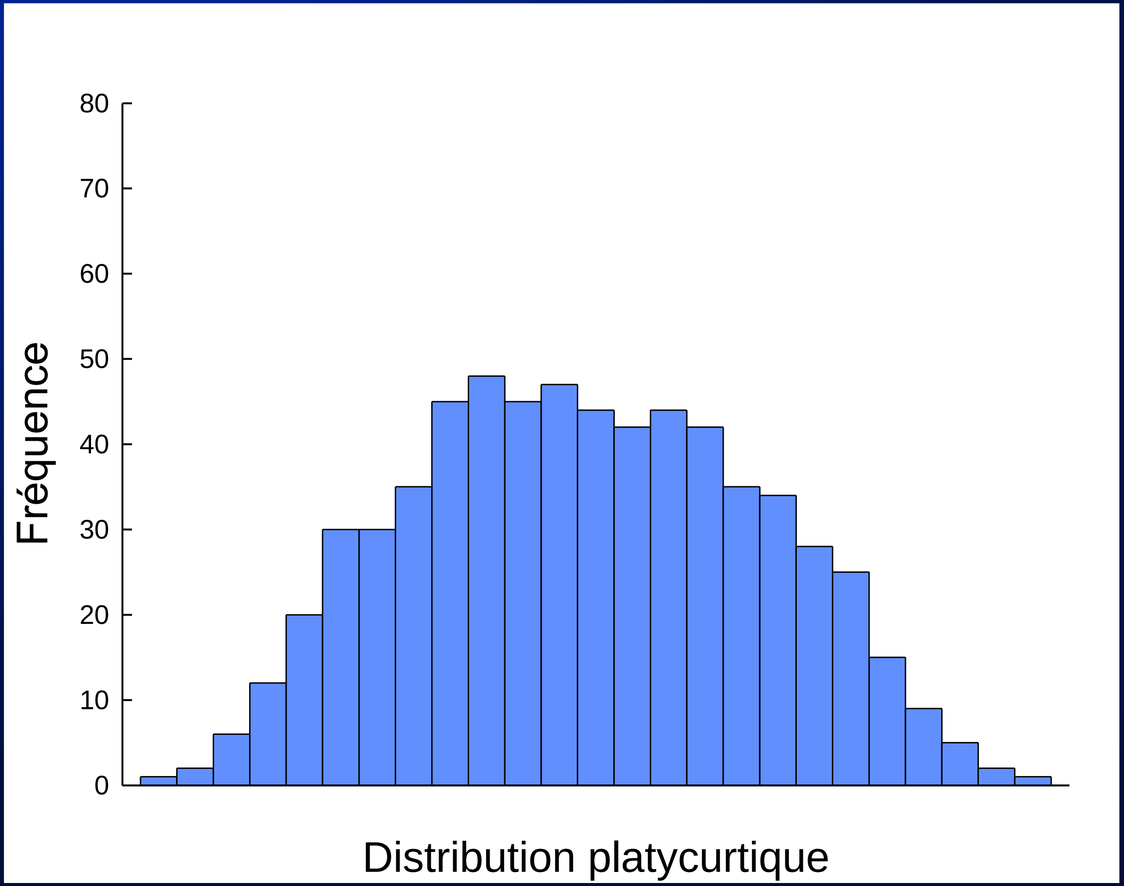
Asymétrie négative



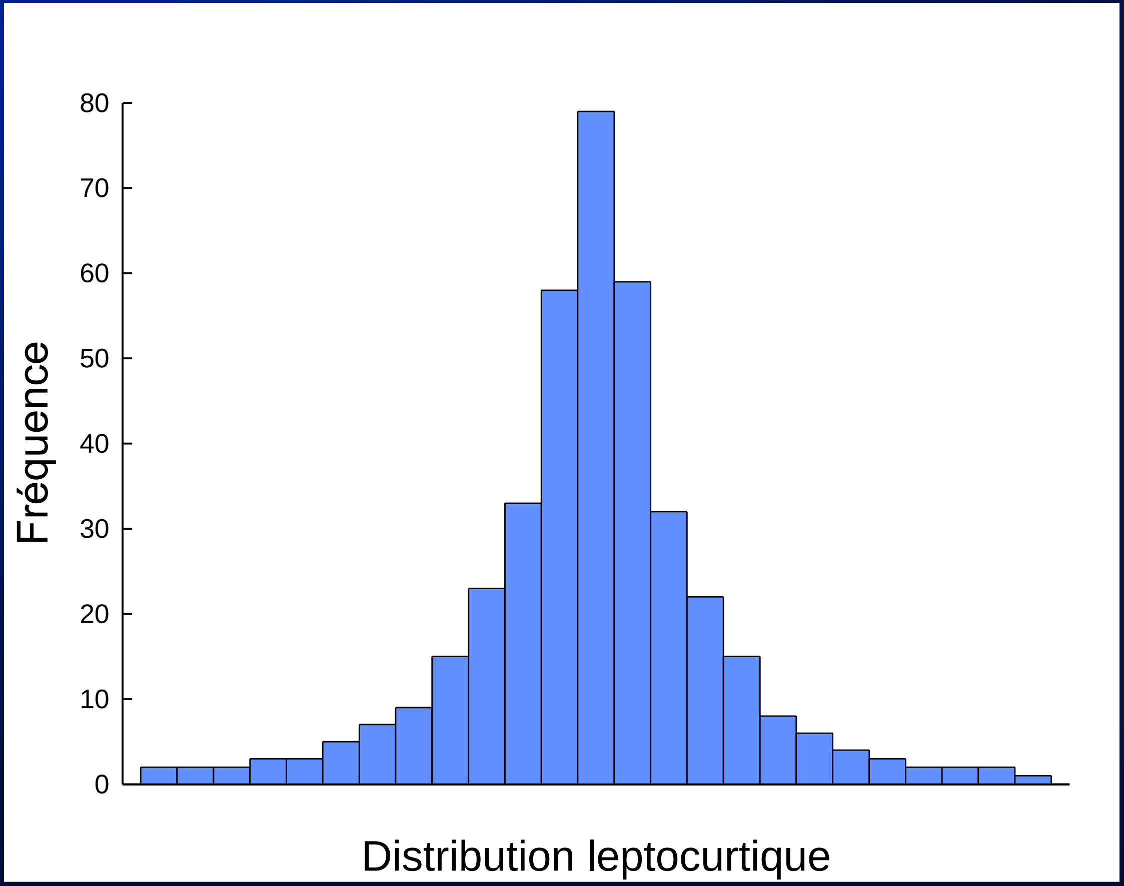
Asymétrie positive



Distribution platycurtique

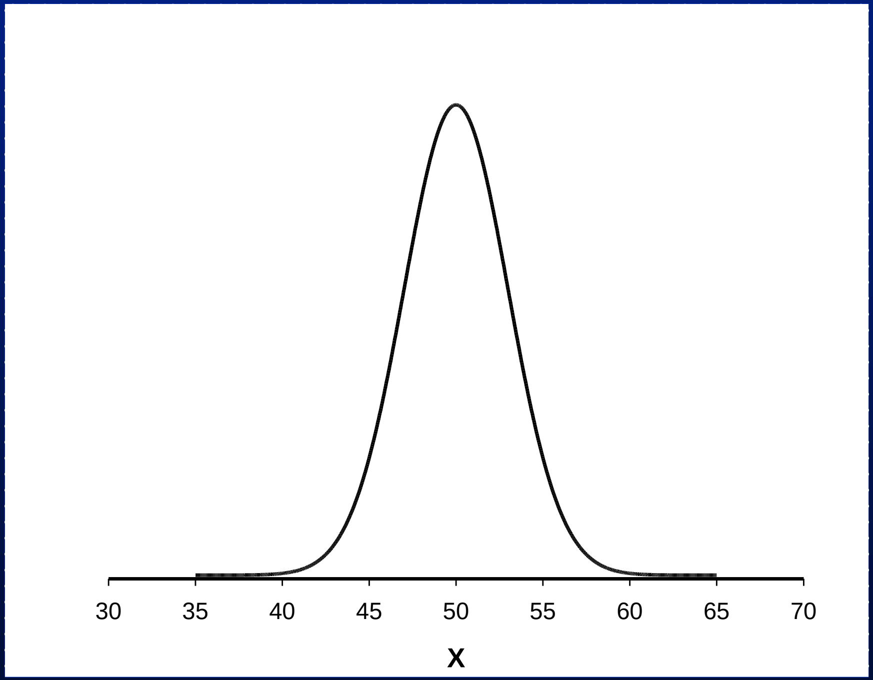


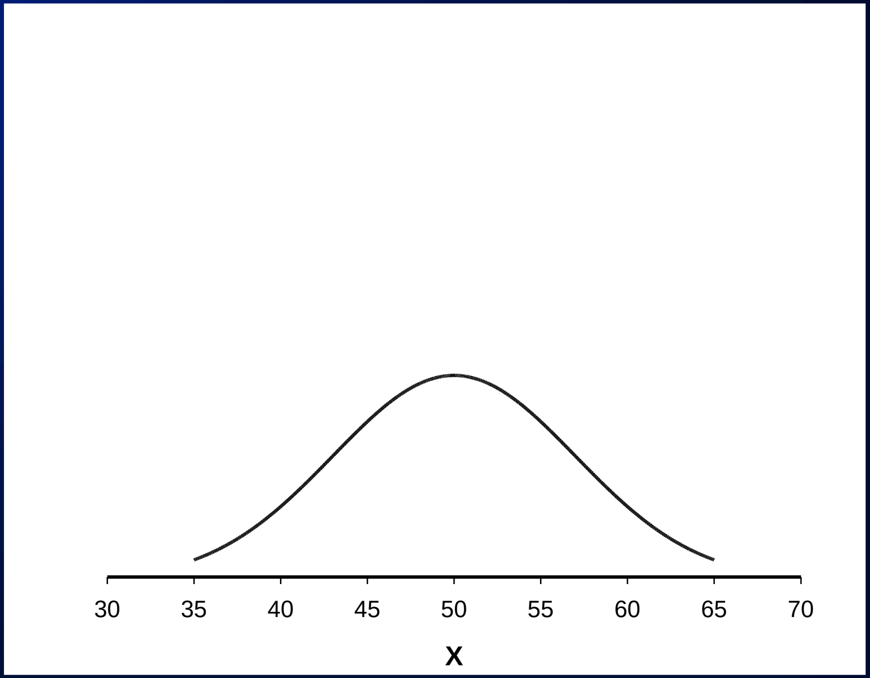
Distribution leptocurtique



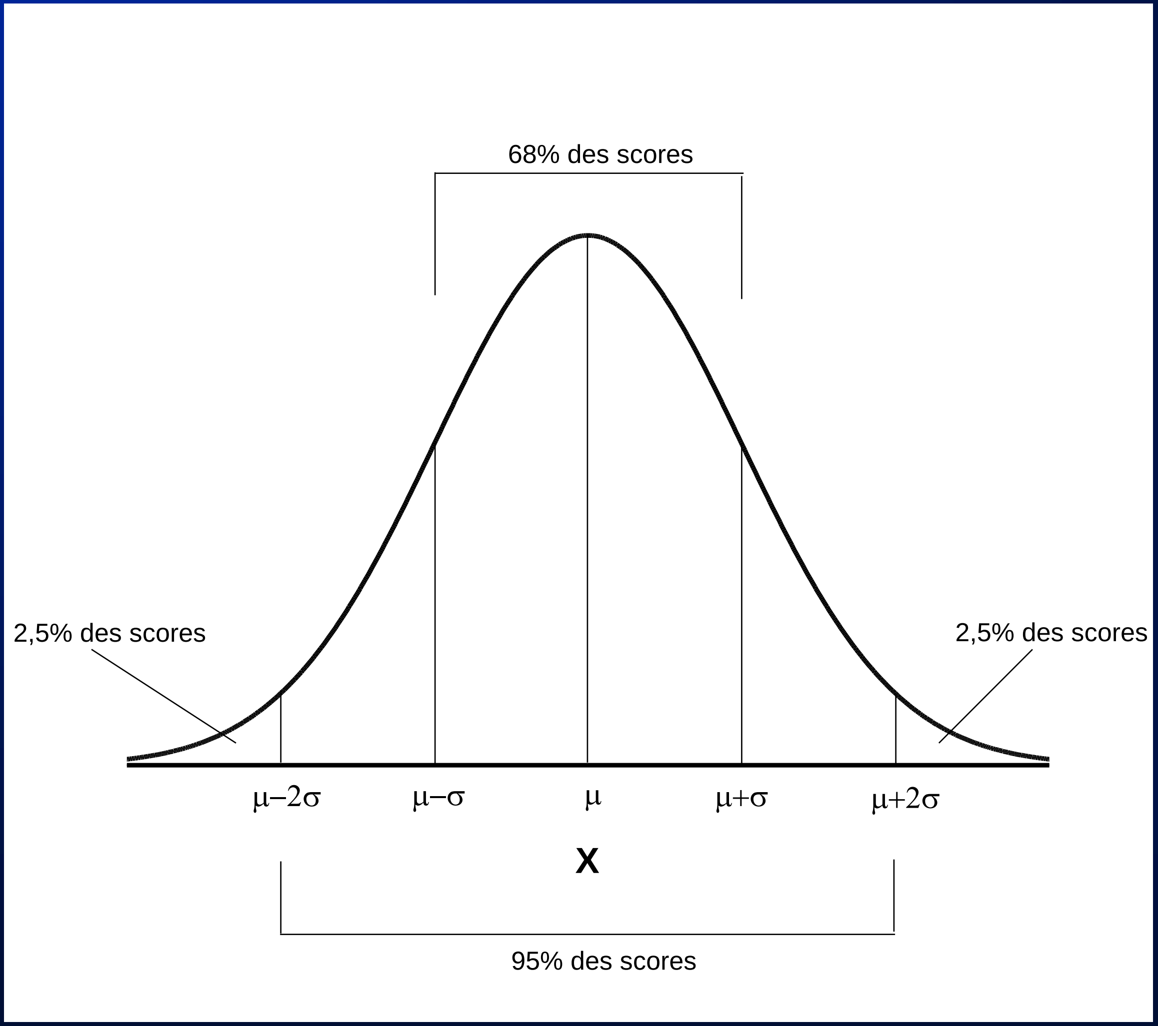
Une infinité de distributions normales

Elles diffèrent par leurs moyennes et écart-types

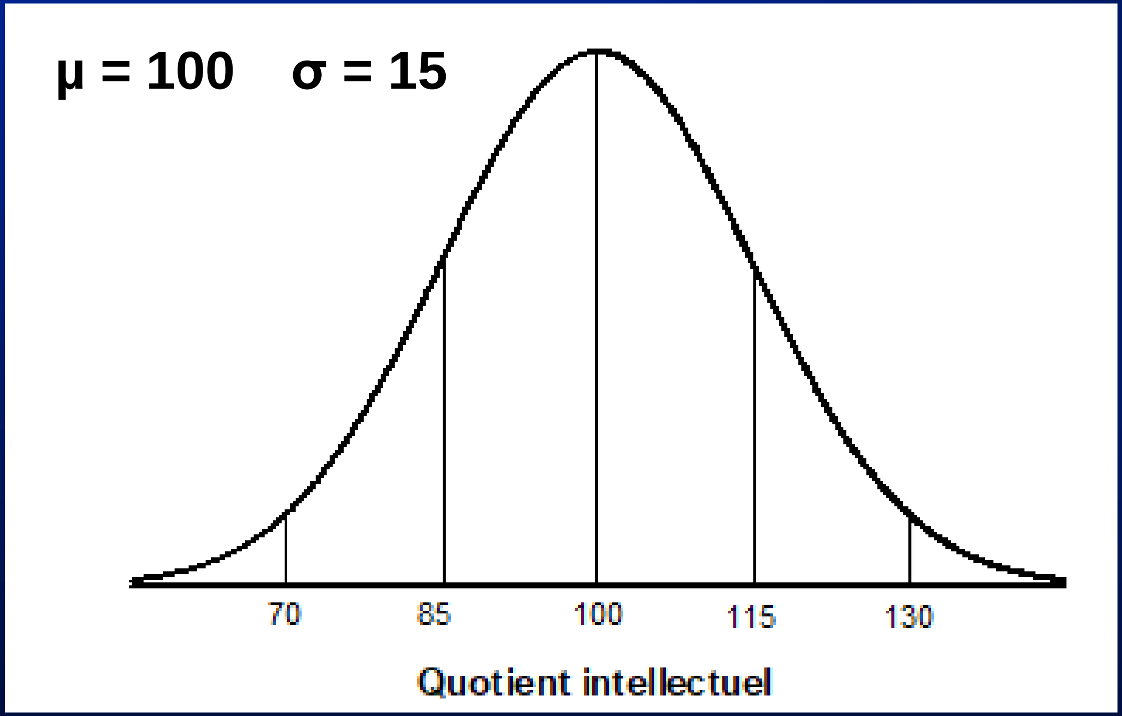




Mais des propriétés communes



Mais des propriétés communes



68 % entre 85 et 115
95 % entre 70 et 130
2,5% < 70
2,5% > 130

Standardisation

$$Z = \frac{X - \mu}{\sigma}$$

Score standardisé ou Score Z

Standardisation

Standardiser = exprimer les données en mesures d'écart-types:

<u>Exemple</u> **moyenne = 10** **$\sigma = 2$**

Score de 10 $\longrightarrow$ **Z = 0**

Score de 12 $\longrightarrow$ **Z = 1**

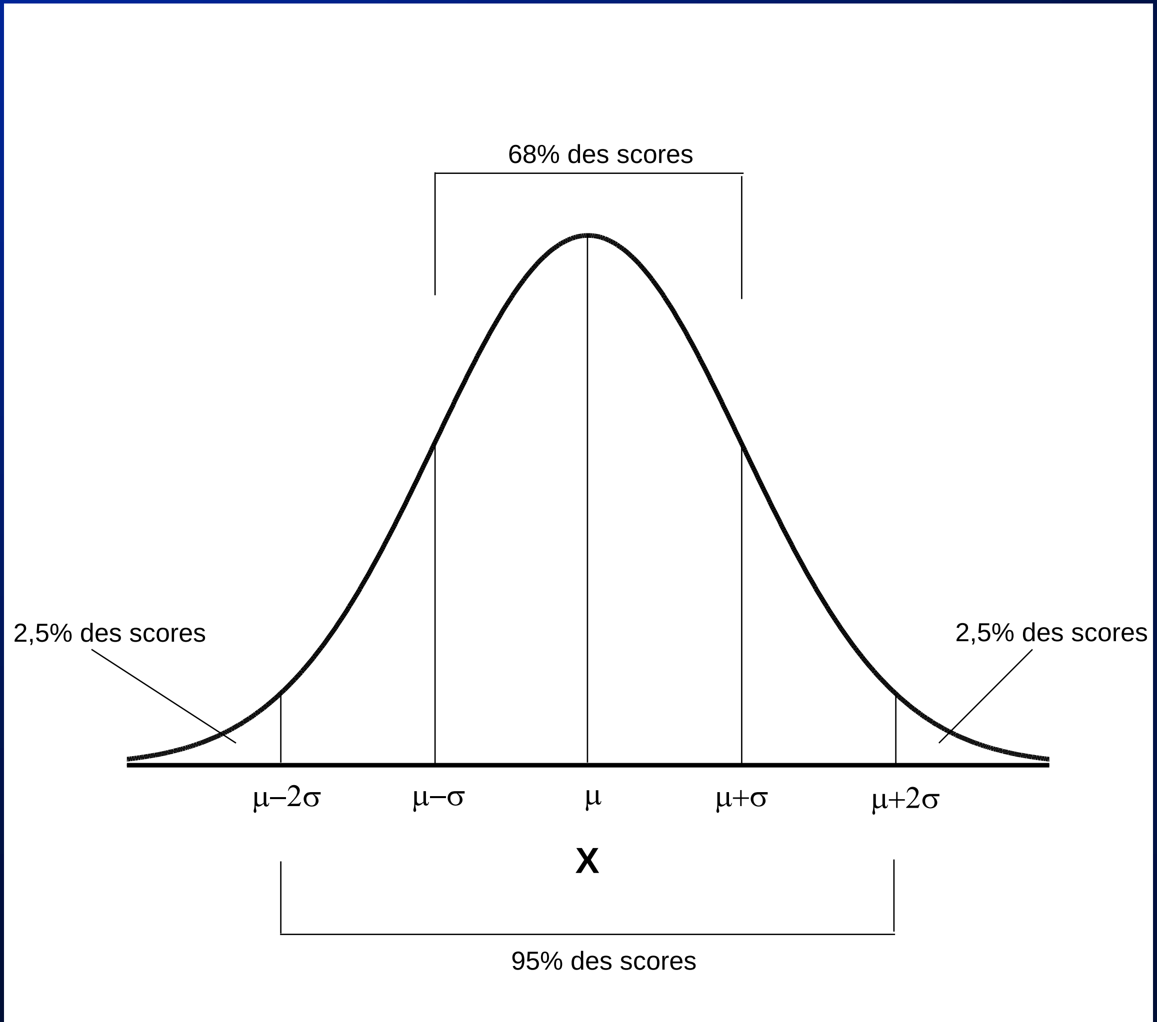
Score de 8 $\longrightarrow$ **Z = -1**

Propriétés des scores Z

$$Z = \frac{X - \mu}{\sigma}$$

- Le signe (+ ou -) détermine si > ou < moyenne
- = valeur réexprimée en écart-types
- \+ le score Z est grand, + l'observation est atypique

Standardisation



Calculer les scores Z

Mesure du QI (μ=100, σ =15)

Sujet	QI	score Z
Jean	112	0,8
Alfred	108	0,53
Jacques	95	-0,33
Claudine	122	1,47
Henry	98	-0,13

$$Z = \frac{X - \mu}{\sigma} = \frac{112 - 100}{15} = \frac{12}{15} = 0{,}8$$

Standardisation

Les scores Z sont utiles comme standards de comparaison

Exemple :

Un enfant obtient un score 82 en langage et un score de 23 en habileté psychomotrice

Standardisation

Les scores Z sont utiles comme standards de comparaison

<u>Exemple </u>:

Un enfant obtient Z = -1,5 en langage et Z = 1 en habileté psychomotrice

Standardisation

Les scores Z sont utiles comme standards de comparaison

Exemple :

Personne testée sur deux échelles d'anxiété:

X = 112

X = 11

Standardisation

Les scores Z sont utiles comme standards de comparaison

<u>Exemple</u> :

Personne testée sur deux échelles d'anxiété:

$Z = 1,2$ ($\mu = 100$ $\sigma = 10$)

$Z = 1$ ($\mu = 10$ $\sigma = 1$)

Score Z moyen = 1,1

Retrouver la valeur à partir du score Z

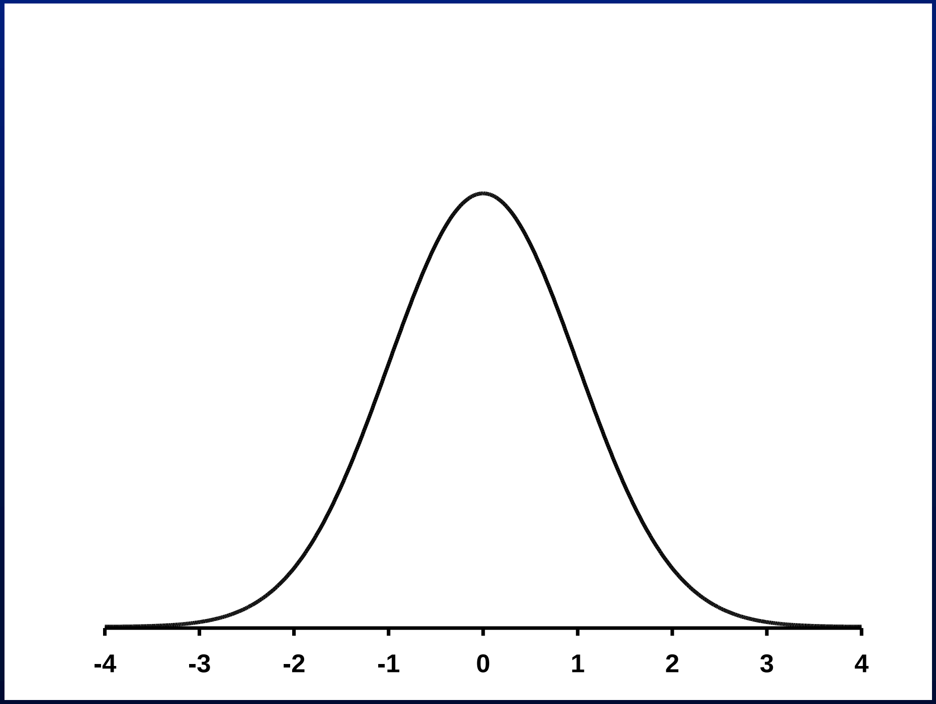
$$X = \mu + Z\sigma$$

Exemple: Score Z de 0,8 pour le QI

QI = 100 + (0,8 X 15) = 100 + 12 = 112

La distribution normale réduite

Standardiser tous les scores pour obtenir une distribution normale réduite :



$\mu = 0$
$\sigma = 1$

Distribution relative des scores reste inchangée