

Biostatistiques

Xavier Penin

Introduction

Mal aimées

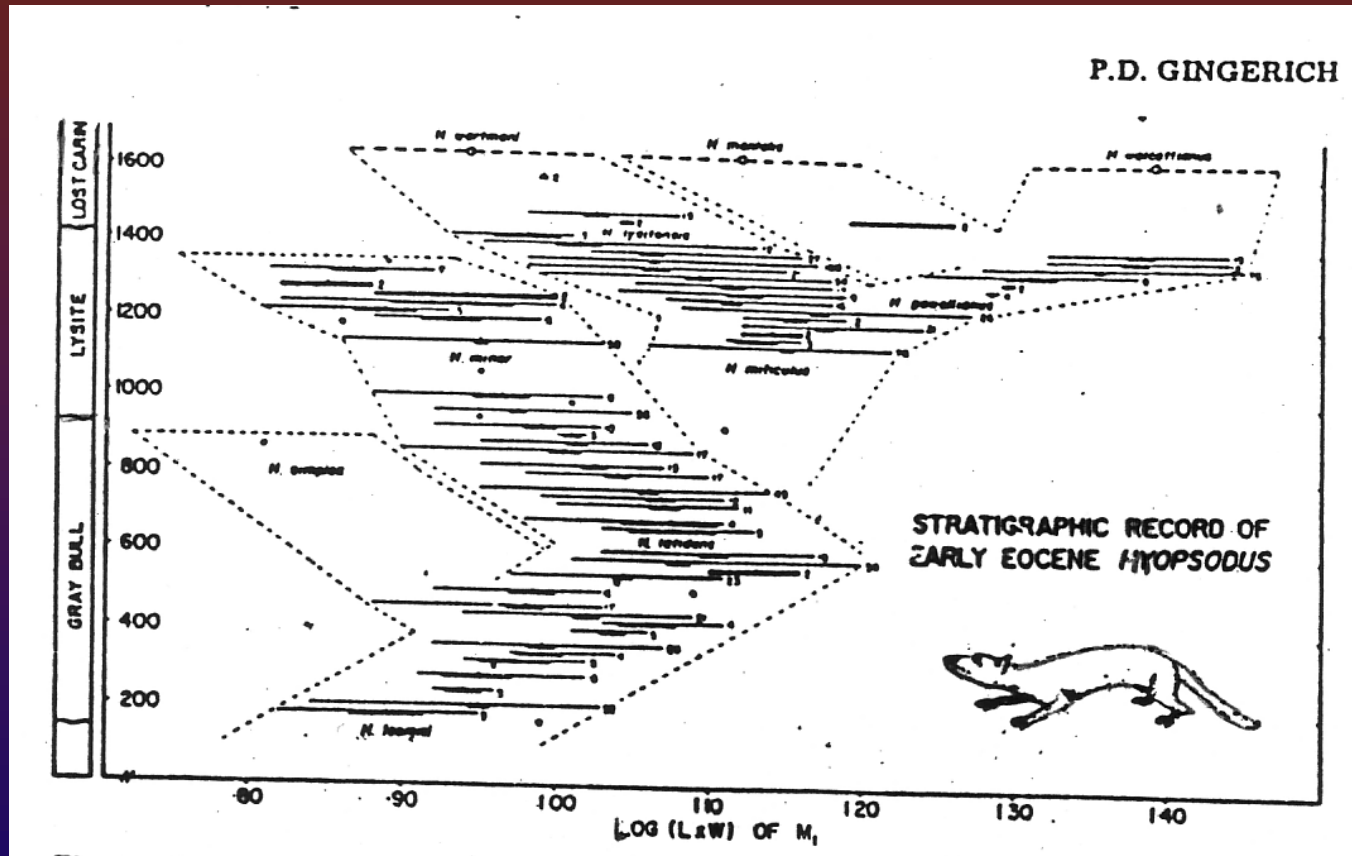
Il existe trois sortes de mensonge: les mensonges, les affreux mensonges et les statistiques.

B. Disraeli

Statistics, that gray area which is not quite a branch of mathematics – and just as surely a branch of science.

W. H. Press

Mais indispensables



Vivant = Variable

Une règle de base en biométrie



Une mesure isolée n'est
pas interprétable

L'échantillon est l'unité de base

L'échantillon est issu d'une population

La loi des grands nombres considère qu'avec l'augmentation des effectifs on peut trouver des valeurs numériques stables caractéristiques de la population malgré la variabilité.

Les estimateurs :

1. La tendance centrale
2. La dispersion

Mesures de tendance centrale

1. La Moyenne
ou espérance mathématique



$$\bar{y} = \frac{\sum_{i=1}^n y}{n}$$

2. La Médiane



Estimateur « Robuste »

3. Le Mode

Mesures de dispersion

La Variance



$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

L'Ecart-Type



$$s = \sqrt{s^2}$$

Qualités des estimateurs

1. Consistant
2. Sans Biais
3. Efficient

Note : La moyenne est plus efficiente que la médiane

La significativité

Les estimateurs décrivent la réalité mais ne permettent pas de savoir si les différences observées sont significatives.

Les embûches du

"Post hoc ergo propter hoc"

A la suite de cela donc à cause de cela

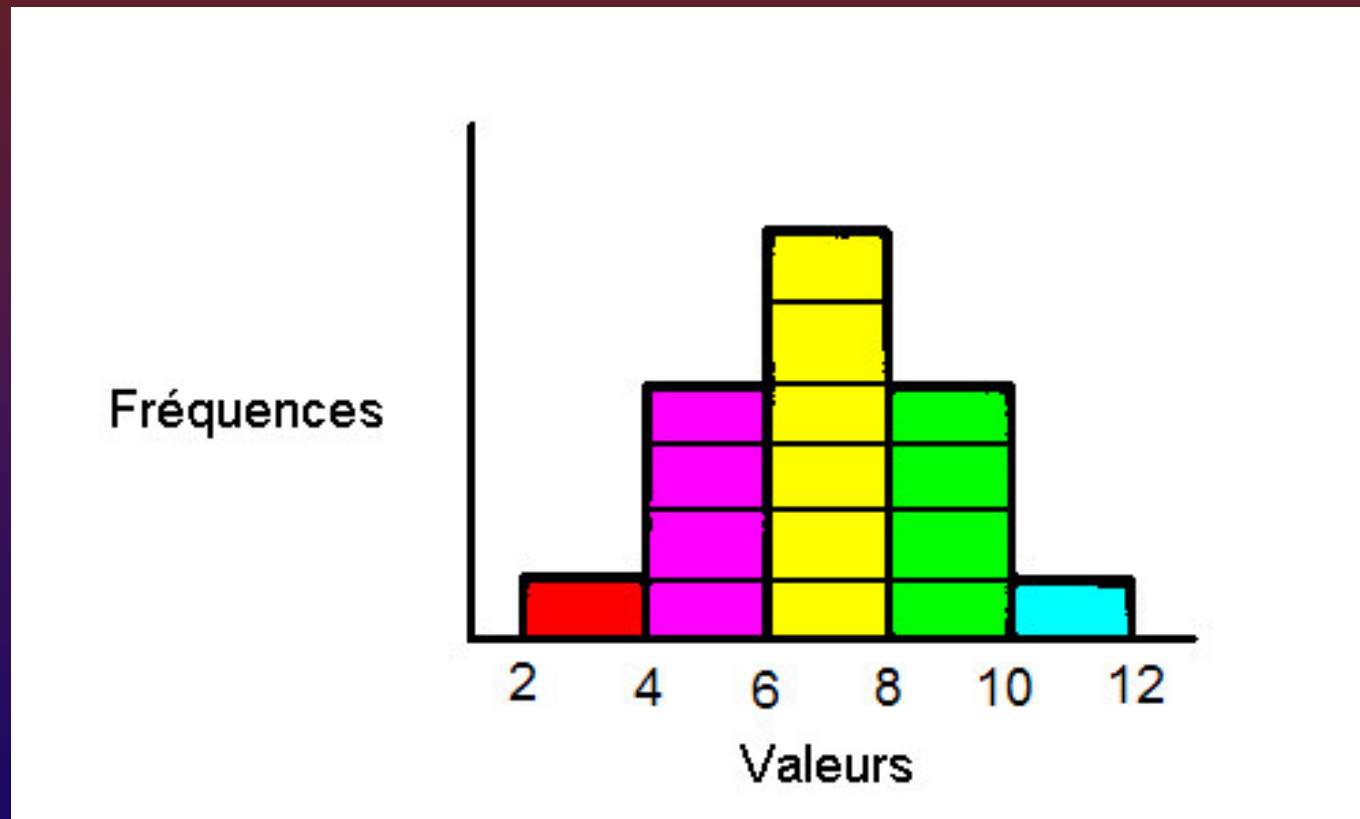
Descriptive  Inférentielle

Mesure  probabilité.

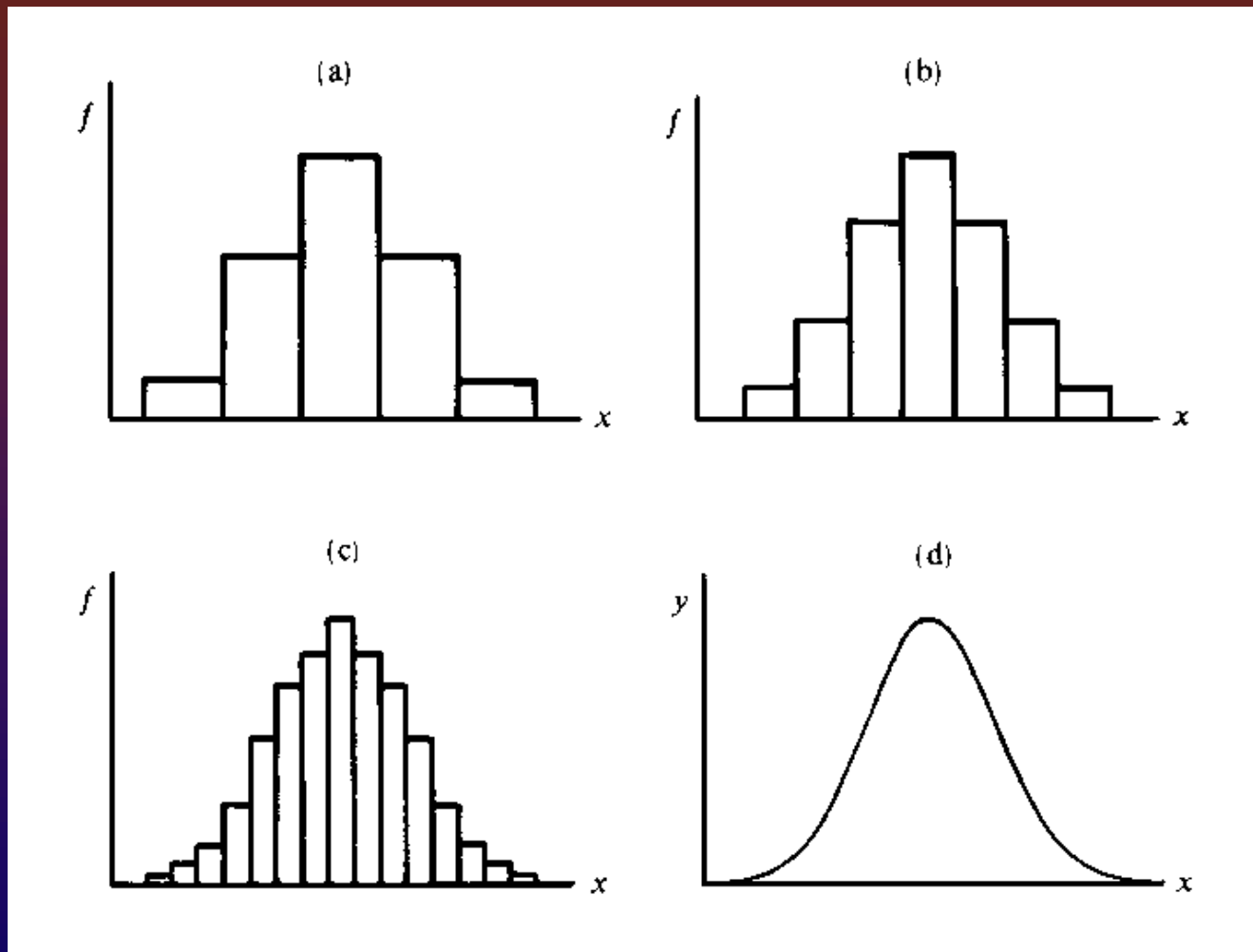
Certitude = 1 Impossibilité = 0

Pour calculer une probabilité il faut connaître la distribution

L'histogramme

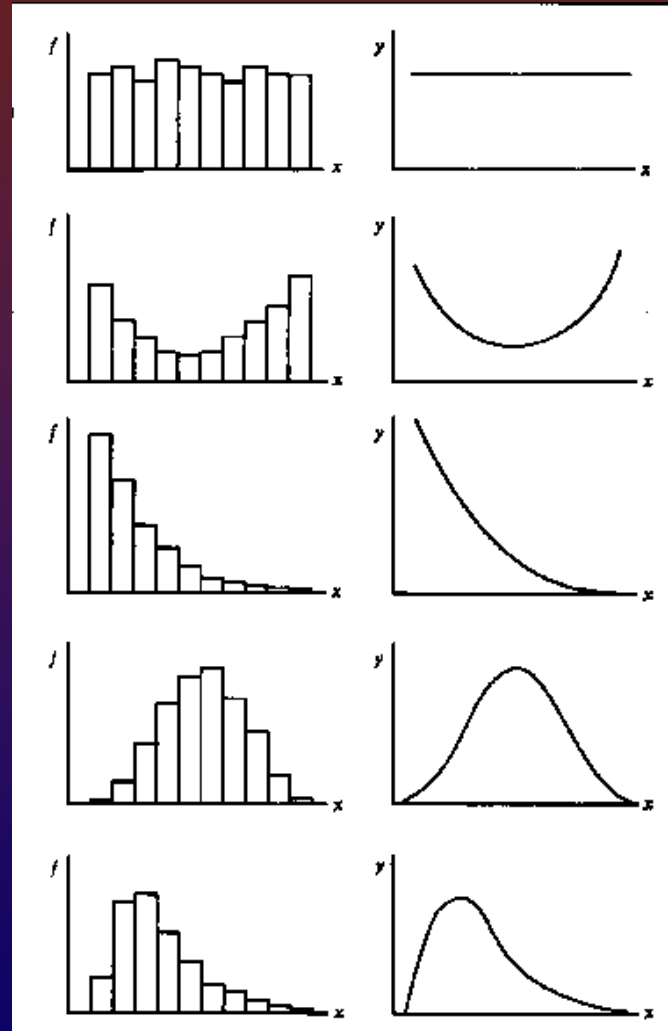


De l'histogramme à la courbe

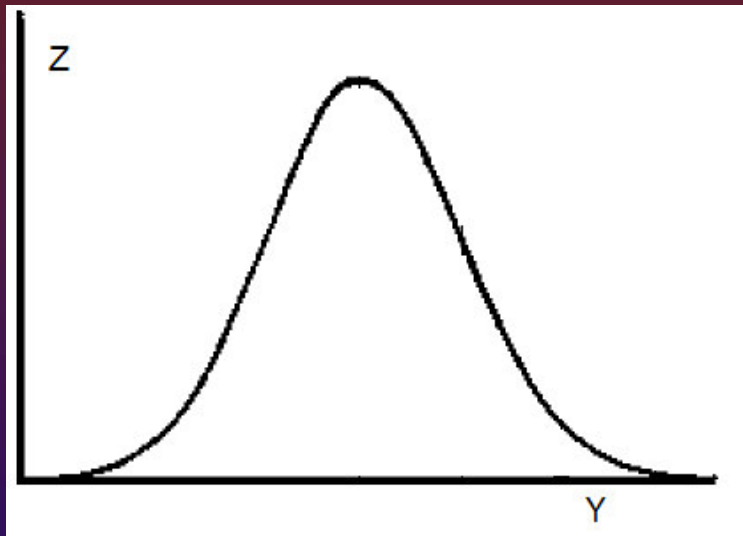


Du discret au continu

Différentes distributions



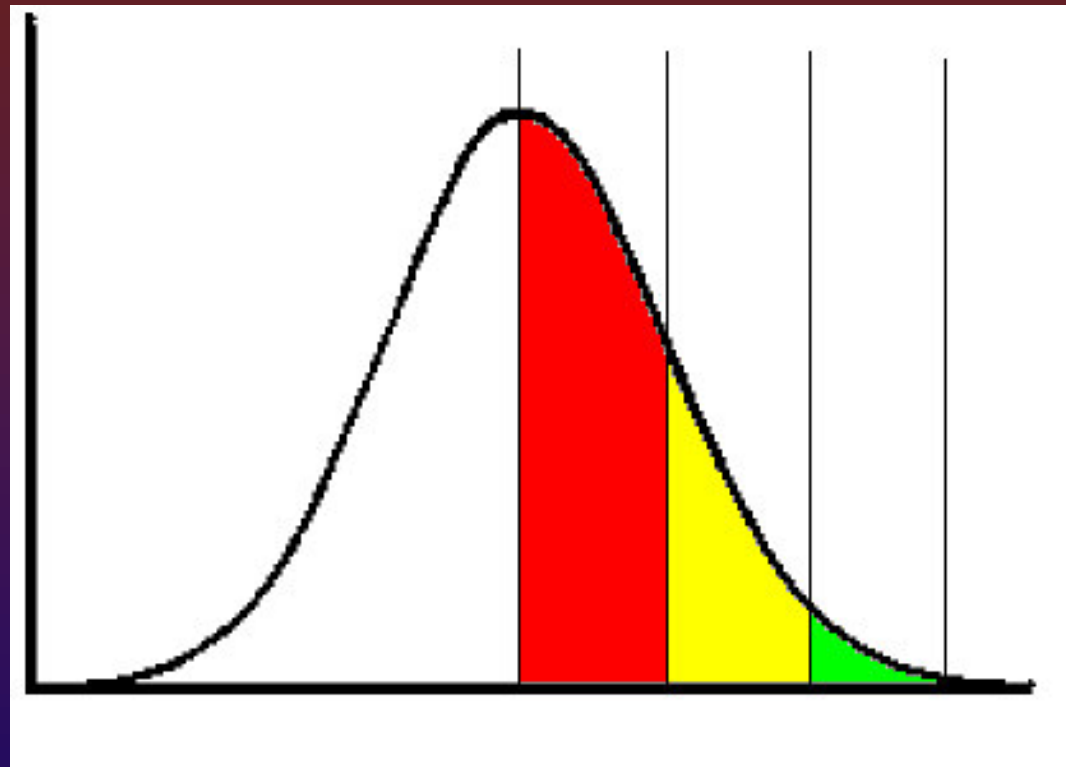
Apport des probabilités



$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left[\frac{(x-\mu)}{\sigma} \right]^2 \right]$$

La courbe est calculée par l'analyse sans données empiriques

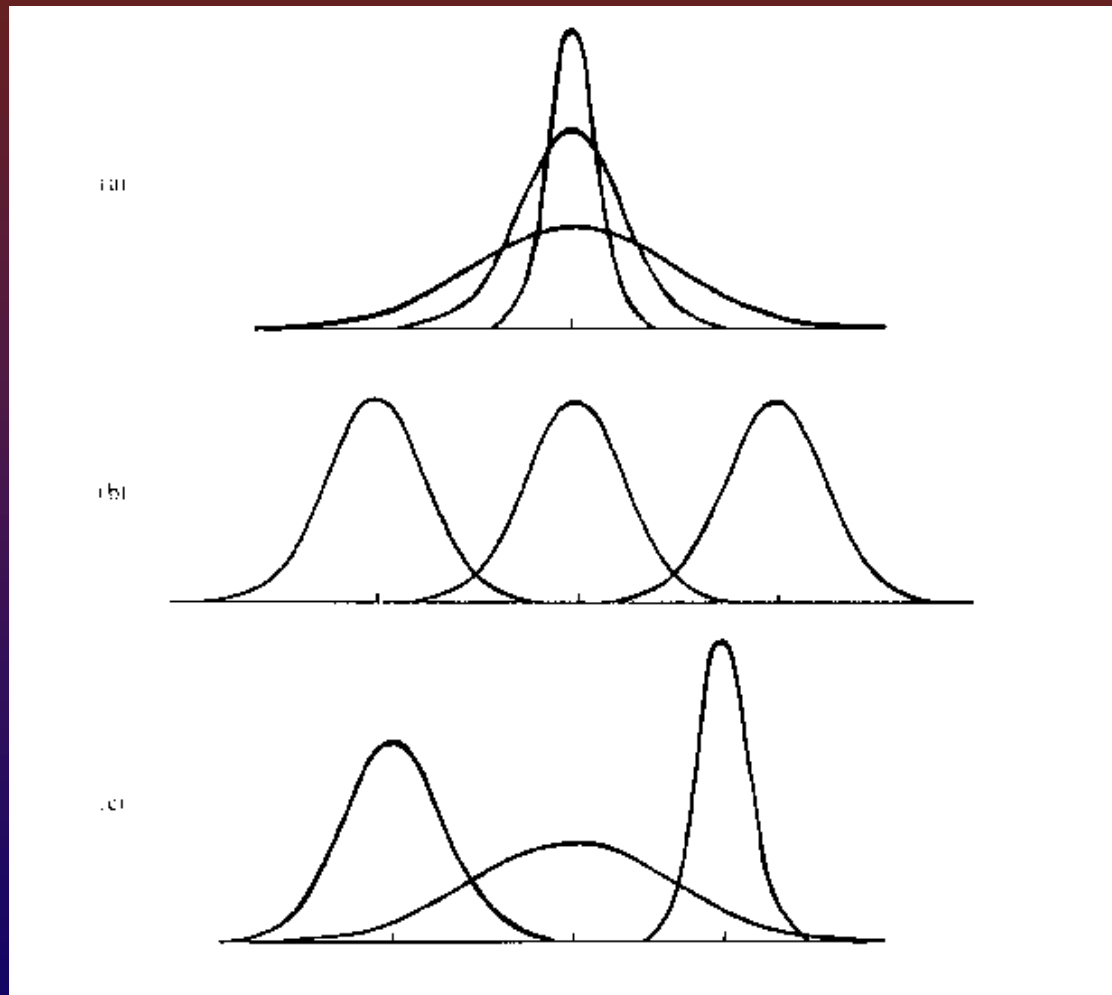
La distribution Normale



μ 1σ 2σ 3σ

50% 34.1% 13.6% 2.2% 0...%

Variations sur la loi Normale



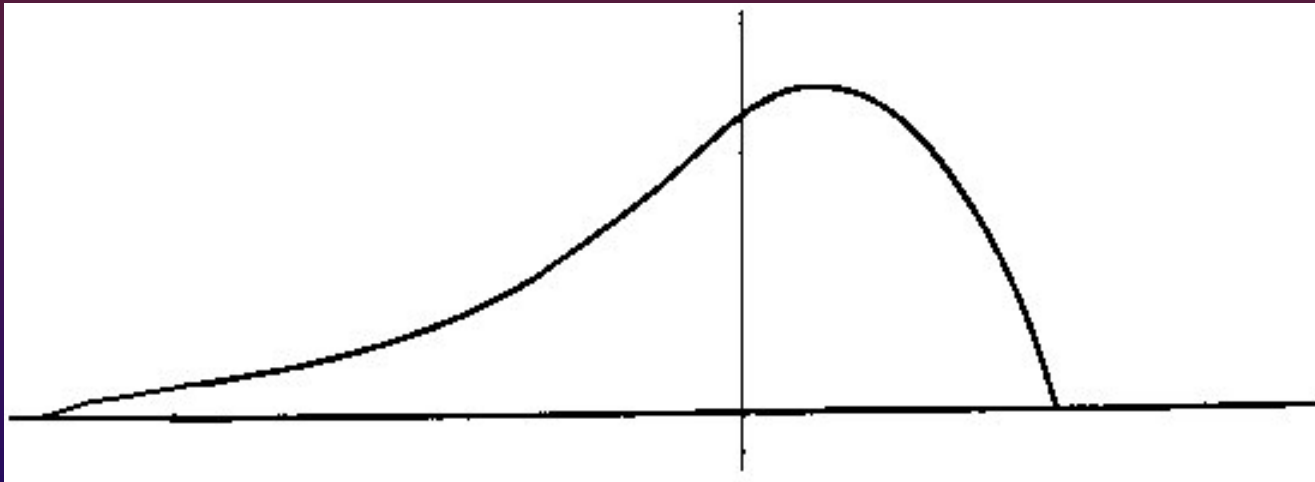
Le théorème central limite (1)

Des caractères biologiques ont une distribution normale si :

1. Les facteurs déterminants sont multiples
2. Ces facteurs sont indépendants
3. Les effets de ces facteurs sont additifs

Conséquence du théorème central limite

Exemple: Poids et alimentation



Le retour à la Normale

Variable x d'une population réelle

Moyenne μ

Ecart-Type σ



Centrée et Normée

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

Le Z de la loi normale théorique

Moyenne 0

Ecart-Type 1



Exemple sur la distribution Z

Soit une longueur mesurée sur une population :

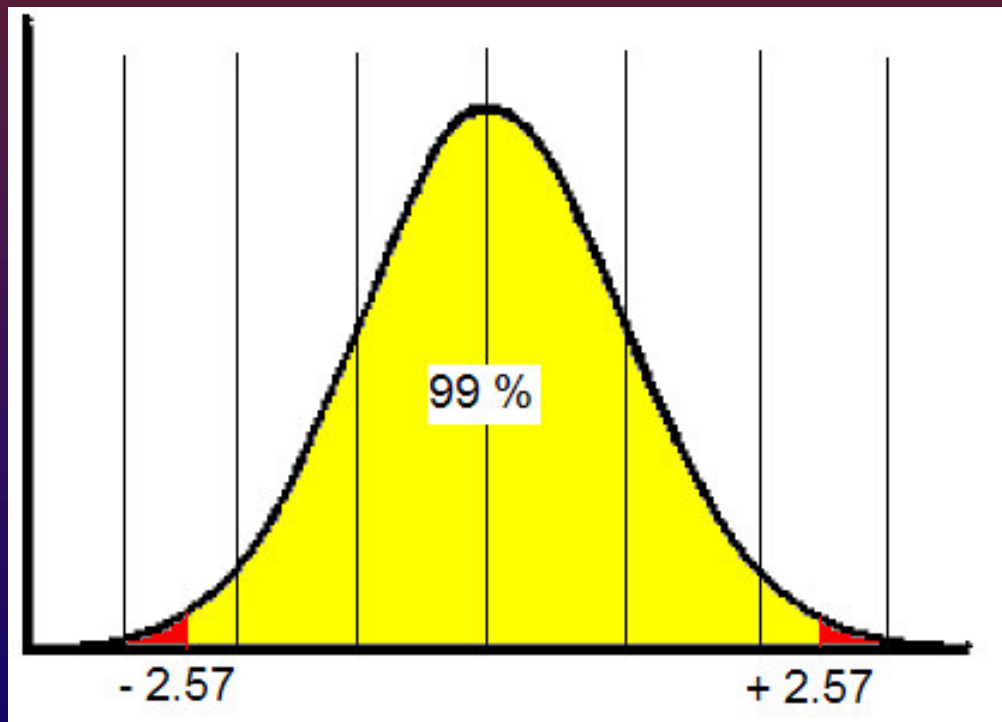
1. de moyenne 11.2 (cm ?)
2. d'écart-type 1.8
3. de distribution Normale

- 1. Entre quelle valeurs trouve t-on 99 % de la population*
- 2. Quel pourcentages de la population mesure plus de 14.8*

Intervalle de confiance

Loi Bilatérale :

Pour 99% le Z est de 2.57



Loi normale centrée réduite bilatérale

Valeur du Z	Probabilité	Complément
0	0.000	1.000
0.1	0.080	0.920
0.2	0.159	0.841
0.3	0.238	0.764
0.4	0.311	0.689
0.5	0.383	0.617
0.6	0.451	0.549
0.674	0.500	0.500
0.7	0.516	0.484
0.8	0.578	0.424
0.9	0.632	0.368
1	0.683	0.317
1.1	0.729	0.271
1.2	0.770	0.230
1.3	0.808	0.194
1.4	0.838	0.162
1.5	0.866	0.134
1.6	0.890	0.110
1.645	0.900	0.100
1.7	0.911	0.089
1.8	0.928	0.072
1.9	0.943	0.057
1.96	0.950	0.050
2	0.954	0.048
2.1	0.964	0.036
2.2	0.972	0.028
2.3	0.979	0.021
2.326	0.980	0.020
2.4	0.984	0.016
2.5	0.988	0.012
2.575	0.990	0.010
2.6	0.991	0.009
2.7	0.993	0.007
2.8	0.995	0.005
2.9	0.996	0.004
3	0.997	0.003
3.5	1.000	0.000

Calcul d'intervalle

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

1. *Entre quelle valeurs trouve t-on 99 % de la population*

$$\mu = 11.2$$

$$\sigma = 1.8$$

Au seuil de 99% on a $Z = 2.57$

$$Z \cdot \sigma = 4.62$$

L'intervalle de confiance à 99 % est de
 11.2 ± 4.62 ; donc il va de 6.58 à 15.82

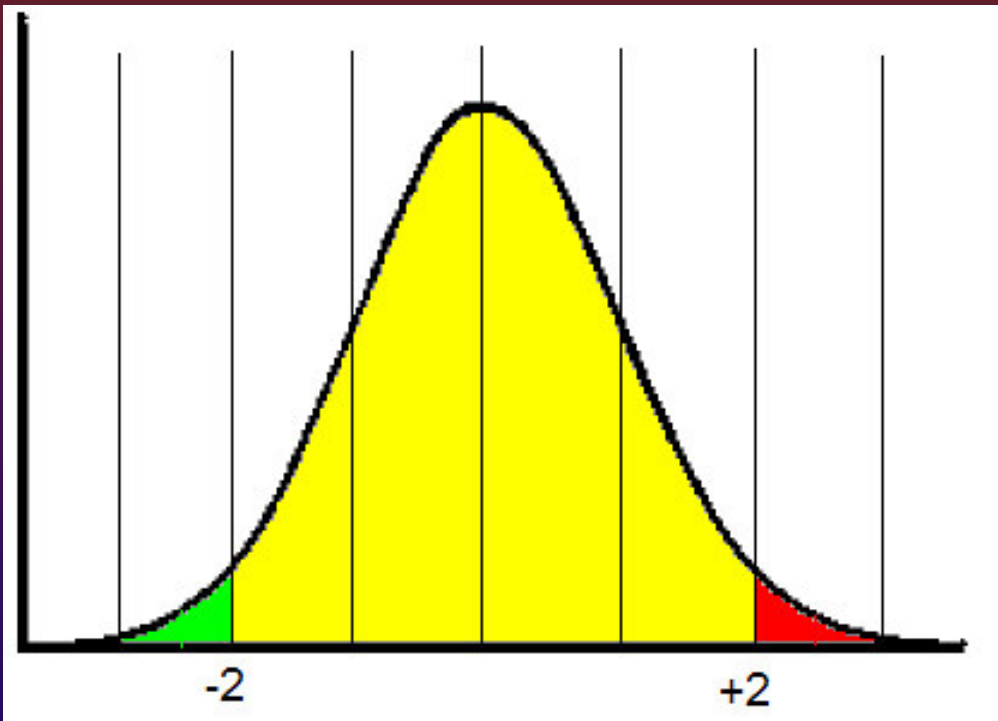
$$Z = \frac{x - \mu}{\sigma}$$

Pourcentages

*Quel pourcentages de la population
de moyenne 11.2
et d'écart type 1.8 mesure plus de 14.8 ?*

$$Z = \frac{14.8 - 11.2}{1.8} = 2$$

Bilatérale / Unilatérale



4.6 % divisé par 2 = 2.3%

Loi normale centrée réduite bilatérale

Valeur du Z	Probabilité	Complément
0	0.000	1.000
0.1	0.080	0.920
0.2	0.159	0.841
0.3	0.238	0.764
0.4	0.311	0.689
0.5	0.383	0.617
0.6	0.451	0.549
0.674	0.500	0.500
0.7	0.516	0.484
0.8	0.578	0.424
0.9	0.632	0.368
1	0.683	0.317
1.1	0.729	0.271
1.2	0.770	0.230
1.3	0.808	0.194
1.4	0.838	0.162
1.5	0.866	0.134
1.6	0.890	0.110
1.645	0.900	0.100
1.7	0.911	0.089
1.8	0.928	0.072
1.9	0.943	0.057
1.96	0.950	0.050
2	0.954	0.048
2.1	0.964	0.036
2.2	0.972	0.028
2.3	0.979	0.021
2.326	0.980	0.020
2.4	0.984	0.016
2.5	0.988	0.012
2.575	0.990	0.010
2.6	0.991	0.009
2.7	0.993	0.007
2.8	0.995	0.005
2.9	0.996	0.004
3	0.997	0.003
3.5	1.000	0.000

Statistiques ou Divination ?

Incertitude initiale	oui	oui
Collecte de données	oui	oui
Analyse des données	oui	oui
Répétabilité des analyses	oui	non
Interprétation des résultats	oui	oui