

Chapitre 4

Statistiques descriptives des variables quantitatives

1. Paramètres de position

José LABARERE, PU-PH

Arnaud Seigneurin, MCU-PH, Bastien Boussat, MCU-PH, Alexandre Bellier, AHU, Patrice François, PU-PH

- Objectifs
- Introduction
- Paramètres de position

Objectifs

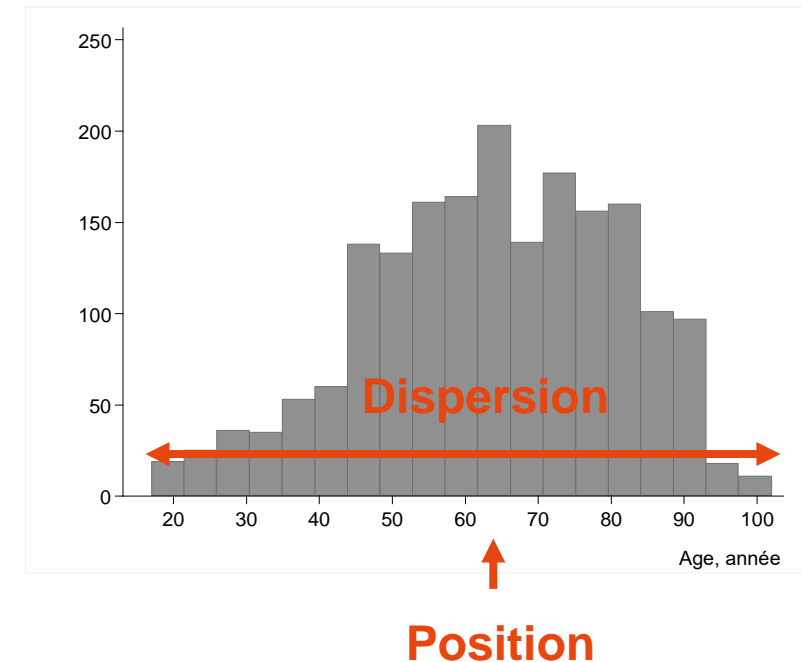
- Expliquer la notion de paramètres de position de la distribution d'une variable quantitative
- Citer, définir, sélectionner, déterminer et interpréter les paramètres de position d'une variable quantitative

Plan

- Objectifs
- **Introduction**
- Paramètres de position

Introduction

- Décrire de manière synthétique la distribution des valeurs d'une variable quantitative sur une série d'individus
- Paramètre = valeur numérique qui résume la distribution d'une variable
 - Paramètres de position (1^{er} ordre)
 - Paramètres de dispersion (2nd ordre)
 - Paramètre mixte



Plan

- Objectifs
- Introduction
- Paramètres de position

Paramètres de position

- Mode
- Moyenne (arithmétique)
- Moyenne (géométrique)
- Médiane
- Quantiles

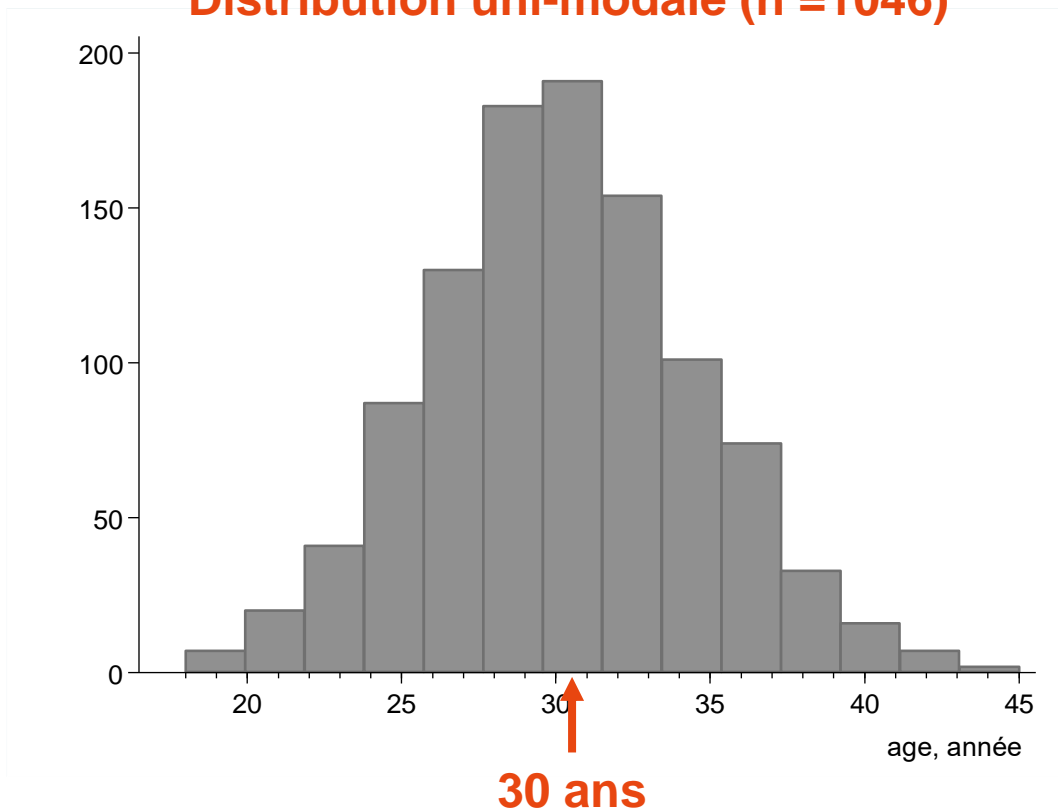


Paramètres de tendance
centrale

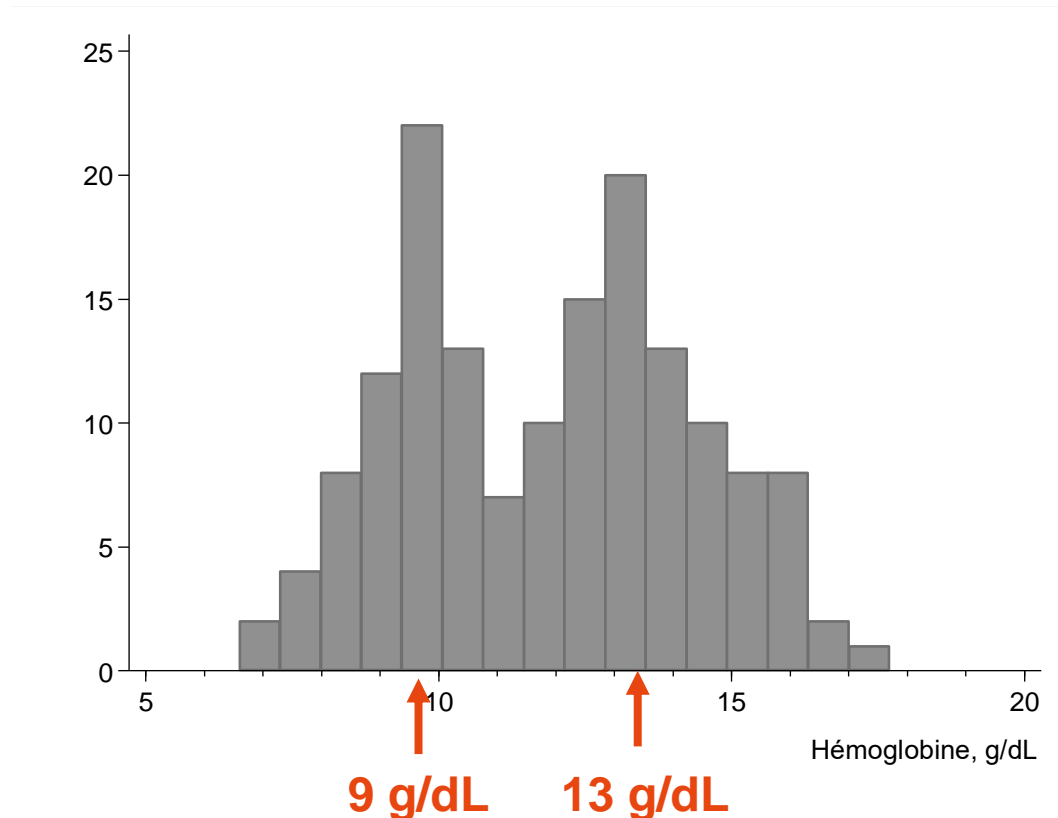
Mode

- Valeur la plus fréquente d'une distribution (pic de distribution)

Distribution uni-modale (n = 1046)



Distribution bi-modale (n = 155)



Moyenne (arithmétique)

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

x_i = valeur de la variable X observée sur l'individu i

\bar{x} = moyenne de la variable X observée sur l'échantillon

n = effectif d'individus dans l'échantillon

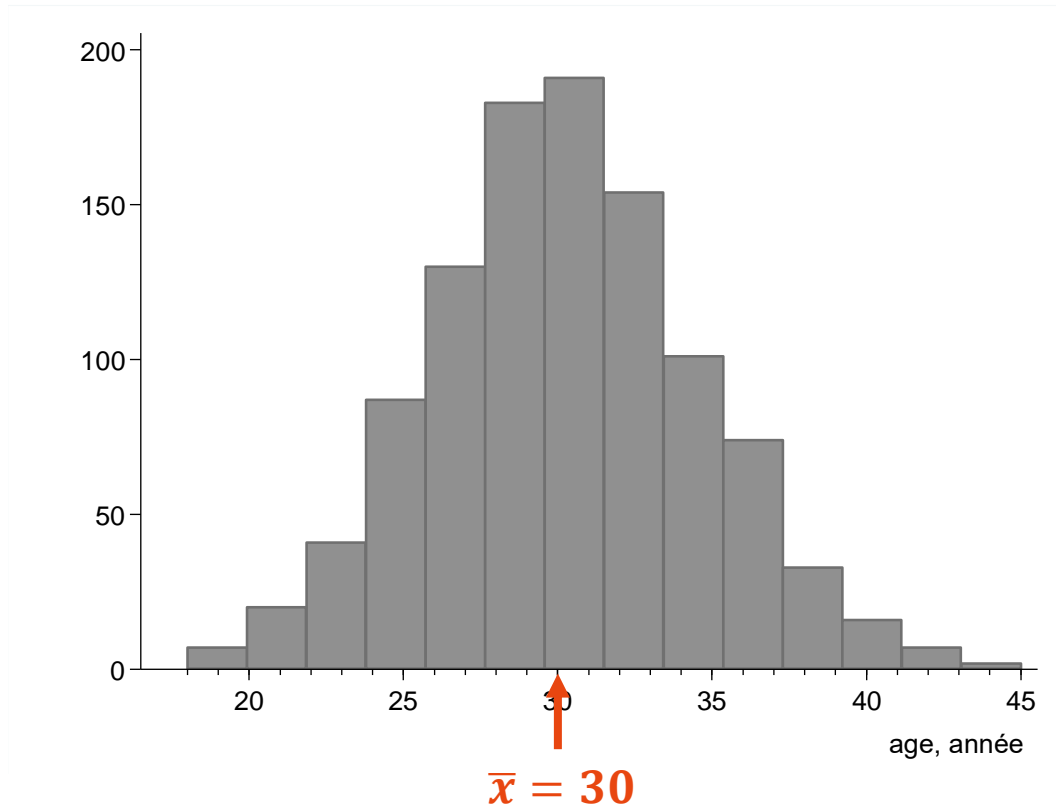
Propriétés algébriques de la moyenne

$$y = ax + b$$

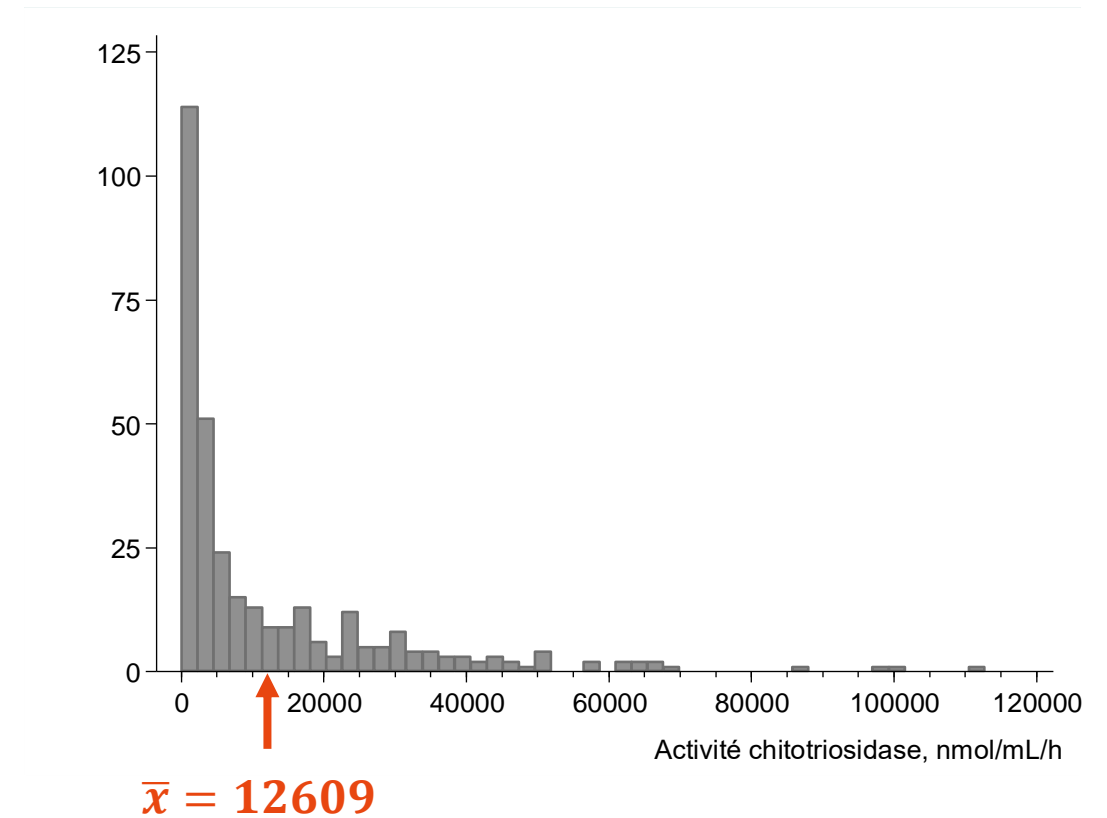
$$\bar{y} = a\bar{x} + b$$

Moyenne (arithmétique)

Distribution symétrique (n = 1046)



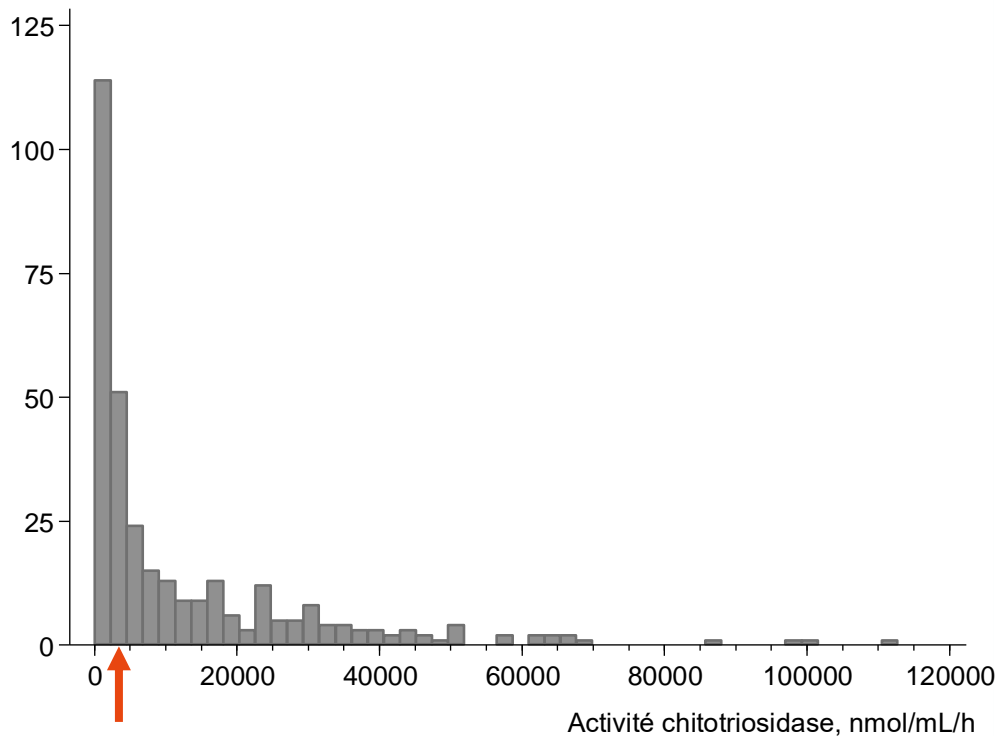
Distribution asymétrique (n = 326)



Moyenne géométrique

$$\overline{x_g} = e^{\frac{\sum_i^n \ln(x_i)}{n}}$$

Distribution asymétrique (n = 326)



$$\overline{x_g} = 4042$$

x_i = valeur de la variable X observée sur l'individu i

$\ln(x)$ = logarithme népérien

\bar{x} = moyenne de la variable X observée sur l'échantillon

n = effectif d'individus dans l'échantillon

Médiane (50^{ème} percentile)

Paramètre de position basé sur les rangs

- Si n est **impair**

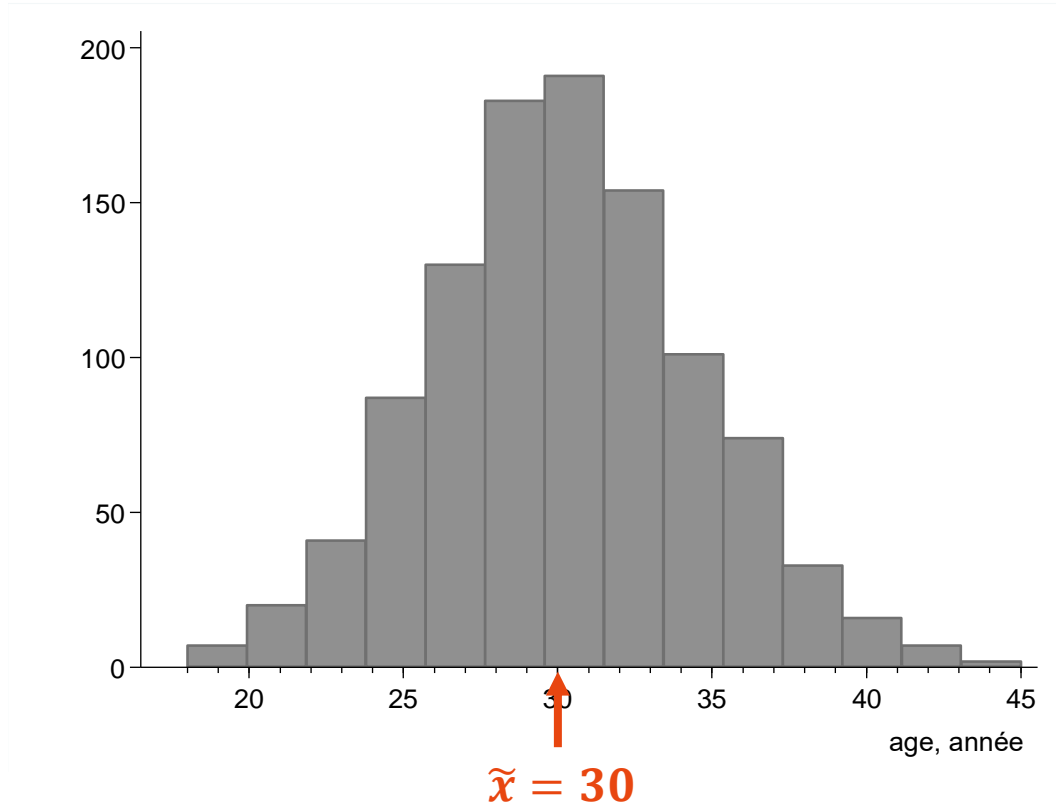
$$\tilde{x} = \frac{x_{(n+1)}}{2}$$

- Si n est **pair**

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}$$

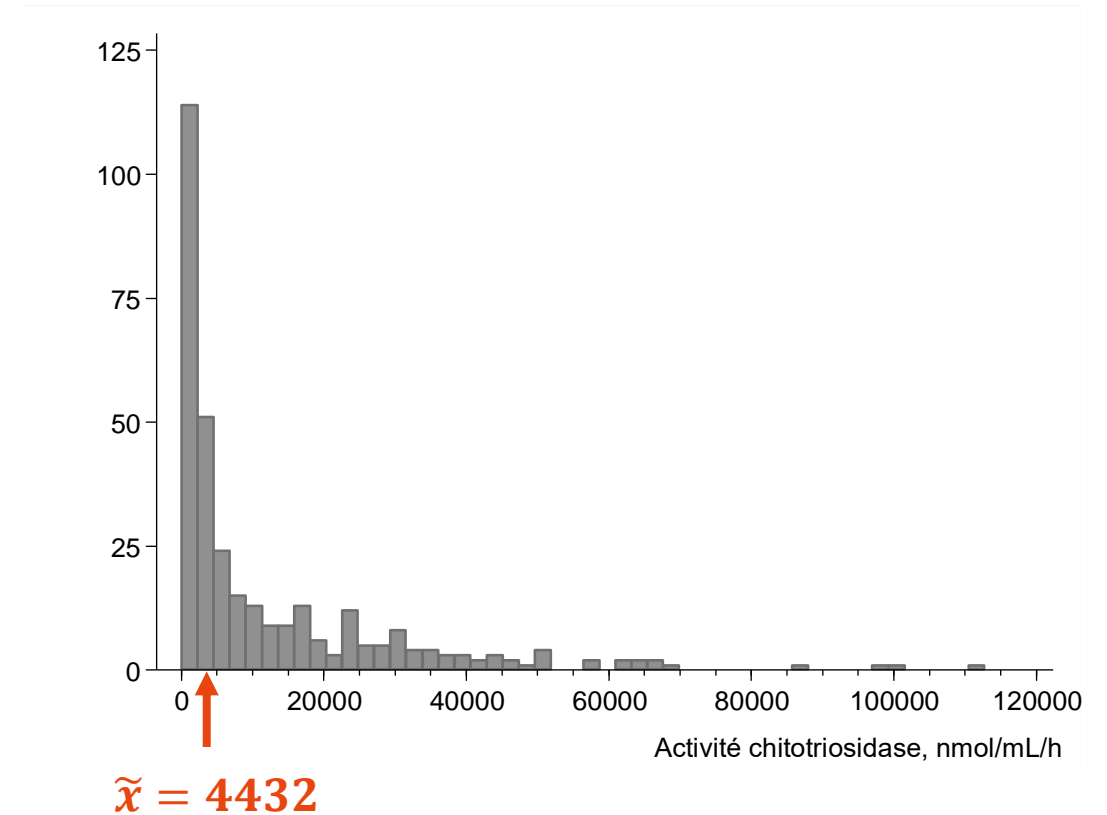
Médiane (50^{ème} percentile)

Distribution symétrique (n = 1046)



Adapté de Raskovalova T, et al. Breastfeeding Assessment Score: Systematic Review and Meta-analysis. Pediatrics 2014.

Distribution asymétrique (n = 326)



Adapté de Raskovalova T, et al. Accuracy of chitotriosidase activity and CCL18 concentration in assessing type I Gaucher disease severity. A systematic review with meta-analysis of individual participant data. Haematologica 2020.

Quantiles : (per)centiles, déciles, quartiles

- Paramètres de position basés sur les rangs
- Q_q = valeur de la variable au dessous de laquelle se situent $q\%$ des observations
 - Q_{10} = 10^{ème} percentile = 1^{er} décile
 - Q_{25} = 25^{ème} percentile = 1^{er} quartile
 - Q_{50} = 50^{ème} percentile = 2^{ème} quartile = médiane
 - Q_{75} = 75^{ème} percentile = 3^{ème} quartile
 - Q_{90} = 90^{ème} percentile = 9^{ème} décile

Quantiles : (per)centiles, déciles, quartiles

- Ordonner les observations par valeur croissante de la variable

$$\{x_r, r = 1, 2, \dots, i, \dots, n\}$$

- Si $\frac{q}{100}(n + 1)$ est un entier i

$$Q_q = x_i$$

- Si $r < \frac{q}{100}(n + 1) < (r + 1)$

$$Q_q = \frac{(x_r + x_{(r+1)})}{2}$$

Messages clés

Paramètre

Avantages

Inconvénients

Moyenne

- Intuitive
- Se prête aux calculs algébriques et aux tests statistiques

- Sensible aux valeurs extrêmes
- Inappropriée pour les variables de distribution asymétrique

Moyenne géométrique

- Robuste aux valeurs extrêmes
- Pertinente pour les variables dont la distribution est étalée vers la droite

- Traitement des valeurs nulles

Médiane

- Robuste aux valeurs extrêmes
- Pertinente pour les variables de distribution asymétrique
- Intéressante pour les variables censurées (survie)

- Se prête mal aux calculs algébriques et aux tests statistiques

Glossaire

- Chitotriosidase : biomarqueur plasmatique dont l'activité enzymatique est augmentée dans la maladie de Gaucher
- Maladie de Gaucher (OMIM#230800) : Maladie de surcharge lysosomale, liée à un déficit enzymatique en glucocérébrosidase et transmise sur un mode autosomique récessif.

Mentions légales

L'ensemble de ce document relève des législations française et internationale sur le droit d'auteur et la propriété intellectuelle. Tous les droits de reproduction de tout ou partie sont réservés pour les textes ainsi que pour l'ensemble des documents iconographiques, photographiques, vidéos et sonores.

Ce document est interdit à la vente ou à la location. Sa diffusion, duplication, mise à disposition du public (sous quelque forme ou support que ce soit), mise en réseau, partielles ou totales, sont strictement réservées à l'Université Grenoble Alpes (UGA).

L'utilisation de ce document est strictement réservée à l'usage privé des étudiants inscrits à l'Université Grenoble Alpes (UGA), et non destinée à une utilisation collective, gratuite ou payante.