

Chapitre 1 : Statistiques descriptives

Romann Szczepaniak

January 11, 2026

1 Introduction

Données ou observations $x_1, \dots, x_n$ où n est la taille de la population

Voici les probabilités :D

Le statisticien cherche à avoir des informations sur la loi des données

On distingue deux statistiques :

- i. Les statistiques descriptives : décrire les données
- ii. statistiques inférentielle : reconstruire la loi

2 Médiane, quantiles, moyenne et variance empirique

Empirique = fonction des données $x_1, \dots, x_n$

2.1 Médiane et quantiles

Définition 1. m est la moyenne de la série statistique $x_1, \dots, x_n$ si $\text{card} \{i : x_i \geq m\} \geq \frac{n}{2}$
et $\text{card} \{i : x_i \leq m\} \geq \frac{n}{2}$

On utilise une statistique d'ordre : $x_{(1)}, \dots, x_{(n)}$ qui sont des données classées par ordre croissant

Convention : Si n est pair, on choisit comme médiane la moyenne des deux valeurs du milieu

Remarque : $e(t) = \frac{1}{n} \sum_{i=1}^n |t - x_i|$ mesure l'écart de t à la série $x_1, \dots, x_n$. La médiane m minimise la fonction $e(t)$

Généralisation

Définition 2. Soit $p \in]0, 1[$, le quantile empirique d'ordre p est $q_p =$

$$\begin{cases} x_{[np]+1} \text{ si } np \notin \mathbb{N} \\ q_p = \frac{x_{(np)} + x_{(np+1)}}{2} \text{ si } np \in \mathbb{N} \end{cases}.$$

$p = \frac{1}{2}$: Médiane $p = \frac{1}{3}$ 1er quartile, ...

2.2 Moyenne et variance empirique

$$v(t) = \frac{1}{n} \sum_{i=1}^n (t - x_i)^2$$

$$v'(t) = \frac{1}{n} \sum_{i=1}^n 2(t - x_i), \text{ donc } v'(t) = 0 \iff t - \frac{1}{n} \sum_{i=1}^n x_i = 0$$

v décroissante puis croissante avec un minimum en $t = \frac{1}{n} \sum_{i=1}^n x_i$

Notation : $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$, c'est la moyenne empirique des $(x_i)_{1 \leq i \leq n}$

Remarque :

$$v(t) = \frac{1}{n} \sum_{i=1}^n (t - x_i)^2 = \frac{1}{n} \sum_{i=1}^n (t - \bar{x}_n)^2 + 2(t - \bar{x}_n)(\bar{x}_n - x_i) + (\bar{x}_n - x_i)^2$$

$$= (t - \bar{x}_n)^2 + \frac{2}{n} (t - \bar{x}_n) \sum_{i=1}^n (\bar{x}_n - x_i) + \frac{1}{n} \sum_{i=1}^n (\bar{x}_n - x_i)^2$$

$v(t) = v_n + (t - \bar{x}_n)^2 \implies$ On retrouve que $\bar{x}_n$ minimise $v(t)$ en $t \in \mathbb{R}$

Comparaison médiane et moyenne :

La médiane est peu sensible aux grandes valeurs par rapport à la moyenne

Variance empirique

$$v_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \text{ mesure la dispersion à la moyenne}$$

Ecart type empirique : $\sigma_n = \sqrt{v_n}$ même unité que les $(x_i)_{1 \leq i \leq n}$

Encadrement

- $x_{(1)} \leq \bar{x}_n \leq x_{(n)}$
- $v_n \leq (x_{(n)} - x_{(1)})^2$
- $\sigma_n \leq x_{(n)} - x_{(1)}$

Car $\forall i \in \{1, \dots, n\} \mid x_i - \bar{x}_n \leq x_{(n)} - x_{(1)}$

$x_{(n)} - x_{(1)}$ est appelée l'étendue de la série

Calcul de v_n .

En développant le carré dans v_n , on obtient :

$$v_n = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2$$

3 Représentation graphique

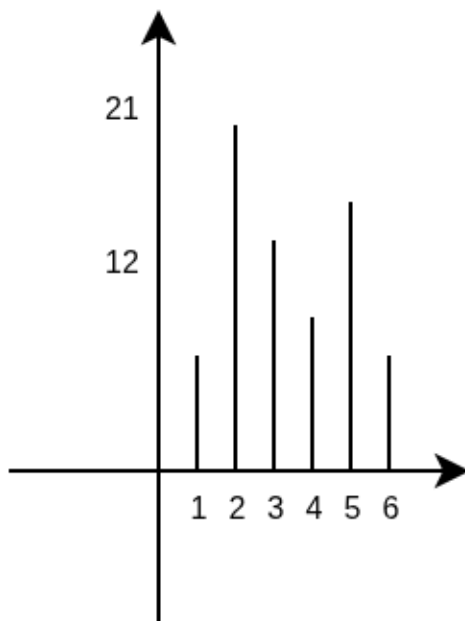
Exemple 1. *Exemple 1 : 100 lancers de dé*

k	1	2	3	4	5	6
n_k	12	21	14	18	13	22

$$n_k = \text{card} \{i : x_i = k\}$$

On peut trouver la médiane et la moyenne avec les formules données ci dessus (facile)

Diagramme en bâton :



mk

$a_0, a_1, \dots, a_n \in \mathbb{A}$