Université Paul Valéry

Arts - Lettres - Langues
Sciences humaines & sociales

Représentations statistiques

Administration **É**conomique et **S**ociale

Mathématiques
XA100M

1.1. Introduction.

Dans de nombreux domaines, on dispose d'une importante quantité de données que l'on souhaite étudier.

La statistique descriptive fournit des outils pour **décrire** toutes ces données.

Exemples :

☞ En démographie : nombre de naissances, de décès, d'immigrants, d'émigrants dans une ville...

☞ En météorologie : température, pression athmosphérique en de nombreux points de la planète.

Données statistiques : ensemble de mesures ou observations décrivant l'**état** ou l'**évolution** d'un phénomène.

La **statistique descriptive** est l'ensemble des outils mathématiques qui permettent de **décrire** ou **résumer** les données statistiques.

☢ L'objet de la statistique descriptive n'est PAS de prédire l'évolution future d'un phénomène à partir de données actuelles.

Population : ensemble sur lequel on effectue les observations.

Échantillon : partie de la population sur laquelle on a effectivement fait les mesures (trop souvent inconnu).

Individu : élément de la population.

Malgré la terminologie, les individus ne sont pas nécessairement des personnes et la population n'est pas nécessairement un ensemble d'êtres humains.

Variable ou **caractère** : ce qui est observé sur la population

Modalités : valeurs observables (valeurs pouvant être prises par la variable).

Exemple 1

On étudie la récolte de blé de chaque pays.

– Population : ensemble des pays ;
– Individus : chaque pays ;
– Variable : quantité de blé récoltée dans chaque pays en tonnes ;
– Modalités : les nombres positifs.

Exemple 2

On veut quantifier la famine.

– Population : tous les êtres humains ;
– Individus : chaque être humain ;
– Variable : quantité de nourriture à laquelle chacun a accès en kilogrammes ;
– Modalités : les nombres positifs.

1.2. Nature des variables.

1.2.1. *Trois exemples.*

1$^{\text{re}}$ étude

– Population : les exploitations agricoles de l'Hérault ;
– Individus : chaque exploitation agricole de l'Hérault ;
– Variable : surface cultivée en hectares ;
– Modalités : les nombres positifs.

Les modalités sont des **nombres** sur lesquels faire des calculs a un sens.

2^{e} **étude**

– Population : les exploitations agricoles de l'Hérault ;
– Individus : chaque exploitation agricole de l'Hérault ;
– Variable : culture dominante ;
– Modalités : blé, orge, vigne.

Les modalités sont des **mots**.

3$^{\text{e}}$ **étude** : on dit qu'une exploitation est de **type 1** si elle peut accueillir des touristes en chambres d'hôtes, de **type 2** sinon.

– Population : les exploitations agricoles de l'Hérault ;
– Individus : chaque exploitation agricole de l'Hérault ;
– Variable : type ;
– Modalités : 1 ou 2.

Les modalités sont des **nombres** sur lesquels faire des calculs n'a aucun sens (qu'est-ce qu'une exploitation de type 1+2 ?)

1.2.2. *Variables quantitatives.*

Variable quantitative : variable dont les modalités sont des nombres sur lesquels faire des calculs a un sens.

La variable « surface de terre cultivée » de la 1re étude (§1.2.1) est quantitative.

La variable « type » de la 3e étude (§1.2.1) n’est pas quantitative.

1.2.3. *Variables qualitatives.*

Variable qualitative : variable dont les modalités sont des mots ou des nombres sur lesquels faire des calculs n'a pas de sens (et qui pourraient être remplacés par des mots).

La variable « culture dominante » de la 2^{e} étude (§1.2.1) est qualitative.

La variable « type » de la 3^{e} étude (§1.2.1) est qualitative. Au lieu de dire de « type 1 » et de « type 2 », on pourrait dire de « type touristique » et de « type non touristique ».

1.2.4. *Continuité d'une variable quantitative.*

Lorsqu'une variable est quantitative, elle est soit **continue** soit **discrète**.

Variable continue : c'est une variable dont les modalités sont des intervalles.

Variable discrète : c'est une variable qui n'est pas continue.

Variable continue : exemple

La variable « âge » est continue. On peut imaginer donner l'âge avec une précision aussi grande que voulue.

Dire de quelqu'un qu'il a 19 ans, c'est dire qu'il est né depuis plus de 19 ans et moins de 20 ans. La modalité correspondante est l'intervalle $[19, 20[$.

Dire qu'une personne est dans la classe des $15 - 20$ ans, c'est dire qu'elle est née depuis plus de 15 ans et moins de 21 ans. La modalité correspondante est l'intervalle $[15, 21[$.

Variable discrète : exemple

La variable « nombre d'enfants d'une famille » est discrète.

Cela n'a aucun sens de parler de $3,5$ enfants.

1.2.5. *Variables ordinales.*

Une variable qualitative ou quantitative est **ordinale** si on peut **classer** les modalités dans un certain ordre.

Une variable qui n'est pas ordinale est **nominale**.

Exemple de variable qualitative ordinale

– Population : les enseignant-chercheurs de l'Université Paul-Valéry ;
– Individus : chaque enseignant-chercheur ;
– Variable : grade ;
– Modalités : maître de conférences MCF , professeur PR.

On peut classer les modalités par ordre hiérarchique :

$$PR > MCF.$$

1.3. Description d'une situation statistique.

Avant toute étude statistique, on précise

- ☞ La population
- ☞ L'échantillon (lorsqu'il est connu) ;
- ☞ La variable observée
  - ☞ Les modalités ;
  - ☞ Quantitative ☞ Discrète ou continue ; ou qualitative ;
  - ☞ Ordinale ou nominale ;

2.1. Représentation en tableau.

On étudie un caractère de modalités $m_1, m_2, \ldots, m_p$.
À la modalité m_1 est associée l'effectif n_1,
à la modalité m_2 est associée l'effectif n_2 *etc.*

Modalités	Effectifs
m_1	n_1
m_2	n_2
$\vdots$	$\vdots$
m_p	n_p

Modalités	m_1	m_2	$\cdots$	m_p
Effectifs	n_1	n_2	$\cdots$	n_p

Si n est l'effectif total de l'échantillon,

$$\begin{aligned} n &= \sum_{i=1}^{p} n_i \\ &= n_1 + n_2 + \cdots + n_p. \end{aligned}$$

Lorsque la variable est ordinale, on classe toujours les modalités de la plus petite à la plus grande :

$$m_1 < m_2 < \cdots < m_p.$$

Modalités	Effectifs
m_1	n_1
m_2	n_2
$\vdots$	$\vdots$
m_p	n_p

Modalités	m_1	m_2	$\cdots$	m_p
Effectifs	n_1	n_2	$\cdots$	n_p

☢ C'est suivant les **modalités** qu'on ordonne, PAS selon les effectifs.

2.2. Représentations graphiques.

2.2.1. *Variable qualitative.*

Dans ce cas, on utilise souvent un **diagramme en secteur**.

Dans ce diagramme, chaque modalité est représentée par un secteur circulaire d'un disque dont la **surface** est proportionnelle à l'effectif.

Modalités	m_1	m_2	$\cdots$	m_p
Effectifs	n_1	n_2	$\cdots$	n_p
Surfaces	S_1	S_2	$\cdots$	S_p

Il existe un nombre $c > 0$ tel que

$$S_1 = cn_1, \quad S_2 = cn_2, \cdots, S_p = cn_p.$$

C'est le même nombre c pour S_1, S_2, …, S_p.

P

Q

La surface totale du disque est

$$S_1 + S_2 + \cdots + S_p = c(n_1 + n_2 + \cdots + n_p) = cn.$$

Si R est le rayon du disque, sa surface est πR^2. On a donc $cn = \pi R^2$ puis

$$c = \frac{\pi R^2}{n}.$$

Pour la modalité m_i, on a donc

$$S_i = cn_i = \pi R^2 \frac{n_i}{n}.$$

Si α_i est l'angle associé au secteur S_i, on a

$$S_i = \pi R^2 \frac{\alpha_i}{2\pi}$$

donc

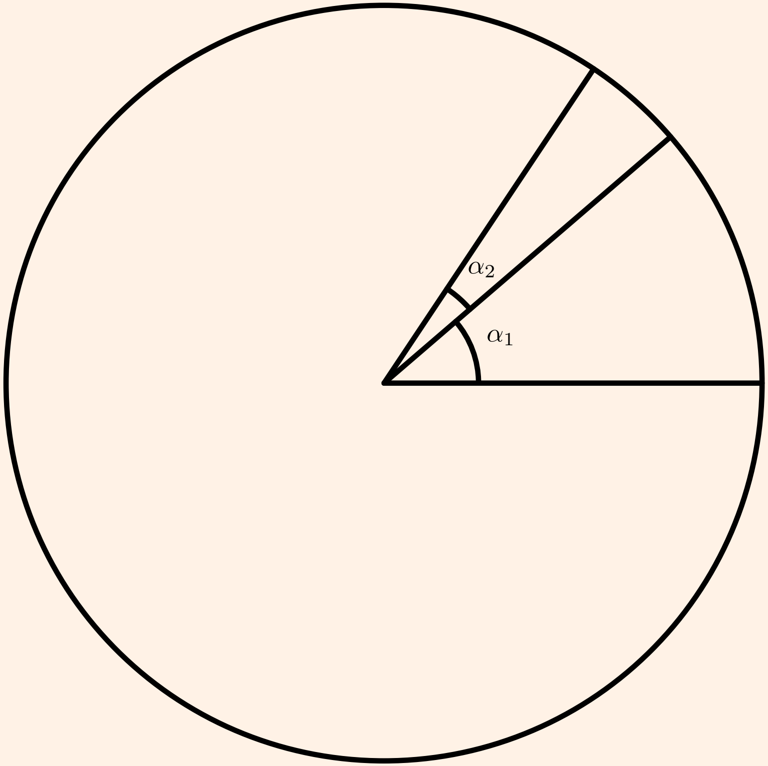
$$\pi R^2 \frac{n_i}{n} = \pi R^2 \frac{\alpha_i}{2\pi}$$

et

$$\alpha_i = 2\pi \frac{n_i}{n} \text{ radians} = 360 \frac{n_i}{n} \text{ degrés.}$$

☞ Dans un diagramme en secteurs, une modalité d'effectif n_i est représentée par un secteur d'angle

$$360\frac{n_i}{n} \text{ degrés.}$$



Exemple

Situation matrimoniale d'un échantillon de 1000 personnes.

Modalités	Mariée	Veuve	Célibataire	Divorcée
Effectifs	400	100	200	300

L'effectif total est

$$n = 400 + 100 + 200 + 300 = 1000.$$

La modalité « mariée » est représentée par un angle (en degrés) de

$$\alpha_{\text{mariée}} = 360 \frac{n_{\text{mariée}}}{n} = 360 \frac{400}{1000} = 144.$$

La modalité « veuve » est représentée par un angle (en degrés) de

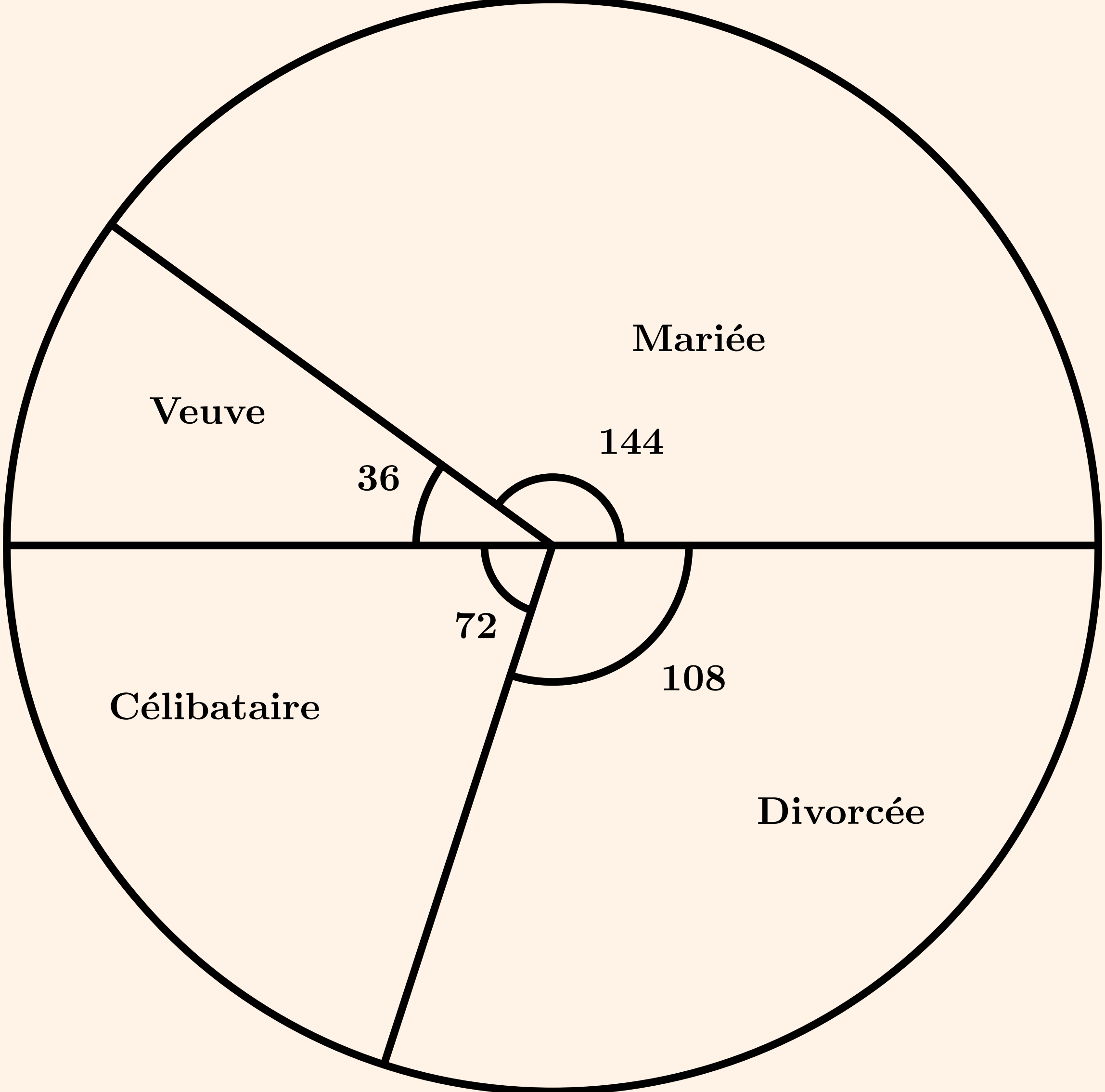
$$\alpha_{\text{veuve}} = 360\frac{n_{\text{veuve}}}{n} = 360\frac{100}{1000} = 36.$$

La modalité « célibataire » est représentée par un angle (en degrés) de

$$\alpha_{\text{célibataire}} = 360\frac{n_{\text{célibataire}}}{n} = 360\frac{200}{1000} = 72.$$

La modalité « divorcée » est représentée par un angle (en degrés) de

$$\alpha_{\text{divorcée}} = 360\frac{n_{\text{divorcée}}}{n} = 360\frac{300}{1000} = 108.$$


Mariée
Veuve
144
36
72
108
Célibataire
Divorcée

On peut aussi utiliser un **diagramme en bâtons**.

Dans ce diagramme, chaque modalité est représentée par un trait de **hauteur** proportionnelle à l'effectif.

Modalités	m_1	m_2	$\cdots$	m_p
Effectifs	n_1	n_2	$\cdots$	n_p
Hauteurs	h_1	h_2	$\cdots$	h_p

Il existe un nombre $K > 0$ tel que

$$h_1 = Kn_1, \quad h_2 = Kn_2, \cdots, h_p = Kn_p.$$

C'est le même nombre K pour h_1, h_2, ..., h_p.

Exemple

Situation matrimoniale d'un échantillon de 1000 personnes.

Modalités	Mariée	Veuve	Célibataire	Divorcée
Effectifs	400	100	200	300

On choisit

$$K = \frac{2}{100} \text{ cm.}$$

La modalité « mariée » est représentée par un trait de hauteur (en cm) de

$$h_{\text{mariée}} = \frac{2}{100} n_{\text{mariée}} = \frac{2}{100} \times 400 = 8.$$

La modalité « veuve » est représentée par un trait de hauteur (en cm) de

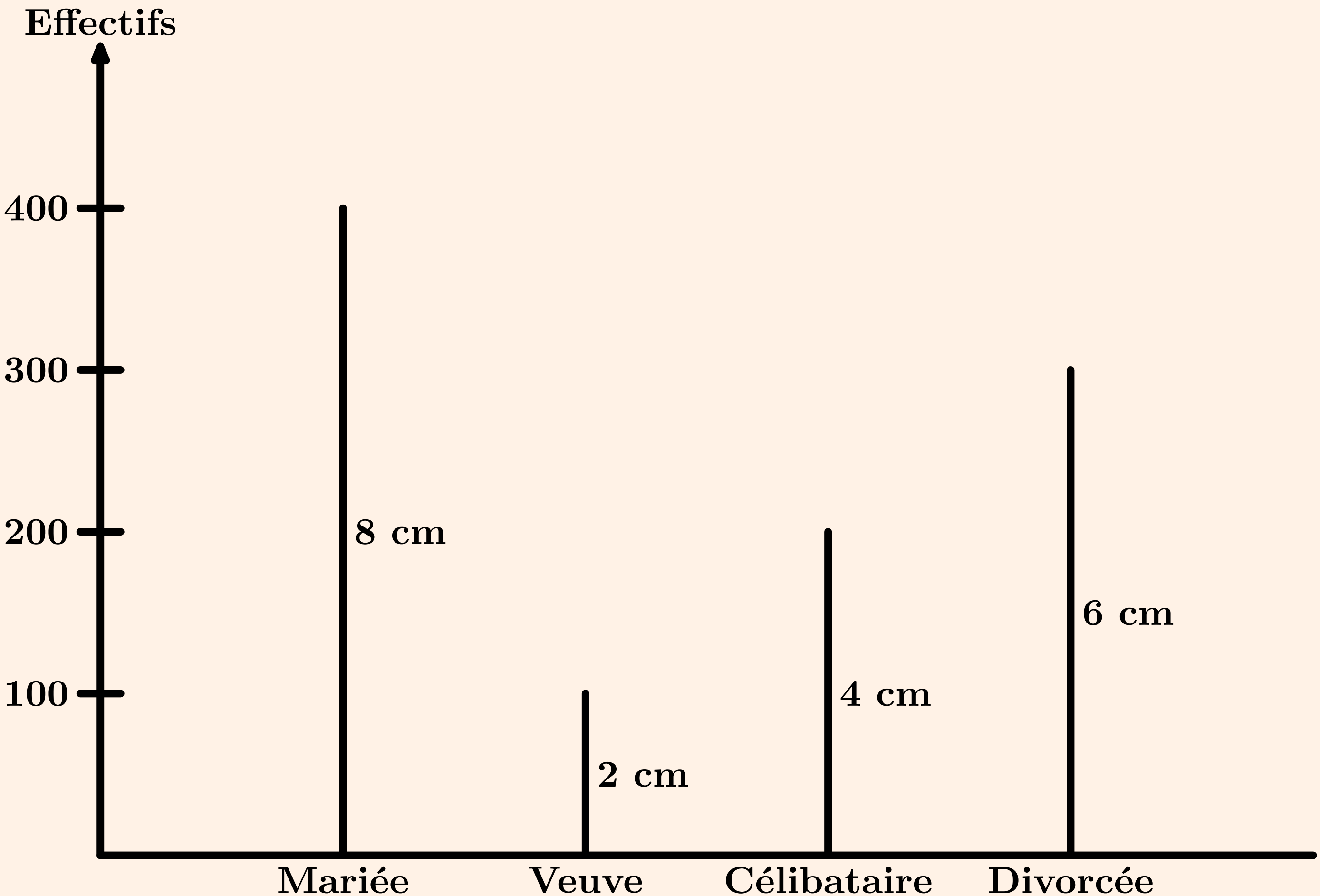
$$h_{\text{veuve}} = \frac{2}{100} n_{\text{veuve}} = \frac{2}{100} \times 100 = 2.$$

La modalité « célibataire » est représentée par un trait de hauteur (en cm) de

$$h_{\text{célibataire}} = \frac{2}{100} n_{\text{célibataire}} = \frac{2}{100} \times 200 = 4.$$

La modalité « divorcée » est représentée par un trait de hauteur (en cm) de

$$h_{\text{divorcée}} = \frac{2}{100} n_{\text{divorcée}} = \frac{2}{100} \times 300 = 6.$$


Effectifs
400
300
200
100
8 cm
2 cm
4 cm
6 cm
Mariée
Veuve
Célibataire
Divorcée

2.2.2. *Variable quantitative.*

Les modalités sont des nombres, les représentations doivent en tenir compte.

Lorsque la variable discrète on utilise un **diagramme en bâtons**.

La hauteur des bâtons est calculée comme au §2.2.1

La **distance** des bâtons à l'origine est proportionnelle à la modalité qu'ils représentent : si les modalités sont

$$m_1 < m_2 < \ldots < m_p$$

il existe $C > 0$ tel que les distances des bâtons à l'origine sont

$$\ell_1 = Cm_1 \quad \ell_2 = Cm_2 \quad \ldots \quad \ell_p = Cm_p.$$

STOP C'est le même nombre C pour $\ell_1, \ell_2, \ldots, \ell_p$.

Exemple

Nombre d'enfants par famille dans un échantillon de 17 familles.

Modalités	0	1	2	3
Effectifs	6	4	5	2

Pour la distance des bâtons à l'origine on choisit

$$C = 2 \text{ cm}.$$

Le bâton associé à la modalité « 0 enfant » est à distance (en cm)

$$\ell_0 = 0 \times 2 = 0$$

de l'origine.

Le bâton associé à la modalité « 1 enfant » est à distance (en cm)

$$\ell_1 = 1 \times 2 = 2$$

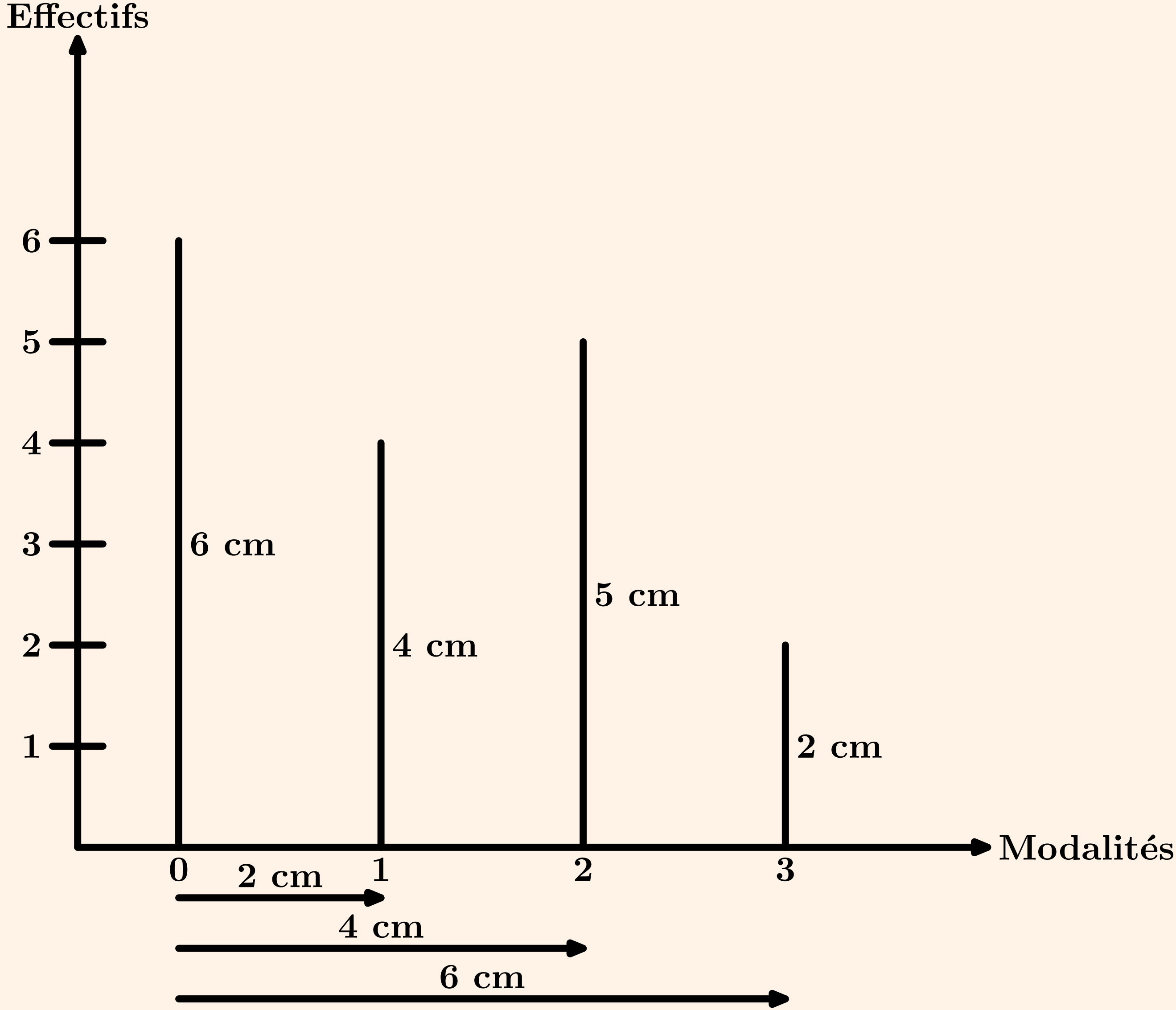
de l'origine.

De même $\ell_2 = 4$ cm et $\ell_3 = 6$ cm.

Pour la hauteur, on choisit $K = 1$ cm.

Le bâton associé à la modalité « 0 enfant » est de hauteur $h_0 = 6$ cm.

De même $h_1 = 4$ cm, $h_2 = 5$ cm et $h_3 = 2$ cm.



Effectifs
6
5
4
3
2
1
6 cm
4 cm
5 cm
2 cm
0
1
2
3
Modalités
2 cm
4 cm
6 cm

Lorsque la variable continue on utilise un **histogramme**.

Les modalités sont des intervalles

$$m_1 = [a_1, a_2[,\ m_2 = [a_2, a_3[, \dots, m_p = [a_p, a_{p+1}[.$$

L'amplitude de m_1 est $a_2 - a_1$, l'amplitude de m_2 est $a_3 - a_2$ *etc.*

À chaque modalité $[a_i, a_{i+1}[$ correspond un rectangle.

Les **largeurs** des rectangles sont proportionnelles aux **amplitudes** des modalités correspondantes.

Les **surfaces** des rectangles sont proportionnelles aux **effectifs** des modalités correspondantes.

Modalités	$[a_1, a_2[$	$[a_2, a_3[$	$\cdots$	$[a_p, a_{p+1}[$
Effectifs	n_1	n_2	$\cdots$	n_p

À chaque modalité $[a_i, a_{i+1}[$, on associe un rectangle de hauteur h_i et de largeur ℓ_i. Sa surface est donc $S_i = h_i \ell_i$.

Il existe un nombre $C > 0$ tel que

$$\ell_1 = C(a_2 - a_1), \ell_2 = C(a_3 - a_2), \ldots, \ell_p = C(a_{p+1} - a_p).$$

Il existe un nombre $K > 0$ tel que

$$S_1 = Kn_1, S_2 = Kn_2, \ldots, S_p = Kn_p.$$

P

Q

Pour tracer un rectangle, il faut connaître sa hauteur et sa largeur :

$$S_i = Kn_i$$

donc

$$h_i \ell_i = Kn_i$$

donc

$$h_i = \frac{Kn_i}{\ell_i}.$$

Or,

$$\ell_i = C(a_{i+1} - a_i)$$

donc

$$h_i = D\frac{n_i}{a_{i+1} - a_i}$$

avec

$$D = \frac{K}{C}.$$

Modalités	$[a_1, a_2[$	$[a_2, a_3[$	$\cdots$	$[a_p, a_{p+1}[$
Effectifs	n_1	n_2	$\cdots$	n_p

☞ Pour tracer un histogramme correspondant à la situation ci-dessus,

- on choisit des nombres $C > 0$ et $D > 0$
- à chaque modalité $[a_i, a_{i+1}[$, on associe un rectangle
  - de largeur $C(a_{i+1} - a_i)$
  - de hauteur $D\dfrac{n_i}{a_{i+1} - a_i}$.

Exemple

Âges d'un échantillon de 1000 personnes.

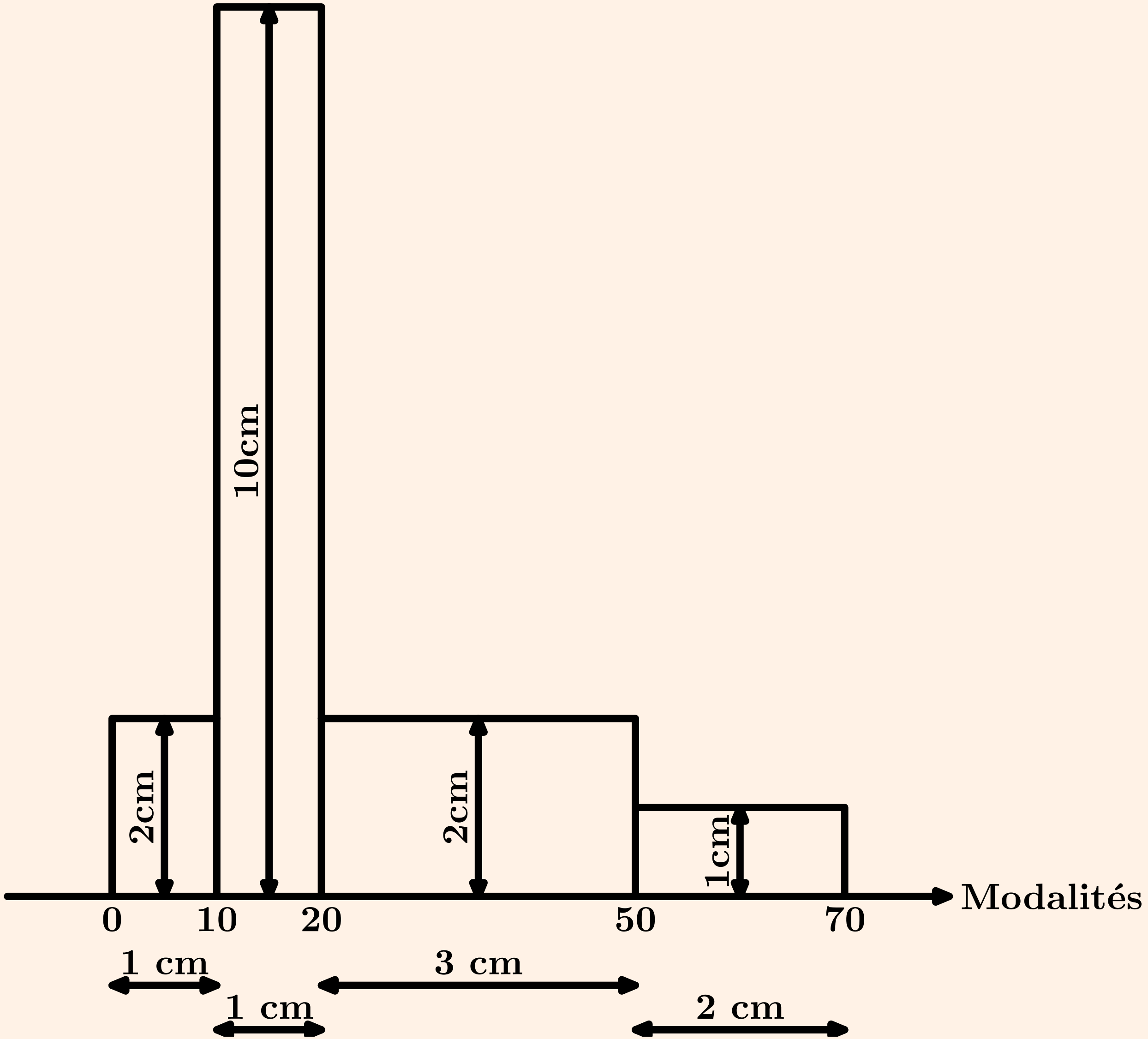
Modalités	0 à 9 ans	10 ans à 19 ans	20 ans à 49 ans	50 ans à 69 ans
Effectifs	100	500	300	100

$$
\begin{array}{llll}
m_1 = [0, 10[& a_1 = 0 & a_2 = 10 & n_1 = 100 \\
m_2 = [10, 20[& a_2 = 10 & a_3 = 20 & n_2 = 500 \\
m_3 = [20, 50[& a_3 = 20 & a_4 = 50 & n_3 = 300 \\
m_4 = [50, 70[& a_4 = 50 & a_5 = 70 & n_1 = 100.
\end{array}
$$

On choisit

$$C = 0,1 \text{ cm} \qquad D = 0,2 \text{ cm.}$$

Modalités	$[0,10[$	$[10,20[$	$[20,50[$	$[50,70[$
Effectifs	100	500	300	100
Largeurs (cm)	$(10-0)\times 0,1 = 1$	$(20-10)\times 0,1 = 1$	$(50-20)\times 0,1 = 3$	$(70-50)\times 0,1 = 2$
Hauteurs (cm)	$\frac{100}{10-0}\times 0,2 = 2$	$\frac{500}{20-10}\times 0,2 = 10$	$\frac{300}{50-20}\times 0,2 = 2$	$\frac{100}{70-50}\times 0,2 = 1$



10cm
2cm
2cm
1cm
Modalités
0
10
20
50
70
1 cm
1 cm
3 cm
2 cm