

Cherif El Msiyah  
Jaouad Madkour

# STATISTIQUE DESCRIPTIVE UNIVARIÉE

Exercices corrigés

Édition 2023



# Table des matières

<b>1</b>	<b>Introduction à la statistique</b>	<b>5</b>
1.1	Tableau statistique . . . . .	5
1.1.1	Variable quantitative discrète . . . . .	5
1.1.2	Variable quantitative continue . . . . .	10
1.2	Représentations graphiques . . . . .	16
1.2.1	Variable quantitative discrète . . . . .	16
1.2.2	Variable quantitative continue . . . . .	37
<b>2</b>	<b>Statistique descriptive univariée</b>	<b>63</b>
2.1	Statistique descriptive basée sur les moments . . . . .	63
2.1.1	Variable quantitative discrète . . . . .	63
2.1.2	Variable quantitative continue . . . . .	83
2.2	Statistique descriptive basée sur les quantiles . . . . .	101
2.2.1	Variable quantitative discrète . . . . .	101
2.2.2	Variable quantitative continue . . . . .	115



# Introduction

Ce manuel "*Statistique descriptive univariée : exercices corrigés*" est une collection d'exercices corrigés de statistique descriptive destinée aux étudiants de l'enseignement supérieur poursuivant leurs études dans des disciplines scientifiques au sens large (mathématiques, physique, chimie, biologie, sciences économiques, gestion...). L'objectif étant d'appliquer les concepts de la statistique sur différents jeux de données. En particulier, dans le chapitre 1 introductif à la statistique seront résolus des exercices relatifs à deux outils indispensables dans la description et la synthèse des données statistiques ; Il s'agit du tableau statistique (section 1.1) et des représentations graphiques (section 1.2). Dans le chapitre 2, seront résolus des exercices sur des mesures numériques basées sur les moments (section 2.1) et des mesures numériques basées sur les quantiles (section 2.2) dans le cas d'une seule variable statistique. Le cas bivarié et le cas multivarié feront l'objet d'éditions futures.



# Chapitre 1

## Introduction à la statistique

Dans ce chapitre introductif à la statistique, on s'intéressera à deux outils indispensables dans la description et la synthèse des données statistiques. Il s'agit du tableau statistique (section 1.1) et des représentations graphiques (section 1.2).

### 1.1 Tableau statistique

L'objet du tableau statistique est de donner des indicateurs clés, tels que les effectifs, les fréquences, les effectifs cumulés et les fréquences cumulées, sur chaque modalité d'une variable quantitative discrète (sous-section 1.1.1) ou sur chaque classe d'une variable quantitative continue (sous-section 1.1.2).

#### 1.1.1 Variable quantitative discrète

Une variable quantitative discrète  $X$  est une variable statistique qui prend en ensemble fini et dénombrable de valeurs numériques notées  $x_i$ .

##### Exercice 1 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	9	19	32
$n_i$	3	22	19

1. Construire le tableau statistique de la variable  $X$  ;
2. Quel est le nombre d'observations égales à 32 ?
3. Quelle est la proportion d'observations égales à 19 ?
4. Quel est le nombre d'observations au plus égales à 19 ?
5. Quel est le nombre d'observations au moins égales à 32 ?
6. Quelle est la proportion d'observations au plus égales à 9 ?
7. Quelle est la proportion d'observations au moins égales à 9 ?

##### Solution :

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
9	3	0.07	44	3	1	0.07
19	22	0.5	41	25	0.93	0.57
32	19	0.43	19	44	0.43	1
$\Sigma$	44	1	–	–	–	–

- Le nombre d'observations égales à la modalité  $x_3 = 32$  est donné par l'effectif de cette modalité, c'est à dire  $n_3 = 19$ .
- La proportion d'observations égales à la modalité  $x_2 = 19$  est donnée par la fréquence de cette modalité, c'est à dire  $f_2 = 0.5$ .
- Le nombre d'observations au plus égales à la modalité  $x_2 = 19$  est donné par l'effectif cumulé croissant de cette modalité, c'est à dire  $N_2^+ = 25$ .
- Le nombre d'observations au moins égales à la modalité  $x_3 = 32$  est donné par l'effectif cumulé décroissant de cette modalité, c'est à dire  $N_3^- = 19$ .
- La proportion d'observations au plus égales à la modalité  $x_1 = 9$  est donnée par la fréquence cumulée croissante de cette modalité, c'est à dire  $F_1^+ = 0.07$ .
- La proportion d'observations au moins égales à la modalité  $x_1 = 9$  est donnée par la fréquence cumulée décroissante de cette modalité, c'est à dire  $F_1^- = 1.0$ .

### Exercice 2 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	5	16	28	44
$n_i$	7	15	21	12

- Construire le tableau statistique de la variable  $X$  ;
- Quel est le nombre d'observations égales à 44 ?
- Quelle est la proportion d'observations égales à 44 ?
- Quel est le nombre d'observations au plus égales à 28 ?
- Quel est le nombre d'observations au moins égales à 28 ?
- Quelle est la proportion d'observations au plus égales à 16 ?
- Quelle est la proportion d'observations au moins égales à 28 ?

### Solution :

- Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
5	7	0.13	55	7	1	0.13
16	15	0.27	48	22	0.87	0.4
28	21	0.38	33	43	0.6	0.78
44	12	0.22	12	55	0.22	1
$\Sigma$	55	1	–	–	–	–

- Le nombre d'observations égales à la modalité  $x_4 = 44$  est donné par l'effectif de cette modalité, c'est à dire  $n_4 = 12$ .

- La proportion d'observations égales à la modalité  $x_4 = 44$  est donnée par la fréquence de cette modalité, c'est à dire  $f_4 = 0.22$ .
- Le nombre d'observations au plus égales à la modalité  $x_3 = 28$  est donné par l'effectif cumulé croissant de cette modalité, c'est à dire  $N_3^+ = 43$ .
- Le nombre d'observations au moins égales à la modalité  $x_3 = 28$  est donné par l'effectif cumulé décroissant de cette modalité, c'est à dire  $N_3^- = 33$ .
- La proportion d'observations au plus égales à la modalité  $x_2 = 16$  est donnée par la fréquence cumulée croissante de cette modalité, c'est à dire  $F_2^+ = 0.4$ .
- La proportion d'observations au moins égales à la modalité  $x_3 = 28$  est donnée par la fréquence cumulée décroissante de cette modalité, c'est à dire  $F_3^- = 0.6$ .

**Exercice 3 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	4	16	25	47	54
$n_i$	3	27	31	13	5

- Construire le tableau statistique de la variable  $X$  ;
- Quel est le nombre d'observations égales à 25 ?
- Quelle est la proportion d'observations égales à 16 ?
- Quel est le nombre d'observations au plus égales à 4 ?
- Quel est le nombre d'observations au moins égales à 25 ?
- Quelle est la proportion d'observations au plus égales à 25 ?
- Quelle est la proportion d'observations au moins égales à 16 ?

**Solution :**

- Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
4	3	0.04	79	3	1	0.04
16	27	0.34	76	30	0.96	0.38
25	31	0.39	49	61	0.62	0.77
47	13	0.16	18	74	0.23	0.93
54	5	0.07	5	79	0.07	1
$\Sigma$	79	1	-	-	-	-

- Le nombre d'observations égales à la modalité  $x_3 = 25$  est donné par l'effectif de cette modalité, c'est à dire  $n_3 = 31$ .
- La proportion d'observations égales à la modalité  $x_2 = 16$  est donnée par la fréquence de cette modalité, c'est à dire  $f_2 = 0.34$ .
- Le nombre d'observations au plus égales à la modalité  $x_1 = 4$  est donné par l'effectif cumulé croissant de cette modalité, c'est à dire  $N_1^+ = 3$ .

5. Le nombre d'observations au moins égales à la modalité  $x_3 = 25$  est donné par l'effectif cumulé décroissant de cette modalité, c'est à dire  $N_3^- = 49$ .
6. La proportion d'observations au plus égales à la modalité  $x_3 = 25$  est donnée par la fréquence cumulée croissante de cette modalité, c'est à dire  $F_3^+ = 0.77$ .
7. La proportion d'observations au moins égales à la modalité  $x_2 = 16$  est donnée par la fréquence cumulée décroissante de cette modalité, c'est à dire  $F_2^- = 0.96$ .

**Exercice 4 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	9	18	38
$n_i$	7	35	37

1. Construire le tableau statistique de la variable  $X$  ;
2. Quel est le nombre d'observations égales à 18 ?
3. Quelle est la proportion d'observations égales à 18 ?
4. Quel est le nombre d'observations au plus égales à 18 ?
5. Quel est le nombre d'observations au moins égales à 18 ?
6. Quelle est la proportion d'observations au plus égales à 38 ?
7. Quelle est la proportion d'observations au moins égales à 9 ?

**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
9	7	0.09	79	7	1	0.09
18	35	0.44	72	42	0.91	0.53
38	37	0.47	37	79	0.47	1
$\Sigma$	79	1	–	–	–	–

2. Le nombre d'observations égales à la modalité  $x_2 = 18$  est donné par l'effectif de cette modalité, c'est à dire  $n_2 = 35$ .
3. La proportion d'observations égales à la modalité  $x_2 = 18$  est donnée par la fréquence de cette modalité, c'est à dire  $f_2 = 0.44$ .
4. Le nombre d'observations au plus égales à la modalité  $x_2 = 18$  est donné par l'effectif cumulé croissant de cette modalité, c'est à dire  $N_2^+ = 42$ .
5. Le nombre d'observations au moins égales à la modalité  $x_2 = 18$  est donné par l'effectif cumulé décroissant de cette modalité, c'est à dire  $N_2^- = 72$ .
6. La proportion d'observations au plus égales à la modalité  $x_3 = 38$  est donnée par la fréquence cumulée croissante de cette modalité, c'est à dire  $F_3^+ = 1.0$ .
7. La proportion d'observations au moins égales à la modalité  $x_1 = 9$  est donnée par la fréquence cumulée décroissante de cette modalité, c'est à dire  $F_1^- = 1.0$ .

**Exercice 5 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	4	11	32	41
$n_i$	3	19	23	11

1. Construire le tableau statistique de la variable  $X$  ;
2. Quel est le nombre d'observations égales à 4 ?
3. Quelle est la proportion d'observations égales à 32 ?
4. Quel est le nombre d'observations au plus égales à 4 ?
5. Quel est le nombre d'observations au moins égales à 11 ?
6. Quelle est la proportion d'observations au plus égales à 32 ?
7. Quelle est la proportion d'observations au moins égales à 11 ?

**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
4	3	0.05	56	3	1	0.05
11	19	0.34	53	22	0.95	0.39
32	23	0.41	34	45	0.61	0.8
41	11	0.2	11	56	0.2	1
$\Sigma$	56	1	-	-	-	-

2. Le nombre d'observations égales à la modalité  $x_1 = 4$  est donné par l'effectif de cette modalité, c'est à dire  $n_1 = 3$ .
3. La proportion d'observations égales à la modalité  $x_3 = 32$  est donnée par la fréquence de cette modalité, c'est à dire  $f_3 = 0.41$ .
4. Le nombre d'observations au plus égales à la modalité  $x_1 = 4$  est donné par l'effectif cumulé croissant de cette modalité, c'est à dire  $N_1^+ = 3$ .
5. Le nombre d'observations au moins égales à la modalité  $x_2 = 11$  est donné par l'effectif cumulé décroissant de cette modalité, c'est à dire  $N_2^- = 53$ .
6. La proportion d'observations au plus égales à la modalité  $x_3 = 32$  est donnée par la fréquence cumulée croissante de cette modalité, c'est à dire  $F_3^+ = 0.8$ .
7. La proportion d'observations au moins égales à la modalité  $x_2 = 11$  est donnée par la fréquence cumulée décroissante de cette modalité, c'est à dire  $F_2^- = 0.95$ .

**Exercice 6 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	7	17	27	44	59
$n_i$	3	12	15	19	2

1. Construire le tableau statistique de la variable  $X$  ;

2. Quel est le nombre d'observations égales à 17 ?
3. Quelle est la proportion d'observations égales à 17 ?
4. Quel est le nombre d'observations au plus égales à 27 ?
5. Quel est le nombre d'observations au moins égales à 59 ?
6. Quelle est la proportion d'observations au plus égales à 27 ?
7. Quelle est la proportion d'observations au moins égales à 44 ?

**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
7	3	0.06	51	3	1	0.06
17	12	0.24	48	15	0.94	0.3
27	15	0.29	36	30	0.7	0.59
44	19	0.37	21	49	0.41	0.96
59	2	0.04	2	51	0.04	1
$\Sigma$	51	1	–	–	–	–

2. Le nombre d'observations égales à la modalité  $x_2 = 17$  est donné par l'effectif de cette modalité, c'est à dire  $n_2 = 12$ .
3. La proportion d'observations égales à la modalité  $x_2 = 17$  est donnée par la fréquence de cette modalité, c'est à dire  $f_2 = 0.24$ .
4. Le nombre d'observations au plus égales à la modalité  $x_3 = 27$  est donné par l'effectif cumulé croissant de cette modalité, c'est à dire  $N_3^+ = 30$ .
5. Le nombre d'observations au moins égales à la modalité  $x_5 = 59$  est donné par l'effectif cumulé décroissant de cette modalité, c'est à dire  $N_5^- = 2$ .
6. La proportion d'observations au plus égales à la modalité  $x_3 = 27$  est donnée par la fréquence cumulée croissante de cette modalité, c'est à dire  $F_3^+ = 0.59$ .
7. La proportion d'observations au moins égales à la modalité  $x_4 = 44$  est donnée par la fréquence cumulée décroissante de cette modalité, c'est à dire  $F_4^- = 0.41$ .

### 1.1.2 Variable quantitative continue

Une variable quantitative continue  $X$  est une variable statistique qui prend en ensemble infini de valeurs numériques regroupées dans des classes notées  $[x_i^-, x_i^+]$ .

**Exercice 7 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+]$	[5, 16[	[16, 31[	[31, 49[
$n_i$	4	38	1

1. Construire le tableau statistique de la variable  $X$  ;
2. Quel est le nombre d'observations comprises entre 5 et 16 ?

3. Quelle est la proportion d'observations comprises entre 16 et 31 ?
4. Quel est le nombre d'observations au plus égales à 31 ?
5. Quel est le nombre d'observations au moins égales à 16 ?
6. Quelle est la proportion d'observations au plus égales à 31 ?
7. Quelle est la proportion d'observations au moins égales à 16 ?

**Solution :**

1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[5, 16[$	4	0.09	43	4	1	0.09
$[16, 31[$	38	0.88	39	42	0.91	0.97
$[31, 49[$	1	0.03	1	43	0.03	1
$\Sigma$	43	1	—	—	—	—

2. Le nombre d'observations comprises entre 5 et 16 est donné par l'effectif de la classe  $[x_1^-, x_1^+[ = [5, 16[$ , c'est à dire  $n_1 = 4$ .
3. La proportion d'observations comprises entre 16 et 31 est donnée par la fréquence de la classe  $[x_2^-, x_2^+[ = [16, 31[$ , c'est à dire  $f_2 = 0.88$ .
4. Le nombre d'observations au plus égales à 31 est donné par l'effectif cumulé croissant de la classe dont la borne supérieure est égale à 31, en l'occurrence la classe  $[x_2^-, x_2^+[ = [16, 31[$ . C'est à dire  $N_2^+ = 42$ .
5. Le nombre d'observations au moins égales à 16 est donné par l'effectif cumulé décroissant de la classe dont la borne inférieure est égale à 16, en l'occurrence la classe  $[x_2^-, x_2^+[ = [16, 31[$ . C'est à dire  $N_2^- = 39$ .
6. La proportion d'observations au plus égales à 31 est donnée par la fréquence cumulée croissante de la classe dont la borne supérieure est égale à 31, en l'occurrence la classe  $[x_2^-, x_2^+[ = [16, 31[$ . C'est à dire  $F_2^+ = 0.97$ .
7. La proportion d'observations au moins égales à 16 est donnée par la fréquence cumulée décroissante de la classe dont la borne inférieure est égale à 16, en l'occurrence la classe  $[x_2^-, x_2^+[ = [16, 31[$ . C'est à dire  $F_2^- = 0.91$ .

**Exercice 8 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+[$	$[1, 19[$	$[19, 27[$	$[27, 46[$	$[46, 52[$
$n_i$	8	22	25	4

1. Construire le tableau statistique de la variable  $X$  ;
2. Quel est le nombre d'observations comprises entre 27 et 46 ?
3. Quelle est la proportion d'observations comprises entre 19 et 27 ?
4. Quel est le nombre d'observations au plus égales à 19 ?
5. Quel est le nombre d'observations au moins égales à 27 ?

6. Quelle est la proportion d'observations au plus égales à 19 ?
7. Quelle est la proportion d'observations au moins égales à 27 ?

**Solution :**

1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[1, 19[$	8	0.14	59	8	1	0.14
$[19, 27[$	22	0.37	51	30	0.86	0.51
$[27, 46[$	25	0.42	29	55	0.49	0.93
$[46, 52[$	4	0.07	4	59	0.07	1.0
$\Sigma$	59	1	—	—	—	—

2. Le nombre d'observations comprises entre 27 et 46 est donné par l'effectif de la classe  $[x_3^-, x_3^+[ = [27, 46[$ , c'est à dire  $n_3 = 25$ .
3. La proportion d'observations comprises entre 19 et 27 est donnée par la fréquence de la classe  $[x_2^-, x_2^+[ = [19, 27[$ , c'est à dire  $f_2 = 0.37$ .
4. Le nombre d'observations au plus égales à 19 est donné par l'effectif cumulé croissant de la classe dont la borne supérieure est égale à 19, en l'occurrence la classe  $[x_1^-, x_1^+[ = [1, 19[$ . C'est à dire  $N_1^+ = 8$ .
5. Le nombre d'observations au moins égales à 27 est donné par l'effectif cumulé décroissant de la classe dont la borne inférieure est égale à 27, en l'occurrence la classe  $[x_3^-, x_3^+[ = [27, 46[$ . C'est à dire  $N_3^- = 29$ .
6. La proportion d'observations au plus égales à 19 est donnée par la fréquence cumulée croissante de la classe dont la borne supérieure est égale à 19, en l'occurrence la classe  $[x_1^-, x_1^+[ = [1, 19[$ . C'est à dire  $F_1^+ = 0.14$ .
7. La proportion d'observations au moins égales à 27 est donnée par la fréquence cumulée décroissante de la classe dont la borne inférieure est égale à 27, en l'occurrence la classe  $[x_3^-, x_3^+[ = [27, 46[$ . C'est à dire  $F_3^- = 0.49$ .

**Exercice 9 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+[$	$[9, 16[$	$[16, 25[$	$[25, 43[$	$[43, 58[$	$[58, 78[$
$n_i$	3	12	21	18	2

1. Construire le tableau statistique de la variable  $X$  ;
2. Quel est le nombre d'observations comprises entre 9 et 16 ?
3. Quelle est la proportion d'observations comprises entre 9 et 16 ?
4. Quel est le nombre d'observations au plus égales à 43 ?
5. Quel est le nombre d'observations au moins égales à 43 ?
6. Quelle est la proportion d'observations au plus égales à 25 ?
7. Quelle est la proportion d'observations au moins égales à 43 ?

**Solution :**

1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[9, 16[$	3	0.05	56	3	1	0.05
$[16, 25[$	12	0.21	53	15	0.95	0.26
$[25, 43[$	21	0.38	41	36	0.74	0.64
$[43, 58[$	18	0.32	20	54	0.36	0.96
$[58, 78[$	2	0.04	2	56	0.04	1
$\Sigma$	56	1	–	–	–	–

- Le nombre d'observations comprises entre 9 et 16 est donné par l'effectif de la classe  $[x_1^-, x_1^+[ = [9, 16[$ , c'est à dire  $n_1 = 3$ .
- La proportion d'observations comprises entre 9 et 16 est donnée par la fréquence de la classe  $[x_1^-, x_1^+[ = [9, 16[$ , c'est à dire  $f_1 = 0.05$ .
- Le nombre d'observations au plus égales à 43 est donné par l'effectif cumulé croissant de la classe dont la borne supérieure est égale à 43, en l'occurrence la classe  $[x_3^-, x_3^+[ = [25, 43[$ . C'est à dire  $N_3^+ = 36$ .
- Le nombre d'observations au moins égales à 43 est donné par l'effectif cumulé décroissant de la classe dont la borne inférieure est égale à 43, en l'occurrence la classe  $[x_4^-, x_4^+[ = [43, 58[$ . C'est à dire  $N_4^- = 20$ .
- La proportion d'observations au plus égales à 25 est donnée par la fréquence cumulée croissante de la classe dont la borne supérieure est égale à 25, en l'occurrence la classe  $[x_2^-, x_2^+[ = [16, 25[$ . C'est à dire  $F_2^+ = 0.26$ .
- La proportion d'observations au moins égales à 43 est donnée par la fréquence cumulée décroissante de la classe dont la borne inférieure est égale à 43, en l'occurrence la classe  $[x_4^-, x_4^+[ = [43, 58[$ . C'est à dire  $F_4^- = 0.36$ .

**Exercice 10 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+[$	$[1, 18[$	$[18, 37[$	$[37, 46[$
$n_i$	3	32	11

- Construire le tableau statistique de la variable  $X$  ;
- Quel est le nombre d'observations comprises entre 18 et 37 ?
- Quelle est la proportion d'observations comprises entre 18 et 37 ?
- Quel est le nombre d'observations au plus égales à 37 ?
- Quel est le nombre d'observations au moins égales à 18 ?
- Quelle est la proportion d'observations au plus égales à 18 ?
- Quelle est la proportion d'observations au moins égales à 18 ?

**Solution :**

1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[1, 18[$	3	0.07	46	3	1	0.07
$[18, 37[$	32	0.7	43	35	0.93	0.77
$[37, 46[$	11	0.23	11	46	0.23	1
$\Sigma$	46	1	–	–	–	–

- Le nombre d'observations comprises entre 18 et 37 est donné par l'effectif de la classe  $[x_2^-, x_2^+[ = [18, 37[$ , c'est à dire  $n_2 = 32$ .
- La proportion d'observations comprises entre 18 et 37 est donnée par la fréquence de la classe  $[x_2^-, x_2^+[ = [18, 37[$ , c'est à dire  $f_2 = 0.7$ .
- Le nombre d'observations au plus égales à 37 est donné par l'effectif cumulé croissant de la classe dont la borne supérieure est égale à 37, en l'occurrence la classe  $[x_2^-, x_2^+[ = [18, 37[$ . C'est à dire  $N_2^+ = 35$ .
- Le nombre d'observations au moins égales à 18 est donné par l'effectif cumulé décroissant de la classe dont la borne inférieure est égale à 18, en l'occurrence la classe  $[x_2^-, x_2^+[ = [18, 37[$ . C'est à dire  $N_2^- = 43$ .
- La proportion d'observations au plus égales à 18 est donnée par la fréquence cumulée croissante de la classe dont la borne supérieure est égale à 18, en l'occurrence la classe  $[x_1^-, x_1^+[ = [1, 18[$ . C'est à dire  $F_1^+ = 0.07$ .
- La proportion d'observations au moins égales à 18 est donnée par la fréquence cumulée décroissante de la classe dont la borne inférieure est égale à 18, en l'occurrence la classe  $[x_2^-, x_2^+[ = [18, 37[$ . C'est à dire  $F_2^- = 0.93$ .

### Exercice 11 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+[$	$n_i$	$n_i$	$n_i$	$n_i$
$[2, 19[$	2	22	26	8

- Construire le tableau statistique de la variable  $X$  ;
- Quel est le nombre d'observations comprises entre 2 et 19 ?
- Quelle est la proportion d'observations comprises entre 19 et 32 ?
- Quel est le nombre d'observations au plus égales à 19 ?
- Quel est le nombre d'observations au moins égales à 32 ?
- Quelle est la proportion d'observations au plus égales à 32 ?
- Quelle est la proportion d'observations au moins égales à 19 ?

### Solution :

- Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[2, 19[$	2	0.03	58	2	1	0.03
$[19, 32[$	22	0.38	56	24	0.97	0.41
$[32, 47[$	26	0.45	34	50	0.59	0.86
$[47, 59[$	8	0.14	8	58	0.14	1.0
$\Sigma$	58	1	–	–	–	–

- Le nombre d'observations comprises entre 2 et 19 est donné par l'effectif de la classe  $[x_1^-, x_1^+ [= [2, 19[$ , c'est à dire  $n_1 = 2$ .
- La proportion d'observations comprises entre 19 et 32 est donnée par la fréquence de la classe  $[x_2^-, x_2^+ [= [19, 32[$ , c'est à dire  $f_2 = 0.38$ .
- Le nombre d'observations au plus égales à 19 est donné par l'effectif cumulé croissant de la classe dont la borne supérieure est égale à 19, en l'occurrence la classe  $[x_1^-, x_1^+ [= [2, 19[$ . C'est à dire  $N_1^+ = 2$ .
- Le nombre d'observations au moins égales à 32 est donné par l'effectif cumulé décroissant de la classe dont la borne inférieure est égale à 32, en l'occurrence la classe  $[x_3^-, x_3^+ [= [32, 47[$ . C'est à dire  $N_3^- = 34$ .
- La proportion d'observations au plus égales à 32 est donnée par la fréquence cumulée croissante de la classe dont la borne supérieure est égale à 32, en l'occurrence la classe  $[x_2^-, x_2^+ [= [19, 32[$ . C'est à dire  $F_2^+ = 0.41$ .
- La proportion d'observations au moins égales à 19 est donnée par la fréquence cumulée décroissante de la classe dont la borne inférieure est égale à 19, en l'occurrence la classe  $[x_2^-, x_2^+ [= [19, 32[$ . C'est à dire  $F_2^- = 0.97$ .

**Exercice 12 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+ [$	$[3, 11[$	$[11, 31[$	$[31, 41[$	$[41, 55[$	$[55, 78[$
$n_i$	9	16	26	15	3

- Construire le tableau statistique de la variable  $X$  ;
- Quel est le nombre d'observations comprises entre 41 et 55 ?
- Quelle est la proportion d'observations comprises entre 3 et 11 ?
- Quel est le nombre d'observations au plus égales à 55 ?
- Quel est le nombre d'observations au moins égales à 41 ?
- Quelle est la proportion d'observations au plus égales à 55 ?
- Quelle est la proportion d'observations au moins égales à 31 ?

**Solution :**

- Tableau statistique :

$[x_i^-, x_i^+ [$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[3, 11[$	9	0.13	69	9	1	0.13
$[11, 31[$	16	0.23	60	25	0.87	0.36
$[31, 41[$	26	0.38	44	51	0.64	0.74
$[41, 55[$	15	0.22	18	66	0.26	0.96
$[55, 78[$	3	0.04	3	69	0.04	1
$\Sigma$	69	1	-	-	-	-

- Le nombre d'observations comprises entre 41 et 55 est donné par l'effectif de la classe  $[x_4^-, x_4^+ [= [41, 55[$ , c'est à dire  $n_4 = 15$ .

3. La proportion d'observations comprises entre 3 et 11 est donnée par la fréquence de la classe  $[x_1^-, x_1^+ [= [3, 11[$ , c'est à dire  $f_1 = 0.13$ .
4. Le nombre d'observations au plus égales à 55 est donné par l'effectif cumulé croissant de la classe dont la borne supérieure est égale à 55, en l'occurrence la classe  $[x_4^-, x_4^+ [= [41, 55[$ . C'est à dire  $N_4^+ = 66$ .
5. Le nombre d'observations au moins égales à 41 est donné par l'effectif cumulé décroissant de la classe dont la borne inférieure est égale à 41, en l'occurrence la classe  $[x_4^-, x_4^+ [= [41, 55[$ . C'est à dire  $N_4^- = 18$ .
6. La proportion d'observations au plus égales à 55 est donnée par la fréquence cumulée croissante de la classe dont la borne supérieure est égale à 55, en l'occurrence la classe  $[x_4^-, x_4^+ [= [41, 55[$ . C'est à dire  $F_4^+ = 0.96$ .
7. La proportion d'observations au moins égales à 31 est donnée par la fréquence cumulée décroissante de la classe dont la borne inférieure est égale à 31, en l'occurrence la classe  $[x_3^-, x_3^+ [= [31, 41[$ . C'est à dire  $F_3^- = 0.64$ .

## Conclusion

Dans cette section, les effectifs, les fréquences, les effectifs cumulés et les fréquences cumulées de variables quantitatives ont été synthétisés dans des tableaux statistiques. Dans la section suivante, chacune des colonnes du tableau statistique sera représentée par un type particulier de graphique.

## 1.2 Représentations graphiques

Chacune des colonnes du tableau statistique peut être représentée par un graphique particulier. Le choix du graphique est intimement lié à la nature discrète (sous-section 1.2.1) ou continue (sous-section 1.2.2) de la variable statistique étudiée.

### 1.2.1 Variable quantitative discrète

Une variable quantitative discrète  $X$  est une variable statistique qui prend en ensemble fini et dénombrable de valeurs numériques notées  $x_i$ .

#### Exercice 13 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	14	38	56
$n_i$	13	20	15

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;

7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$ .

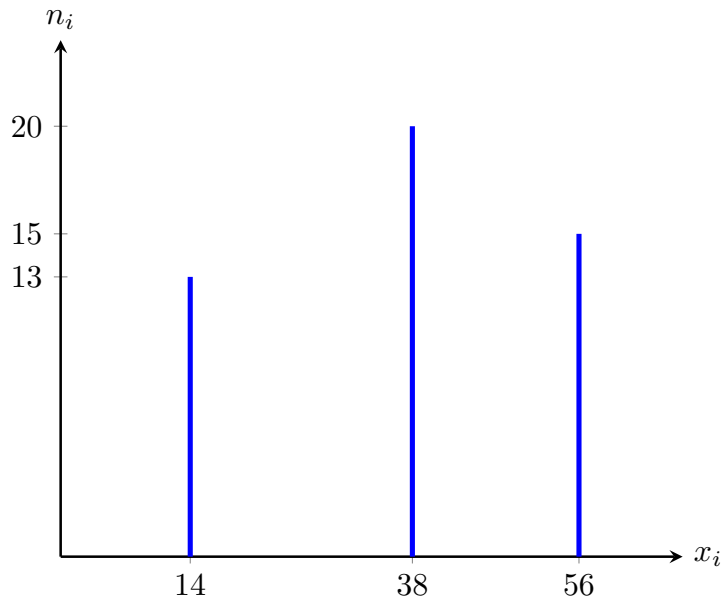
**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
14	13	0.27	48	13	1	0.27
38	20	0.42	35	33	0.73	0.69
56	15	0.31	15	48	0.31	1
$\Sigma$	48	1	—	—	—	—

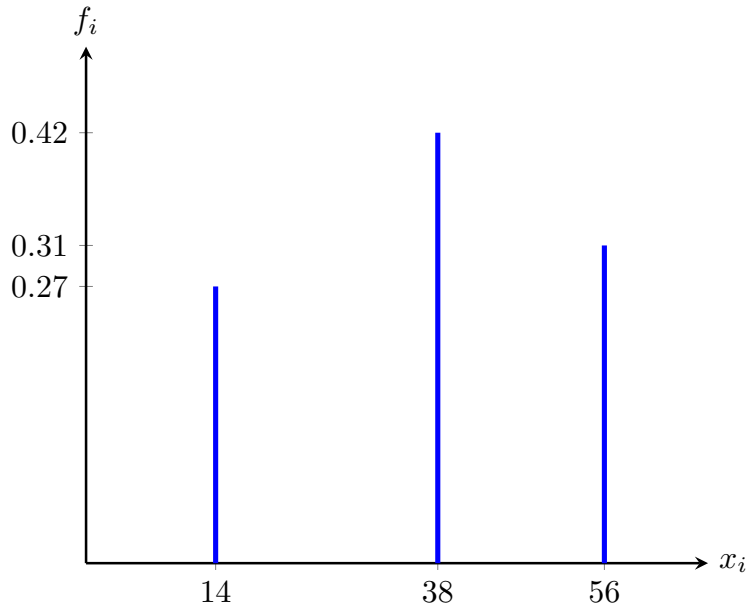
2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide du diagramme en bâtons suivant :



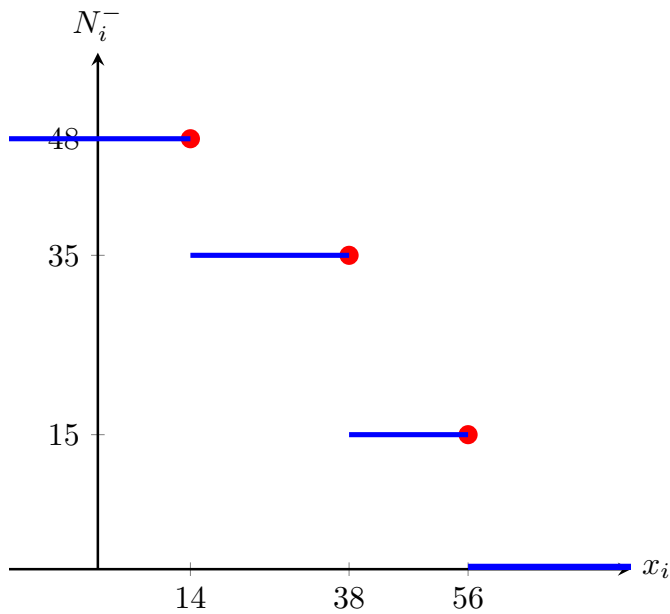
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide du diagramme en bâtons suivant :



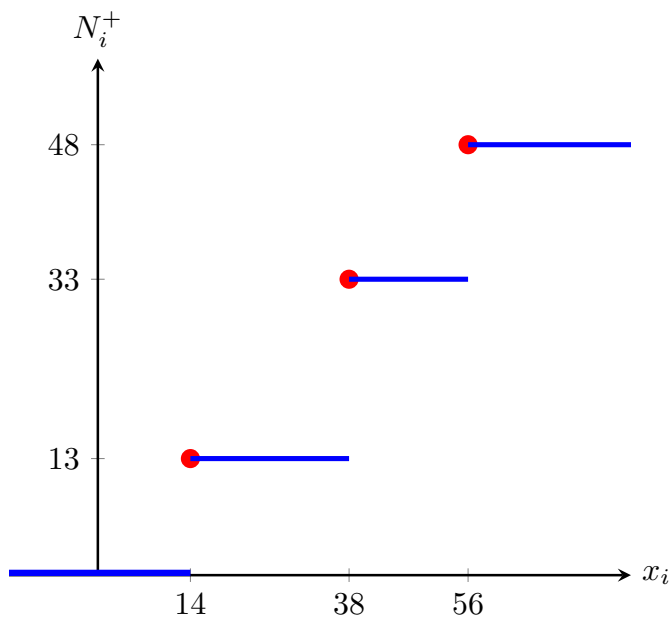
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



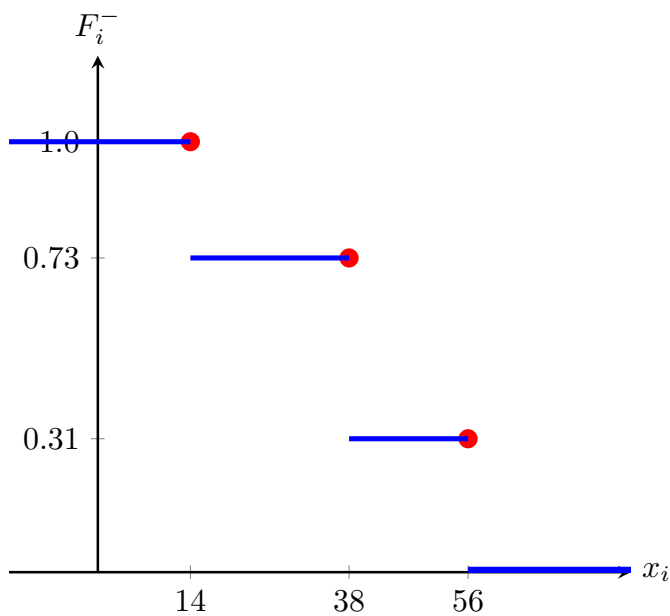
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



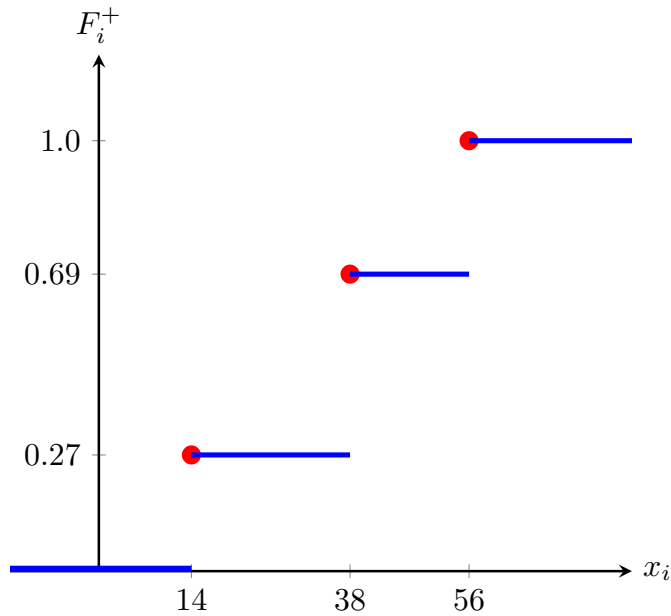
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :

**Exercice 14 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	15	30	42	60
$n_i$	14	29	21	19

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$  .

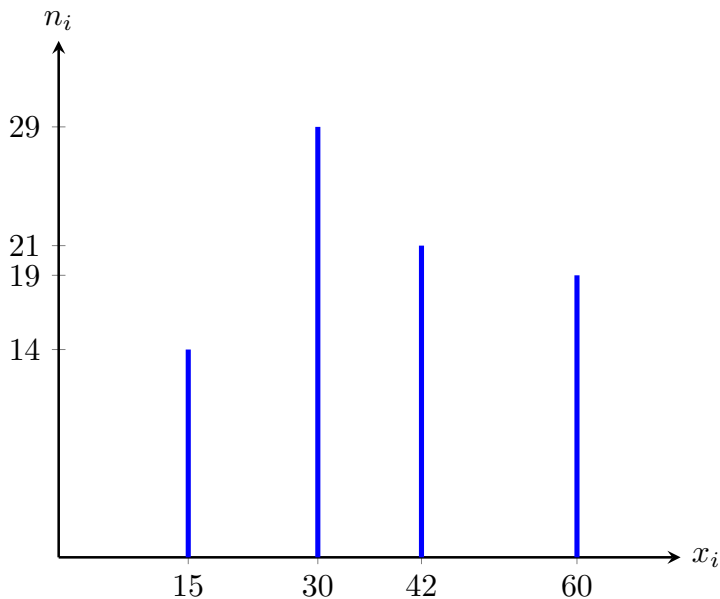
**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
15	14	0.17	83	14	1	0.17
30	29	0.35	69	43	0.83	0.52
42	21	0.25	40	64	0.48	0.77
60	19	0.23	19	83	0.23	1
$\Sigma$	83	1	-	-	-	-

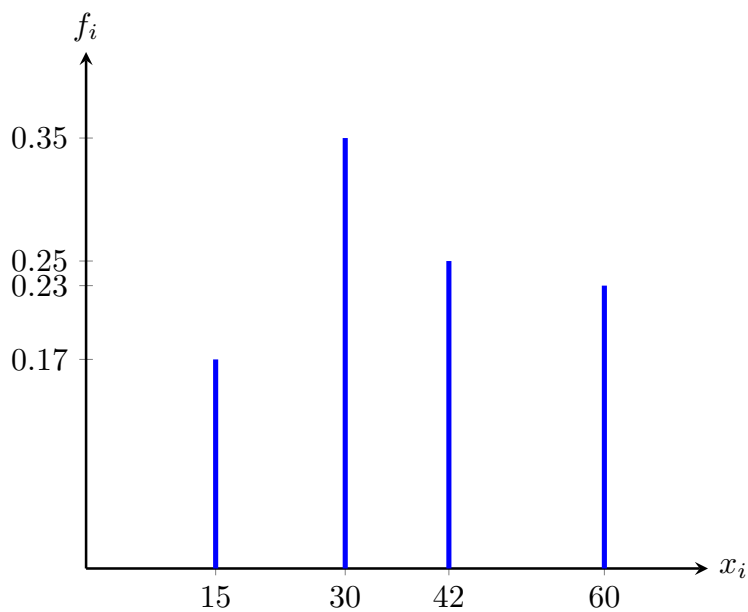
2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide du diagramme en bâtons suivant :



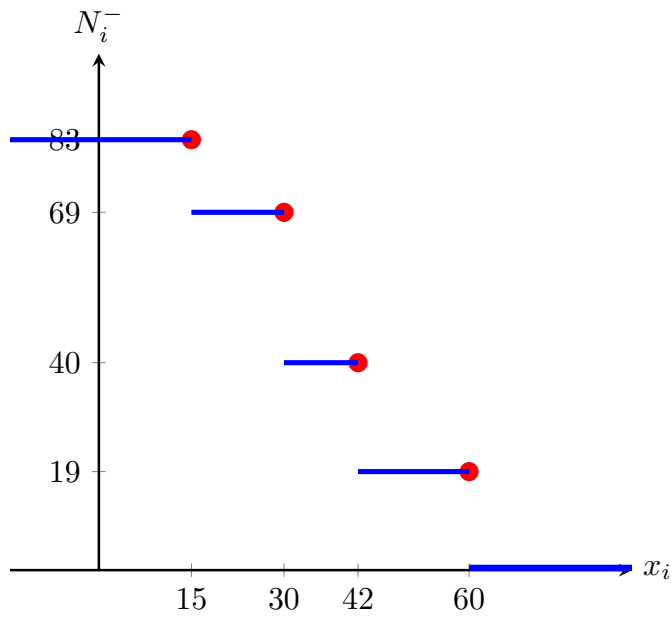
### 3. Représentation graphique des fréquences $f_i$ :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide du diagramme en bâtons suivant :



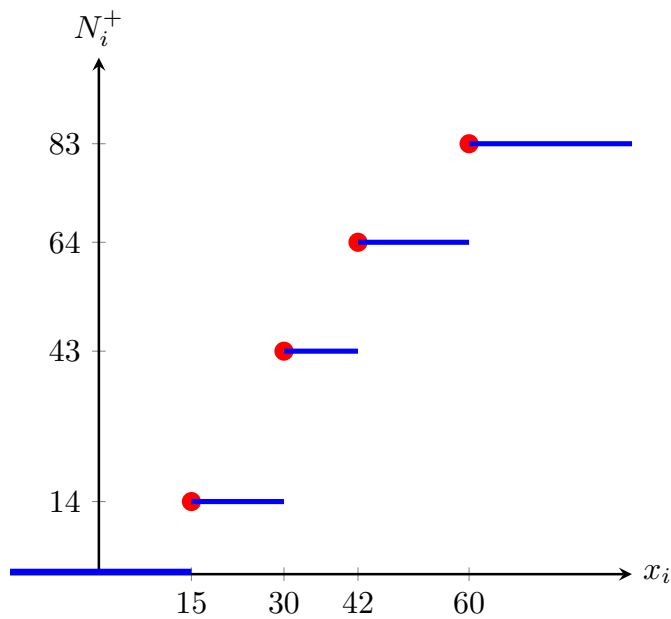
### 4. Représentation graphique des effectifs cumulés décroissants $N_i^-$ :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



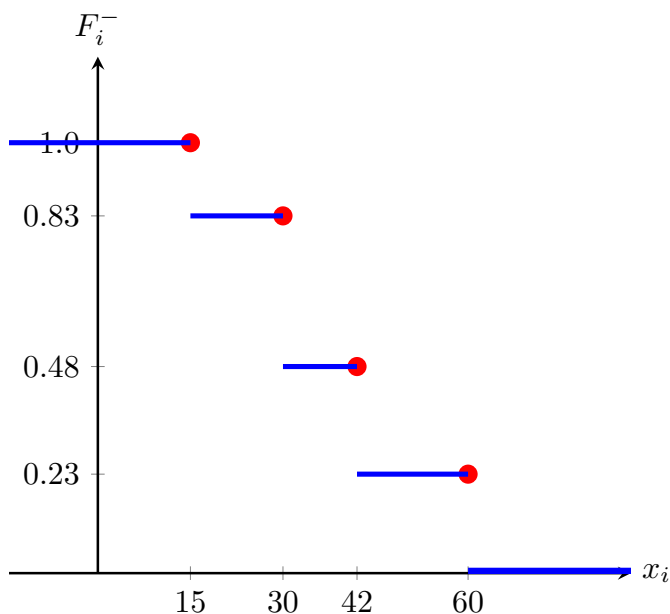
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



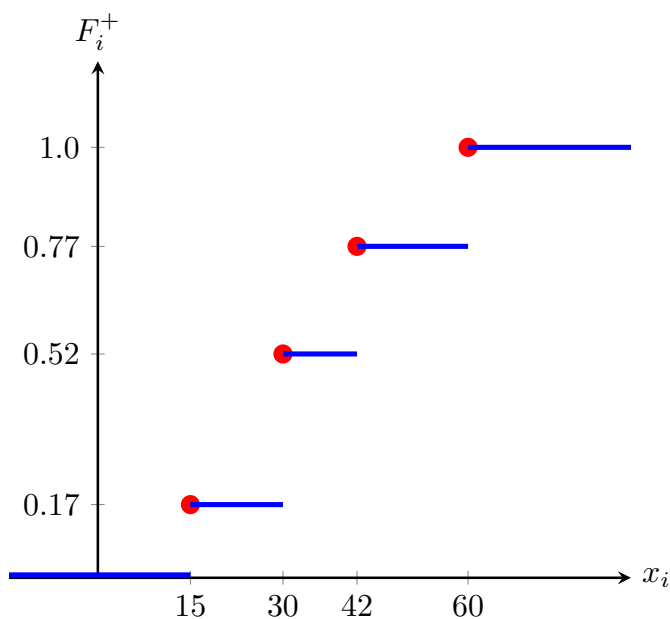
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 15 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	13	32	45	64	74
$n_i$	12	25	31	15	2

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$ .

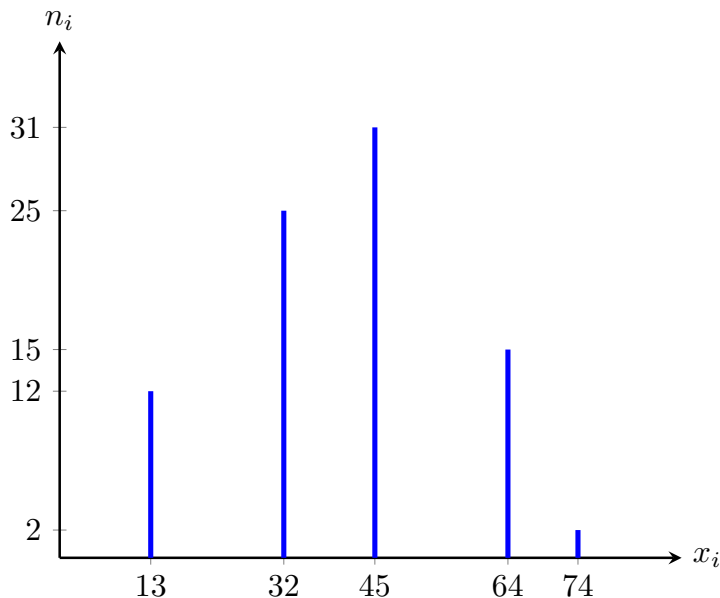
**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
13	12	0.14	85	12	1	0.14
32	25	0.29	73	37	0.86	0.43
45	31	0.36	48	68	0.57	0.79
64	15	0.18	17	83	0.21	0.97
74	2	0.03	2	85	0.03	1
$\Sigma$	85	1	–	–	–	–

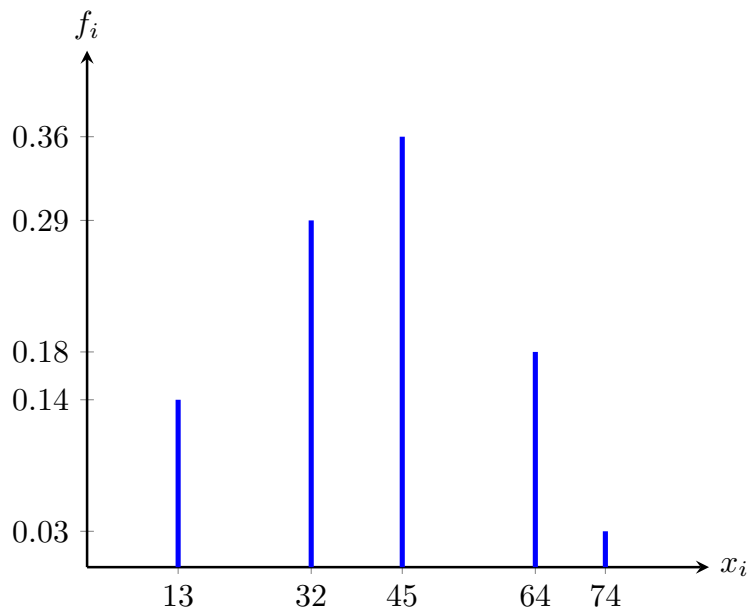
2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide du diagramme en bâtons suivant :



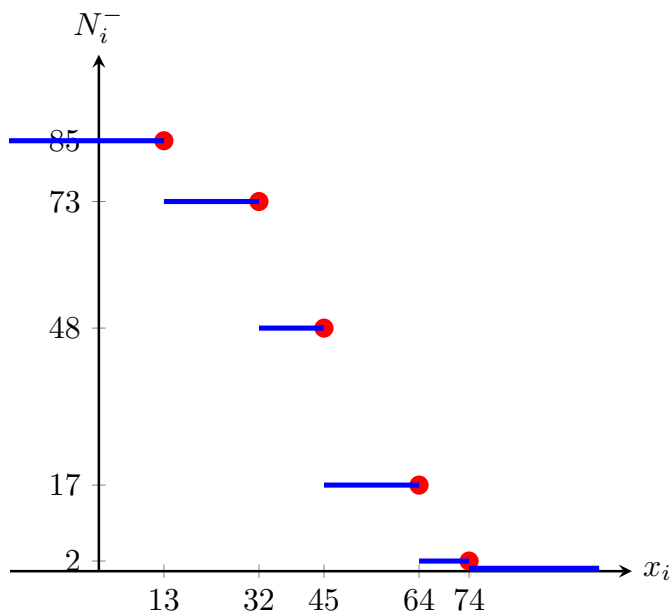
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide du diagramme en bâtons suivant :



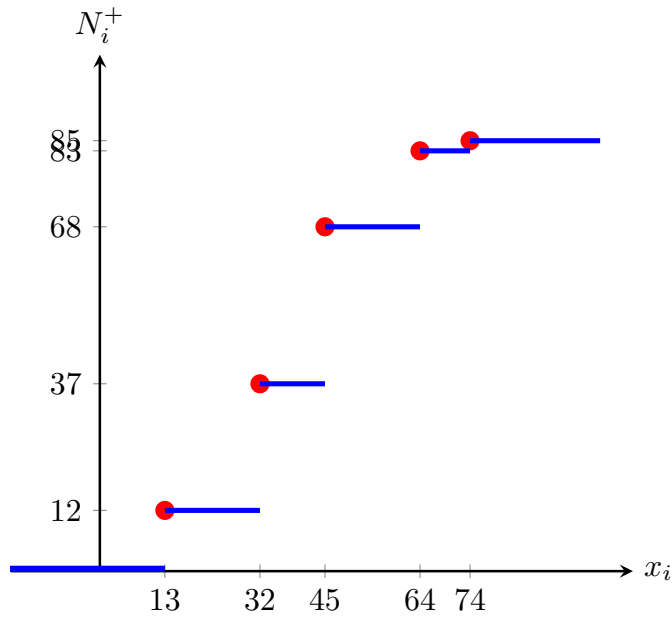
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



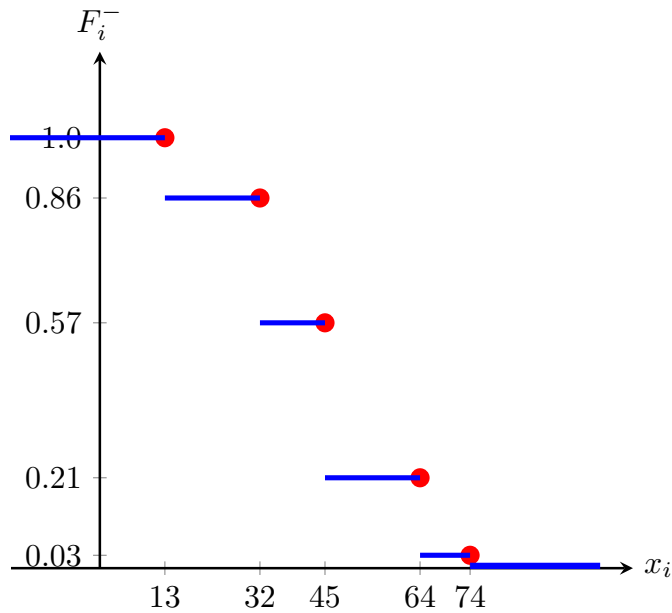
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



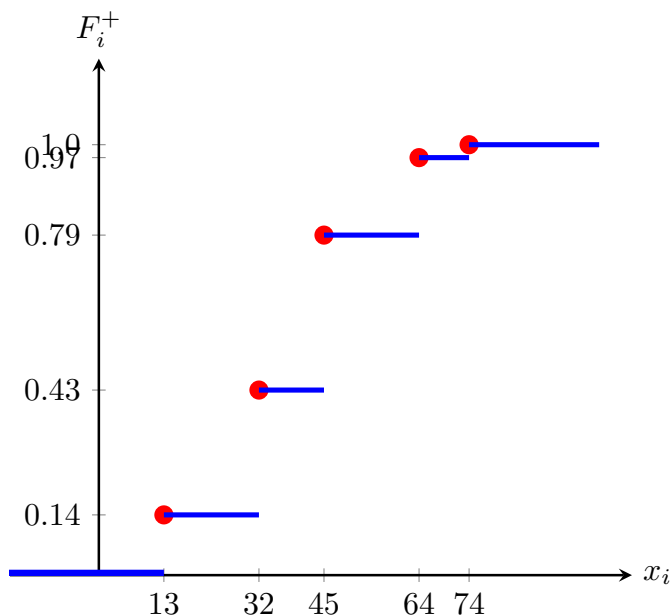
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :

**Exercice 16 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	16	39	55
$n_i$	11	21	19

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$  .

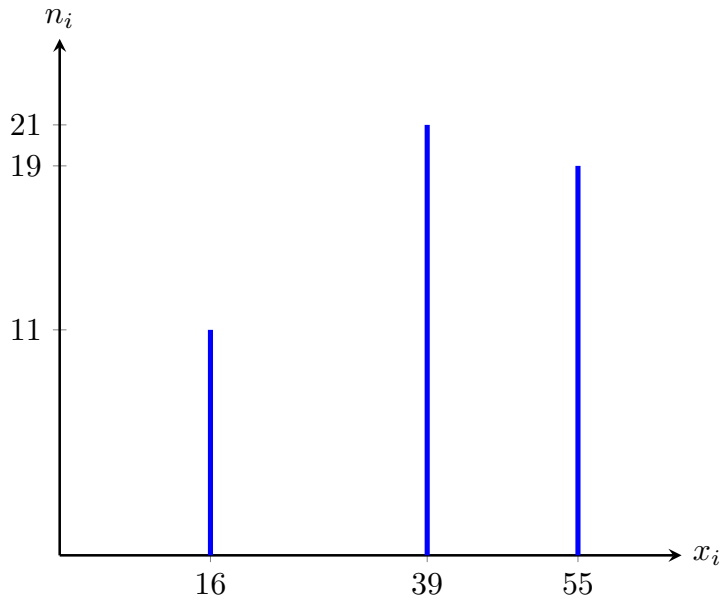
**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
16	11	0.22	51	11	1	0.22
39	21	0.41	40	32	0.78	0.63
55	19	0.37	19	51	0.37	1
$\Sigma$	51	1	—	—	—	—

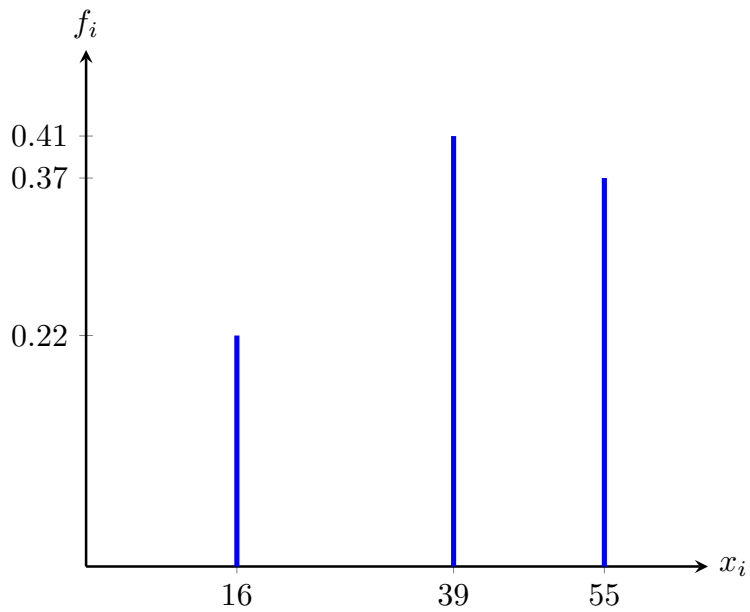
2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide du diagramme en bâtons suivant :



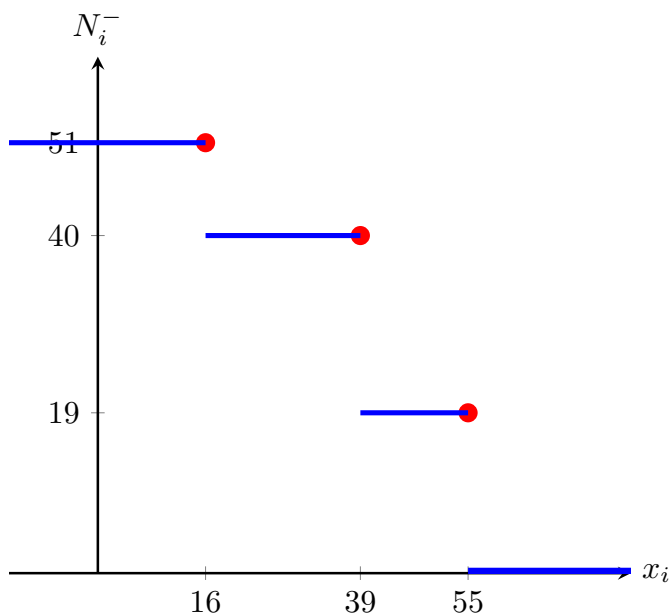
### 3. Représentation graphique des fréquences $f_i$ :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide du diagramme en bâtons suivant :



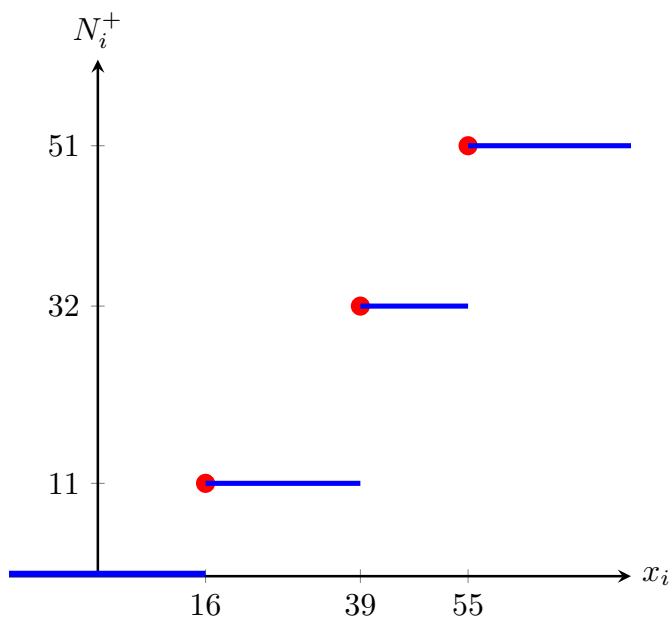
### 4. Représentation graphique des effectifs cumulés décroissants $N_i^-$ :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



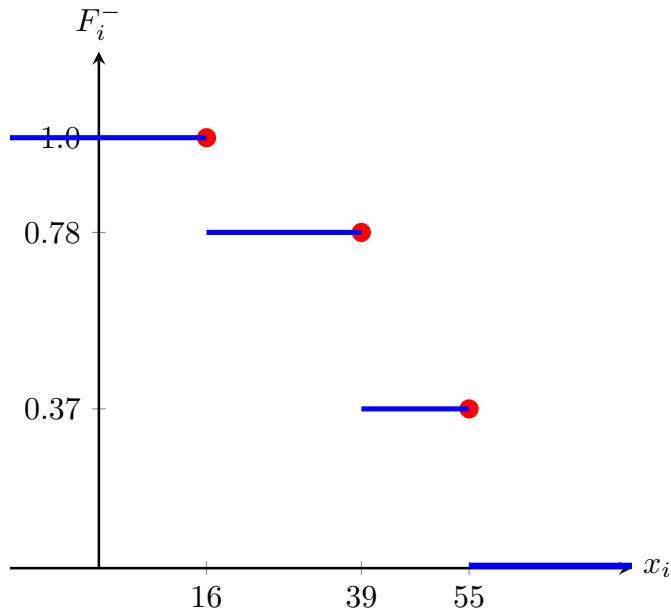
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



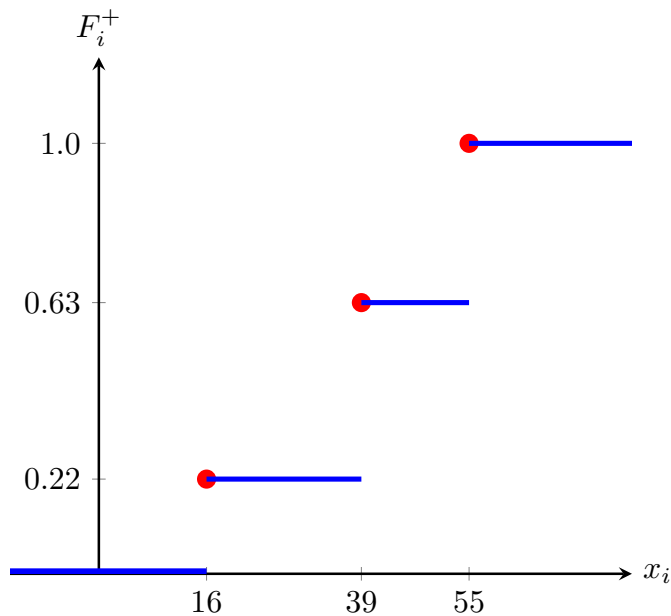
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 17 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	14	32	45	56
$n_i$	13	29	24	16

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$ .

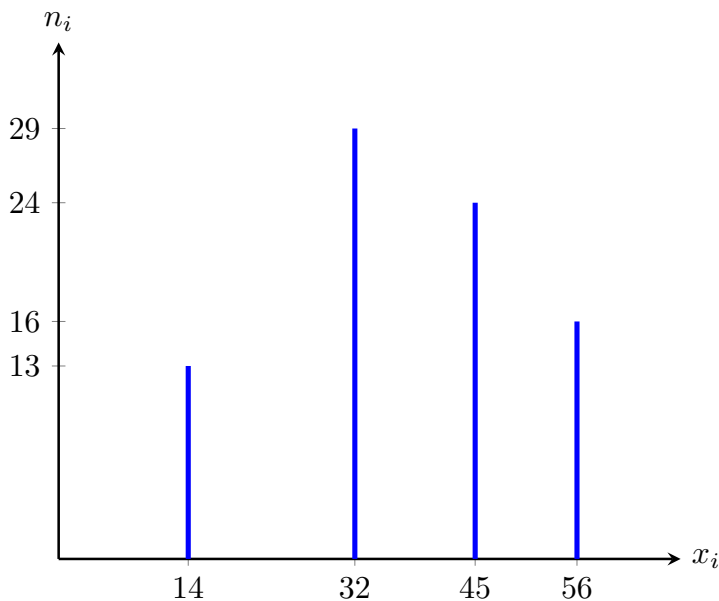
**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
14	13	0.16	82	13	1	0.16
32	29	0.35	69	42	0.84	0.51
45	24	0.29	40	66	0.49	0.8
56	16	0.2	16	82	0.2	1
$\Sigma$	82	1	–	–	–	–

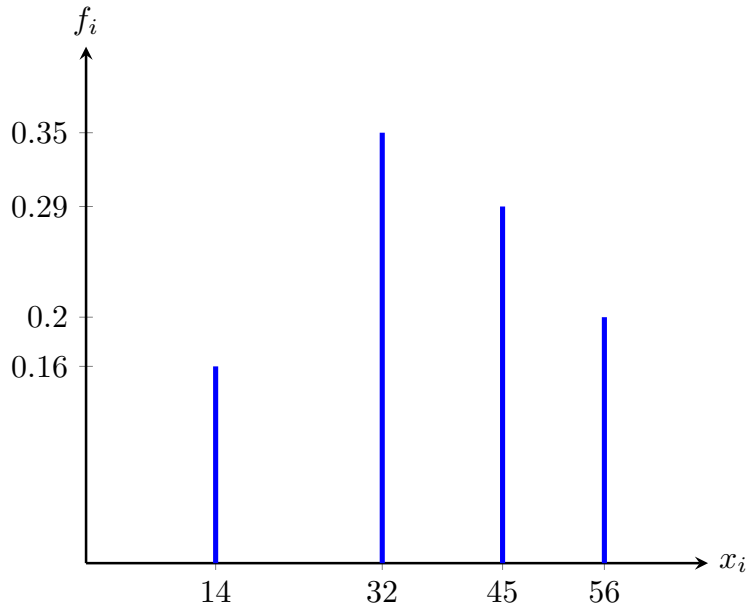
2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide du diagramme en bâtons suivant :



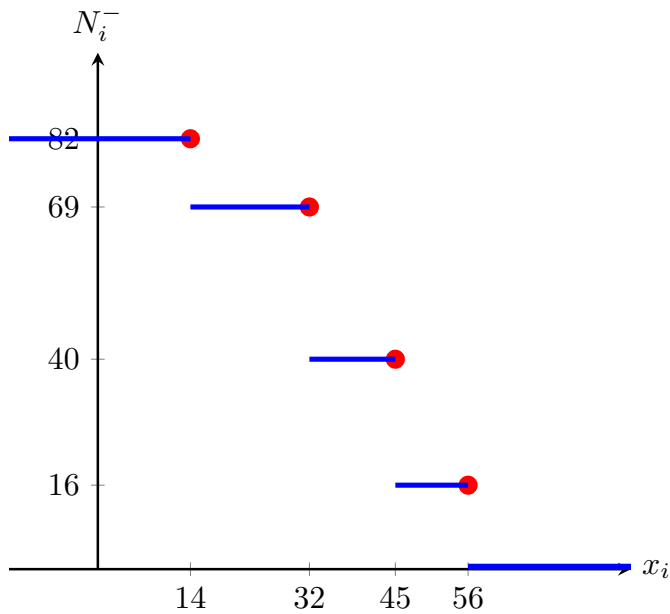
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide du diagramme en bâtons suivant :



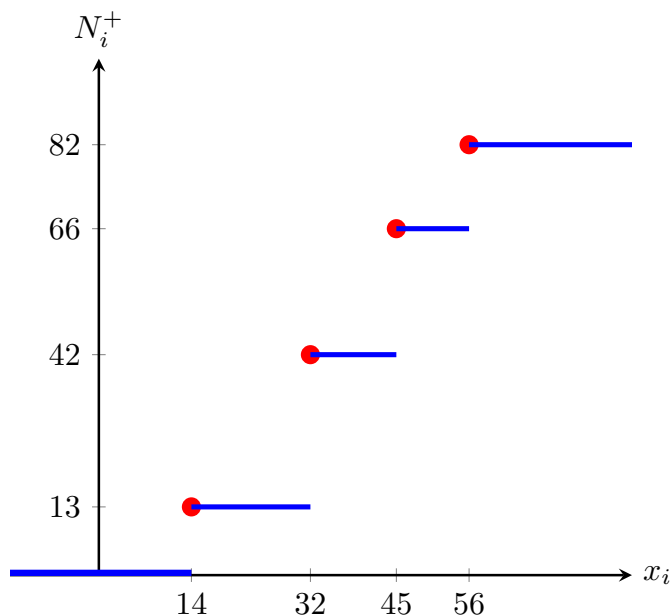
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



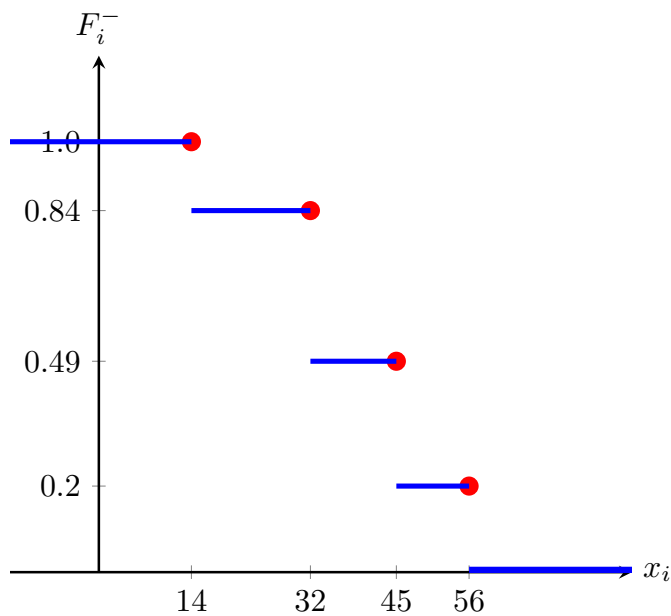
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



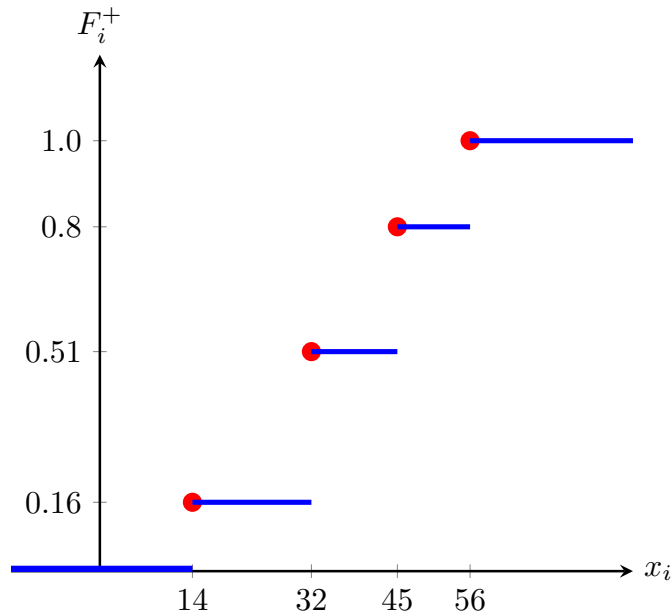
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :

**Exercice 18 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  :

$x_i$	11	27	49	61	70
$n_i$	14	28	32	17	9

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$  .

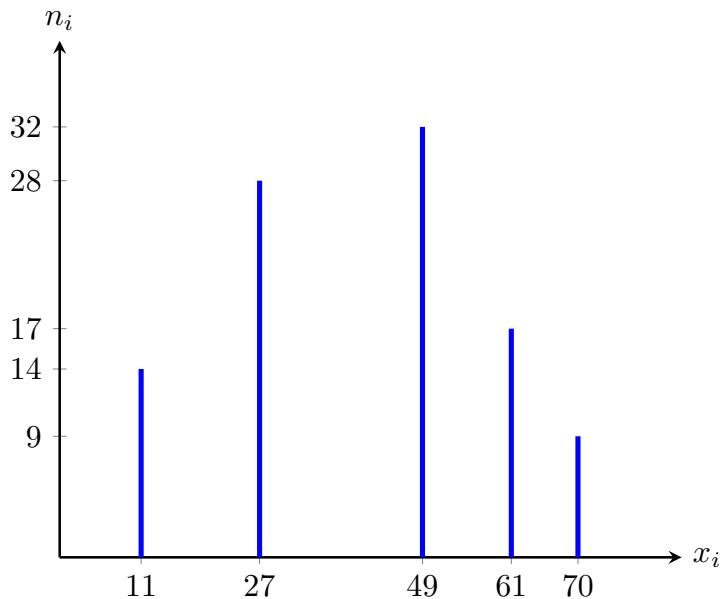
**Solution :**

1. Tableau statistique :

$x_i$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
11	14	0.14	100	14	1	0.14
27	28	0.28	86	42	0.86	0.42
49	32	0.32	58	74	0.58	0.74
61	17	0.17	26	91	0.26	0.91
70	9	0.09	9	100	0.09	1
$\Sigma$	100	1	—	—	—	—

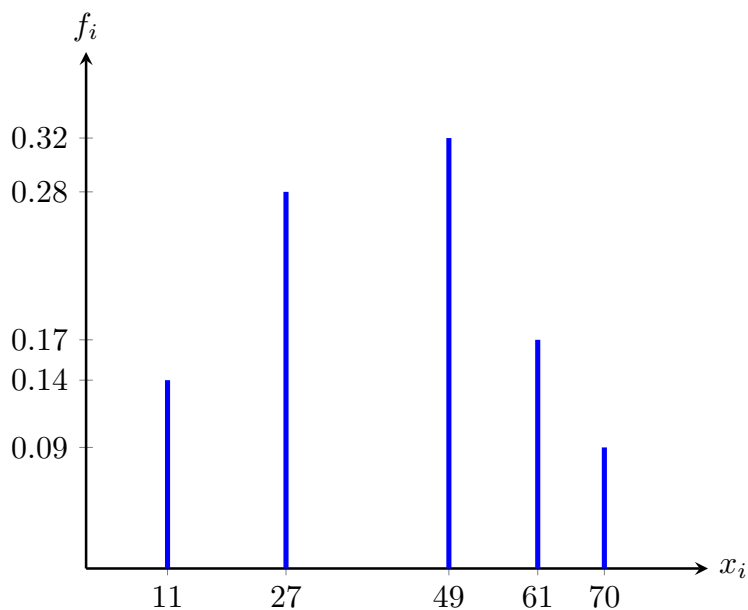
2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide du diagramme en bâtons suivant :



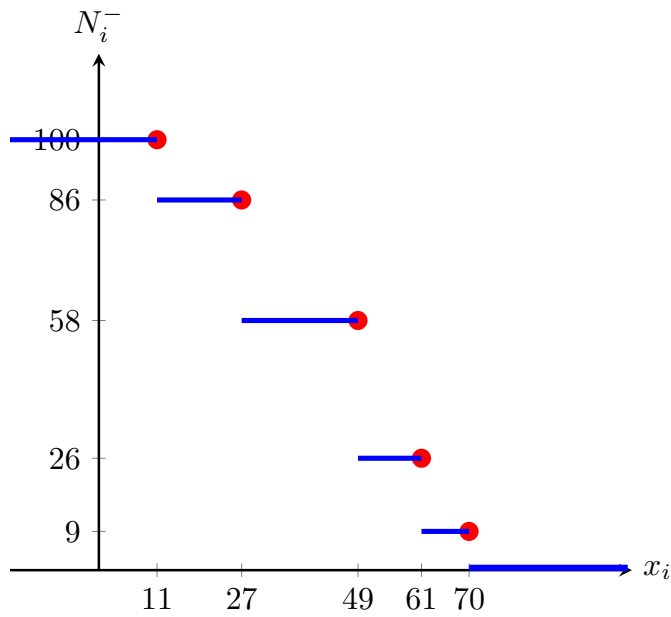
### 3. Représentation graphique des fréquences $f_i$ :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide du diagramme en bâtons suivant :



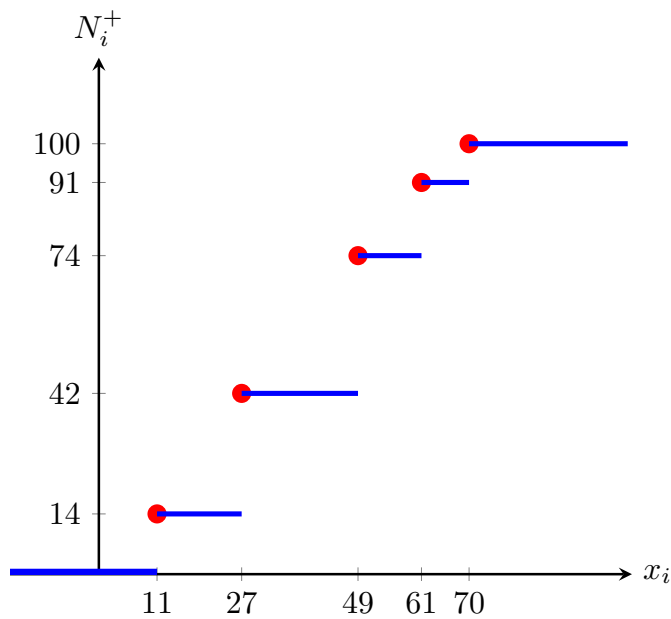
### 4. Représentation graphique des effectifs cumulés décroissants $N_i^-$ :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



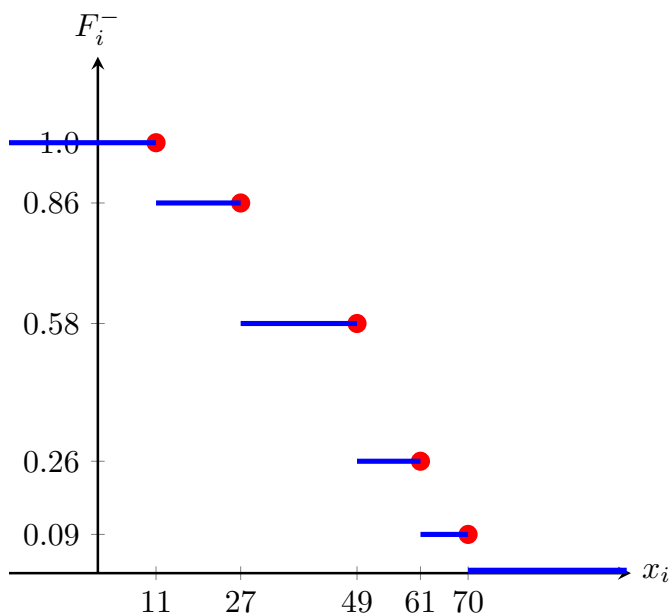
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



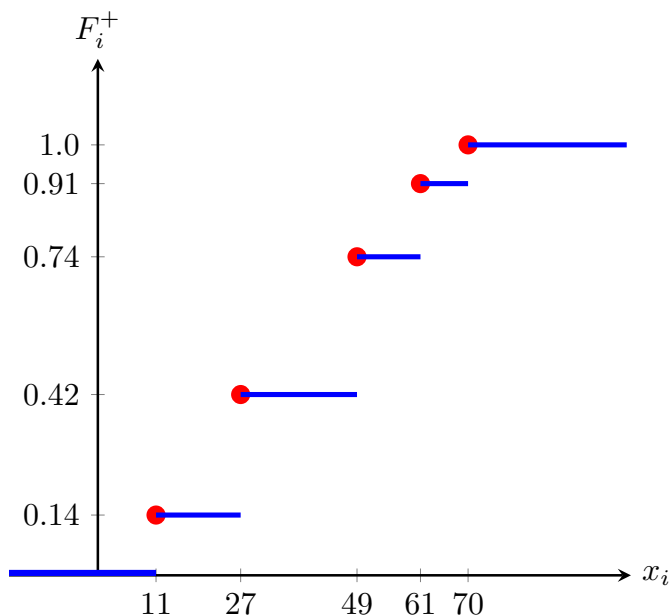
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



### 1.2.2 Variable quantitative continue

Une variable quantitative continue  $X$  est une variable statistique qui prend en ensemble infini de valeurs numériques regroupées dans des classes notées  $[x_i^-, x_i^+]$ .

**Exercice 19 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+]$	$[10, 29[$	$[29, 46[$	$[46, 57[$
$n_i$	11	23	16

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$  .

**Solution :**

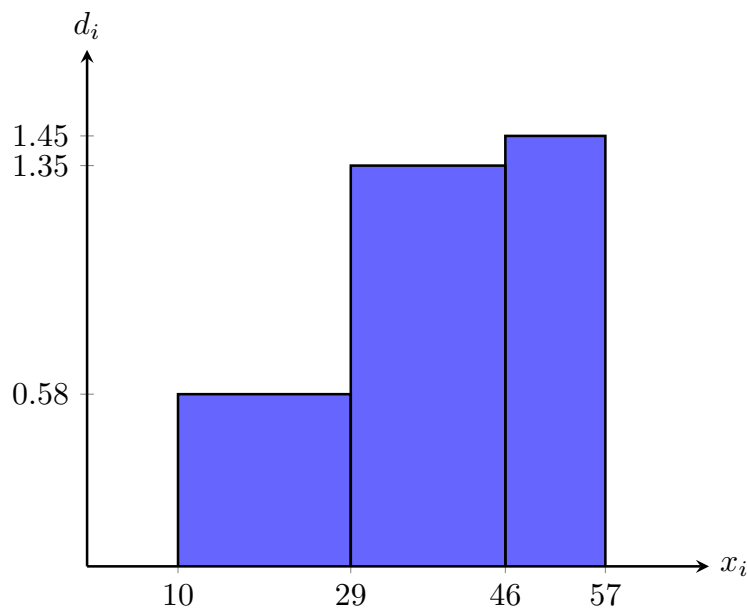
1. Tableau statistique :

$[x_i^-, x_i^+]$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[10, 29[$	11	0.22	50	11	1	0.22
$[29, 46[$	23	0.46	39	34	0.78	0.68
$[46, 57[$	16	0.32	16	50	0.32	1.0
$\Sigma$	50	1	—	—	—	—

2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités d'effectifs  $d_i$  en divisant les effectifs  $n_i$  des classes par leurs amplitudes respectives  $a_i$  :

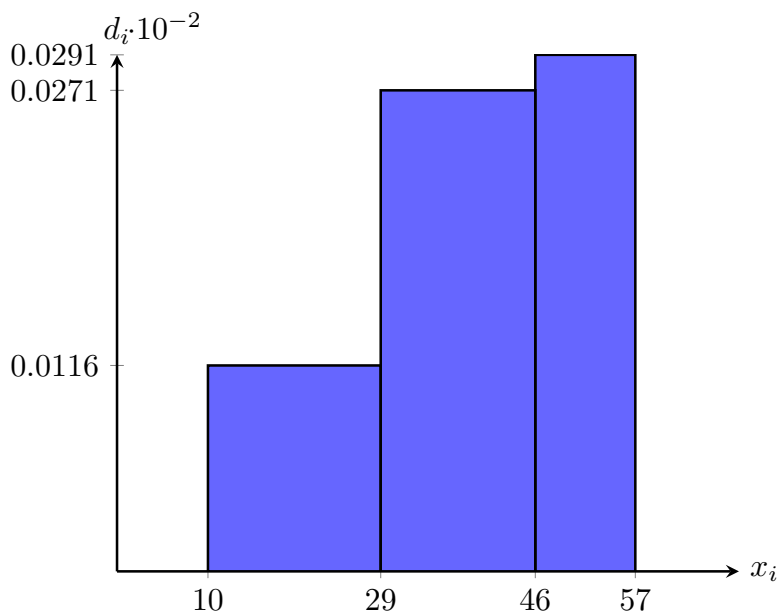
$[x_i^-, x_i^+]$	$[10, 29[$	$[29, 46[$	$[46, 57[$
$n_i$	11	23	16
$a_i$	19	17	11
$d_i$	0.58	1.35	1.45



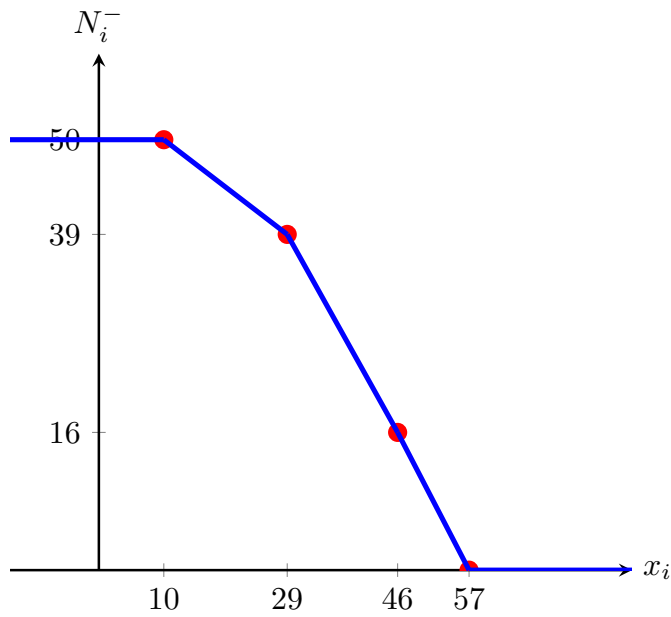
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités de fréquences  $d_i$  en divisant les fréquences  $f_i$  des classes par leurs amplitudes respectives  $a_i$  :

$[x_i^-, x_i^+]$	$[10, 29[$	$[29, 46[$	$[46, 57[$
$f_i$	0.22	0.46	0.32
$a_i$	19	17	11
$d_i$	0.0116	0.0271	0.0291

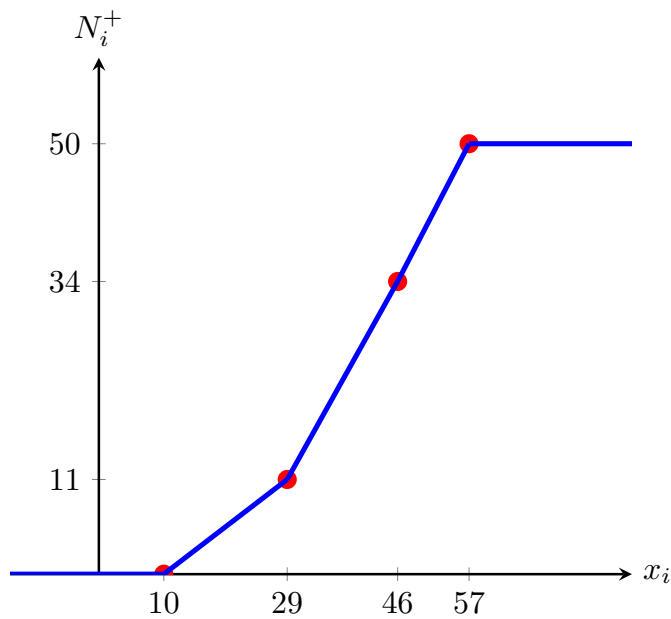
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



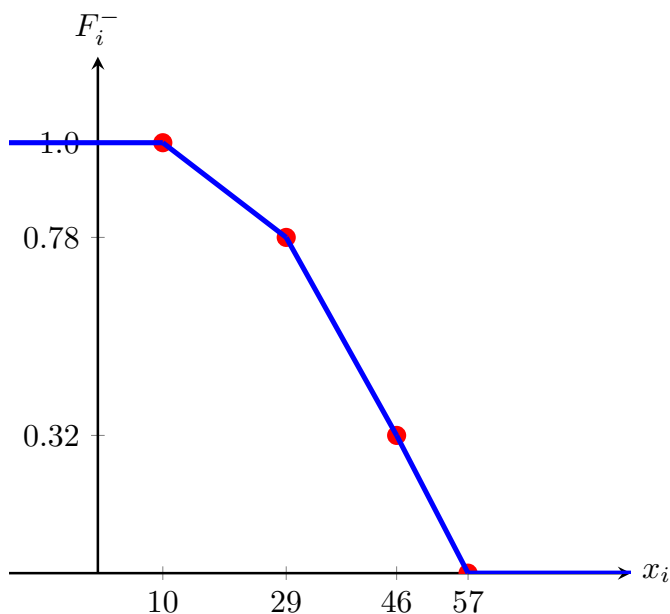
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



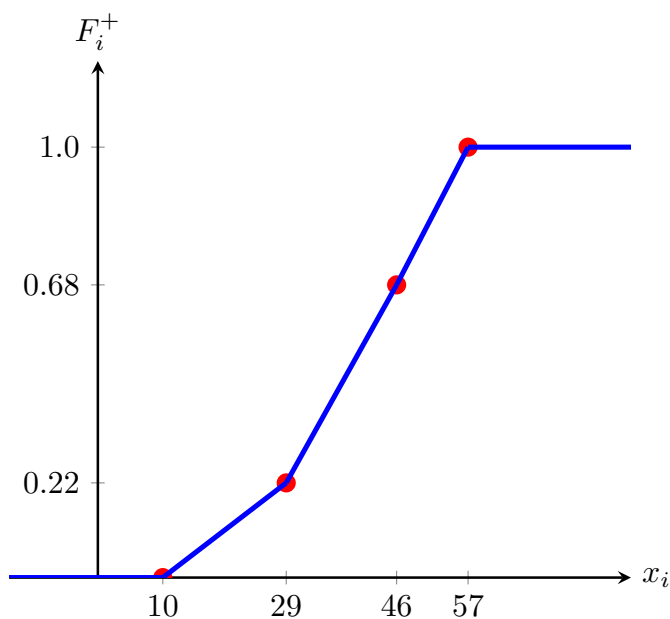
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 20 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+[$	$[18, 25[$	$[25, 48[$	$[48, 60[$	$[60, 73[$
$n_i$	10	28	21	19

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$  .

**Solution :**

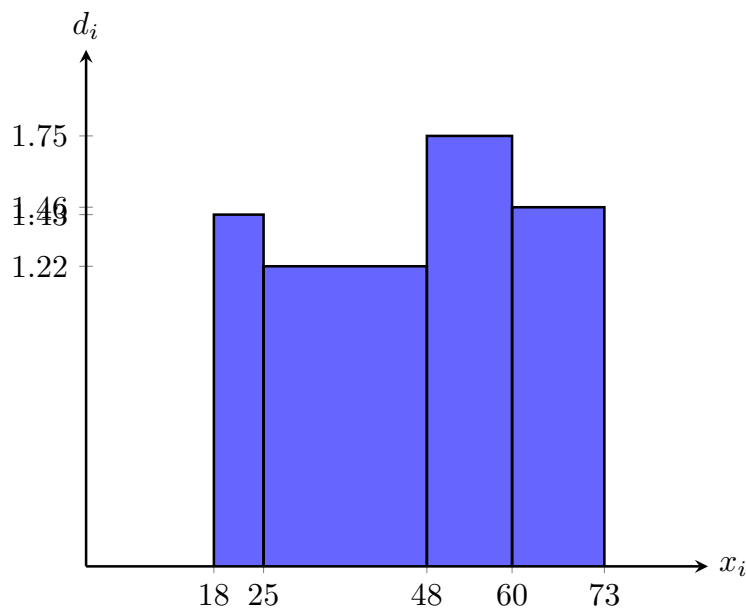
1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[18, 25[$	10	0.13	78	10	1	0.13
$[25, 48[$	28	0.36	68	38	0.87	0.49
$[48, 60[$	21	0.27	40	59	0.51	0.76
$[60, 73[$	19	0.24	19	78	0.24	1.0
$\Sigma$	78	1	–	–	–	–

2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités d'effectifs  $d_i$  en divisant les effectifs  $n_i$  des classes par leurs amplitudes respectives  $a_i$  :

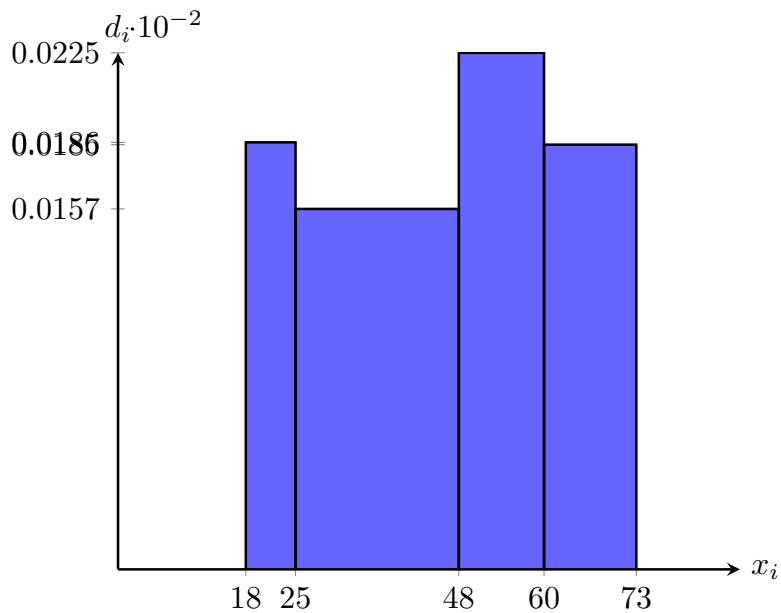
$[x_i^-, x_i^+[$	$[18, 25[$	$[25, 48[$	$[48, 60[$	$[60, 73[$
$n_i$	10	28	21	19
$a_i$	7	23	12	13
$d_i$	1.43	1.22	1.75	1.46



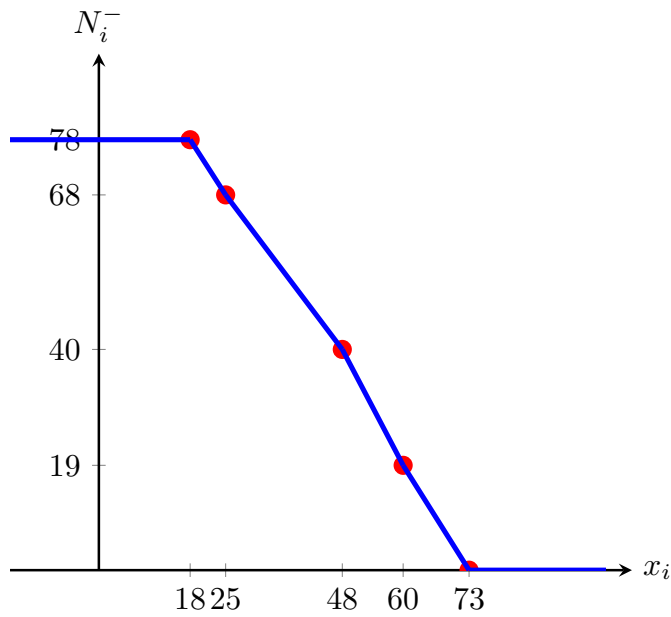
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités de fréquences  $d_i$  en divisant les fréquences  $f_i$  des classes par leurs amplitudes respectives  $a_i$  :

$[x_i^-, x_i^+[$	$[18, 25[$	$[25, 48[$	$[48, 60[$	$[60, 73[$
$f_i$	0.13	0.36	0.27	0.24
$a_i$	7	23	12	13
$d_i$	0.0186	0.0157	0.0225	0.0185

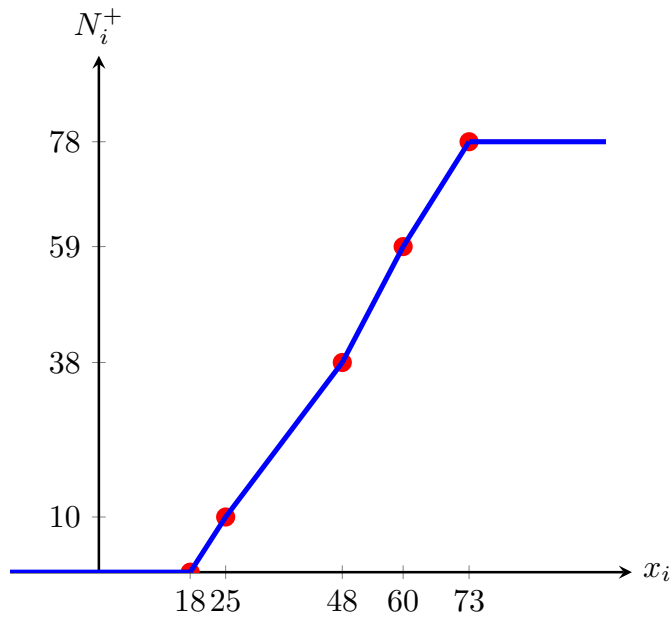
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



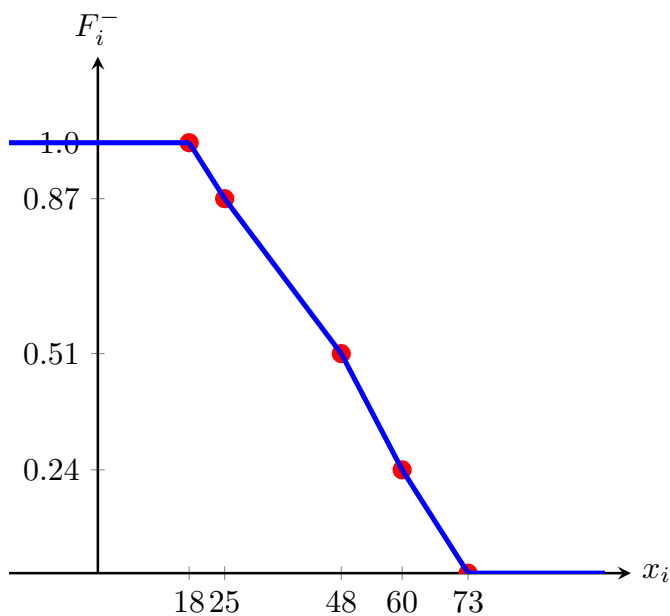
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



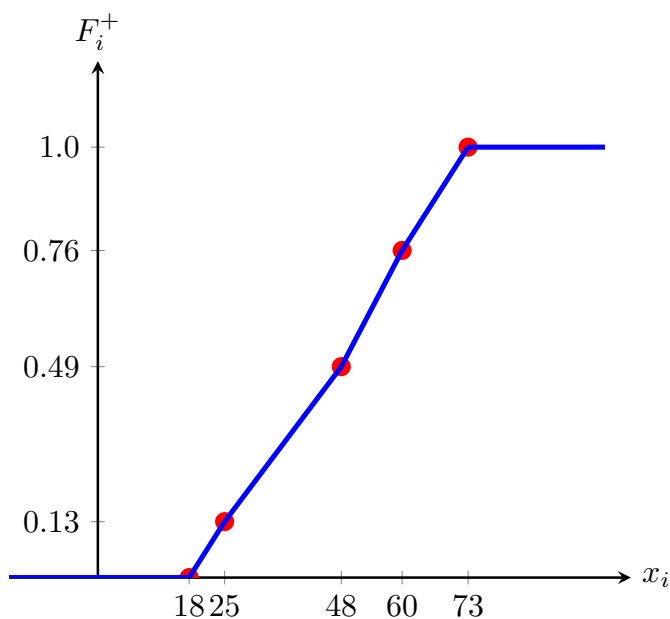
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 21 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+ [$	$[16, 31[$	$[31, 41[$	$[41, 58[$	$[58, 73[$	$[73, 87[$
$n_i$	13	27	34	22	17

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$ .

**Solution :**

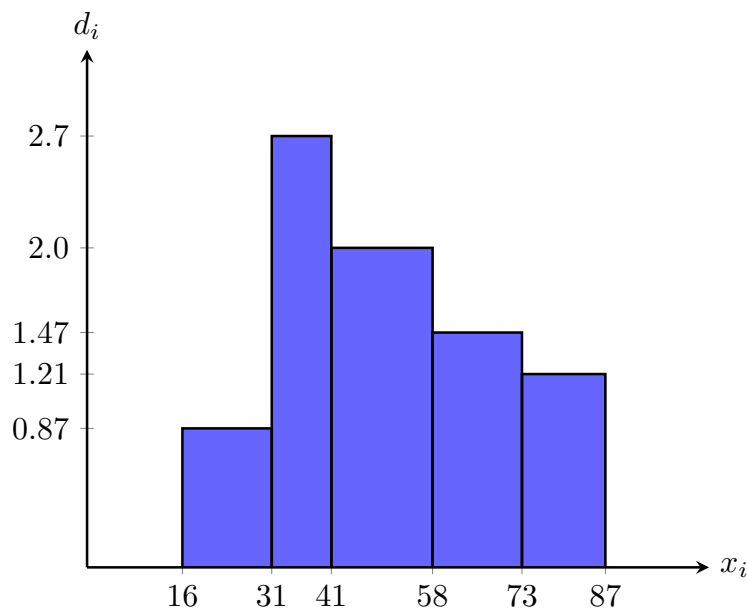
1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[16, 31[$	13	0.12	113	13	1	0.12
$[31, 41[$	27	0.24	100	40	0.88	0.36
$[41, 58[$	34	0.3	73	74	0.64	0.66
$[58, 73[$	22	0.19	39	96	0.34	0.85
$[73, 87[$	17	0.15	17	113	0.15	1
$\Sigma$	113	1	-	-	-	-

2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités d'effectifs  $d_i$  en divisant les effectifs  $n_i$  des classes par leurs amplitudes respectives  $a_i$  :

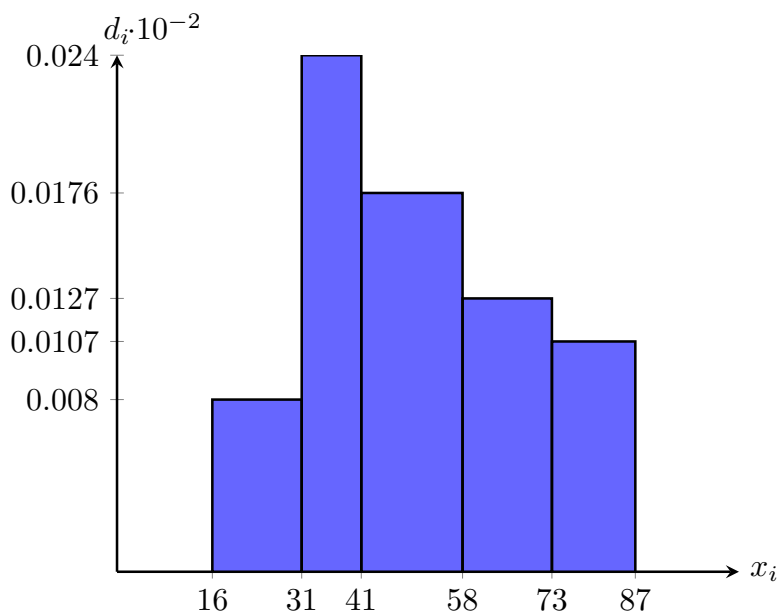
$[x_i^-, x_i^+[$	$[16, 31[$	$[31, 41[$	$[41, 58[$	$[58, 73[$	$[73, 87[$
$n_i$	13	27	34	22	17
$a_i$	15	10	17	15	14
$d_i$	0.87	2.7	2.0	1.47	1.21



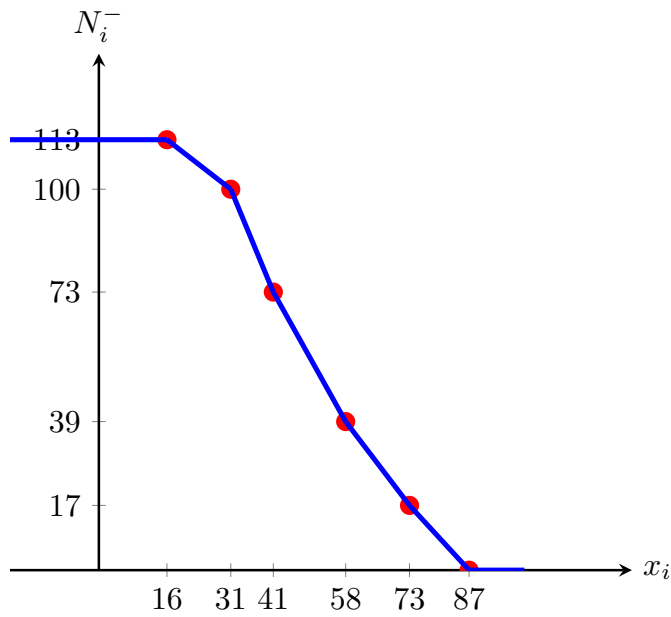
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités de fréquences  $d_i$  en divisant les fréquences  $f_i$  des classes par leurs amplitudes respectives  $a_i$  :

$[x_i^-, x_i^+]$	$[16, 31[$	$[31, 41[$	$[41, 58[$	$[58, 73[$	$[73, 87[$
$f_i$	0.12	0.24	0.3	0.19	0.15
$a_i$	15	10	17	15	14
$d_i$	0.008	0.024	0.0176	0.0127	0.0107

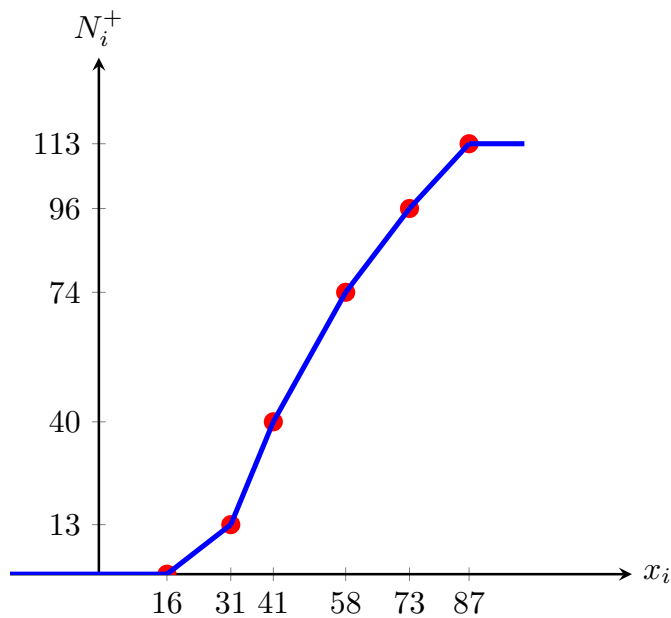
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



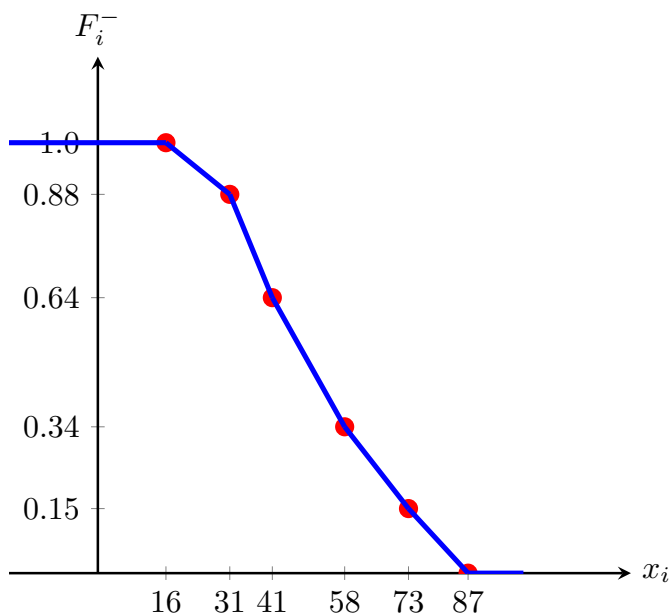
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



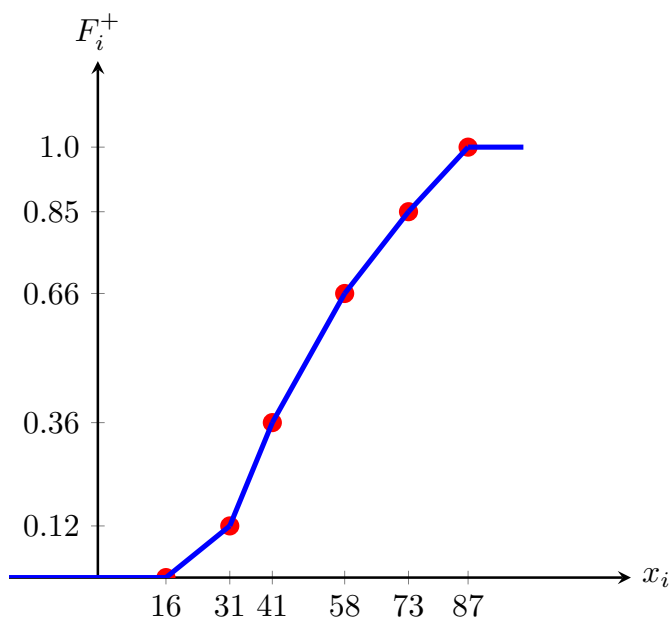
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 22 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+]$	$[13, 25[$	$[25, 45[$	$[45, 56[$
$n_i$	12	21	17

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$ .

**Solution :**

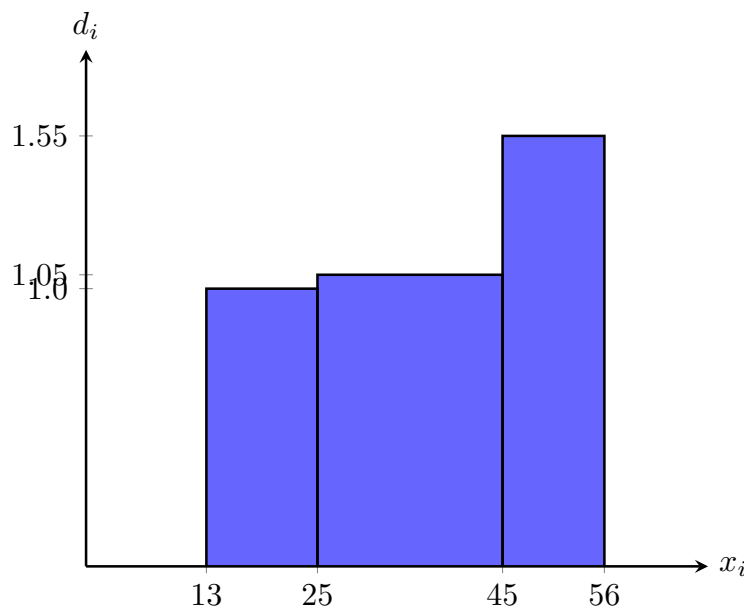
1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[13, 25[$	12	0.24	50	12	1	0.24
$[25, 45[$	21	0.42	38	33	0.76	0.66
$[45, 56[$	17	0.34	17	50	0.34	1.0
$\Sigma$	50	1	–	–	–	–

2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités d'effectifs  $d_i$  en divisant les effectifs  $n_i$  des classes par leurs amplitudes respectives  $a_i$  :

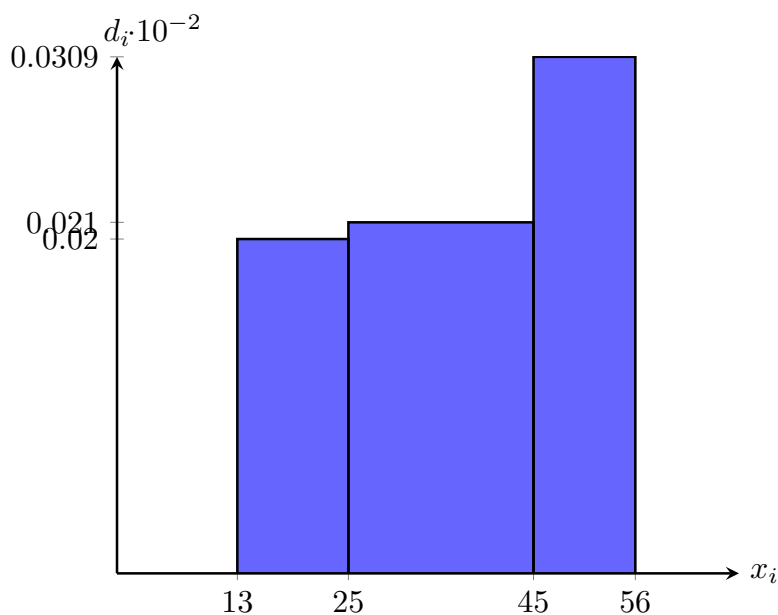
$[x_i^-, x_i^+[$	$[13, 25[$	$[25, 45[$	$[45, 56[$
$n_i$	12	21	17
$a_i$	12	20	11
$d_i$	1.0	1.05	1.55



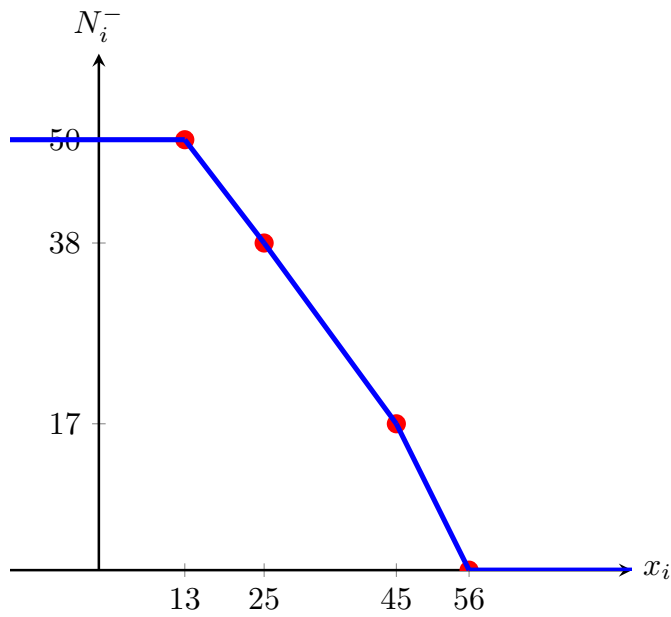
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités de fréquences  $d_i$  en divisant les fréquences  $f_i$  des classes par leurs amplitudes respectives  $a_i$  :

$[x_i^-, x_i^+[$	$[13, 25[$	$[25, 45[$	$[45, 56[$
$f_i$	0.24	0.42	0.34
$a_i$	12	20	11
$d_i$	0.02	0.021	0.0309

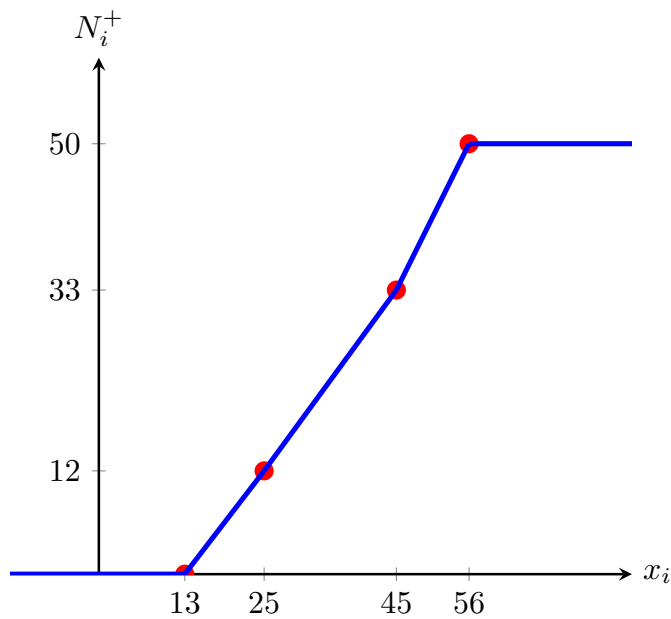
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



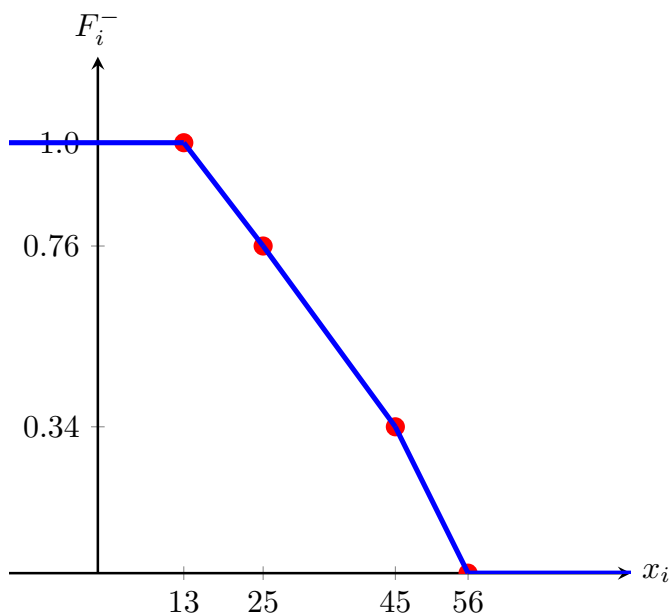
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



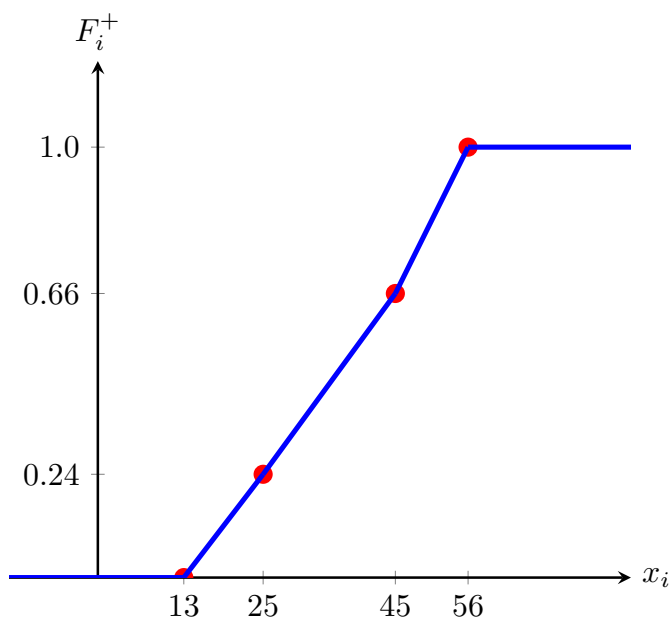
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 23 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+]$	$[16, 28[$	$[28, 49[$	$[49, 55[$	$[55, 74[$
$n_i$	12	28	23	19

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$  .

**Solution :**

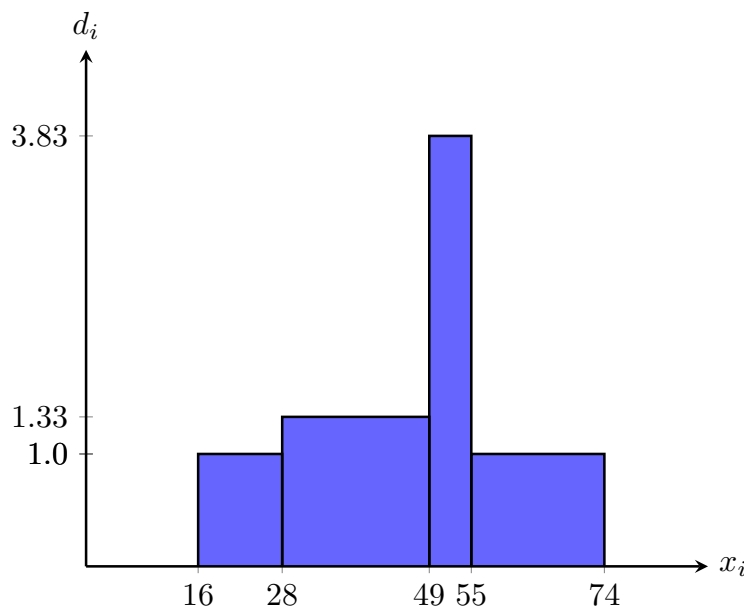
1. Tableau statistique :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[16, 28[$	12	0.15	82	12	1	0.15
$[28, 49[$	28	0.34	70	40	0.85	0.49
$[49, 55[$	23	0.28	42	63	0.51	0.77
$[55, 74[$	19	0.23	19	82	0.23	1.0
$\Sigma$	82	1	–	–	–	–

2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités d'effectifs  $d_i$  en divisant les effectifs  $n_i$  des classes par leurs amplitudes respectives  $a_i$  :

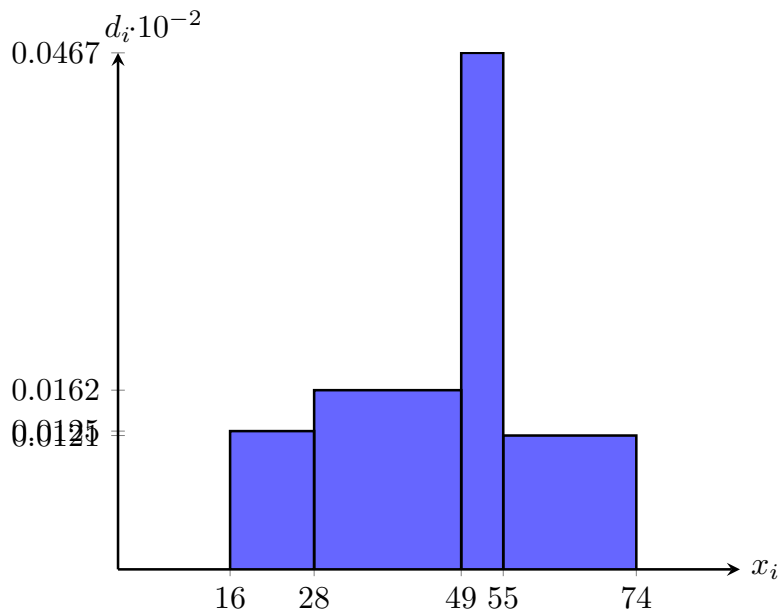
$[x_i^-, x_i^+[$	$[16, 28[$	$[28, 49[$	$[49, 55[$	$[55, 74[$
$n_i$	12	28	23	19
$a_i$	12	21	6	19
$d_i$	1.0	1.33	3.83	1.0



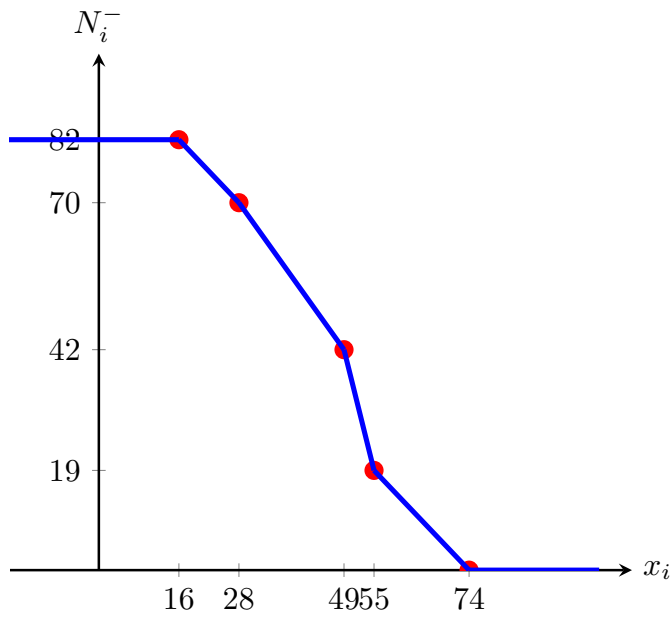
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités de fréquences  $d_i$  en divisant les fréquences  $f_i$  des classes par leurs amplitudes respectives  $a_i$  :

$[x_i^-, x_i^+[$	$[16, 28[$	$[28, 49[$	$[49, 55[$	$[55, 74[$
$f_i$	0.15	0.34	0.28	0.23
$a_i$	12	21	6	19
$d_i$	0.0125	0.0162	0.0467	0.0121

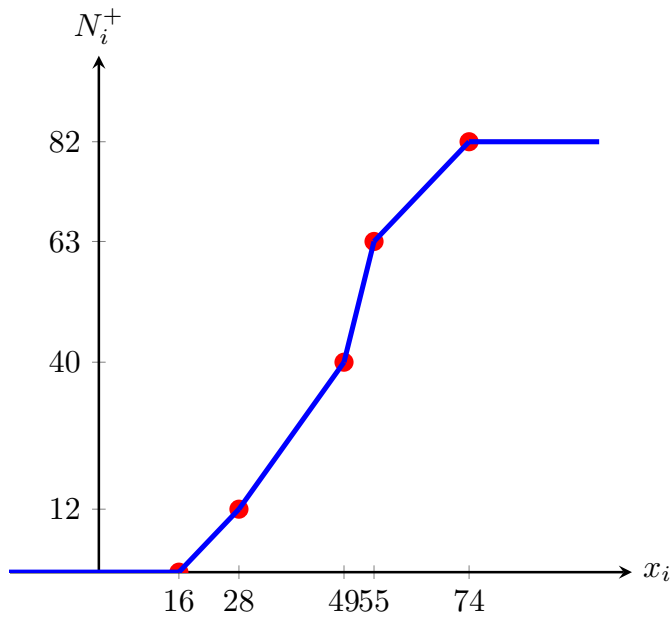
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



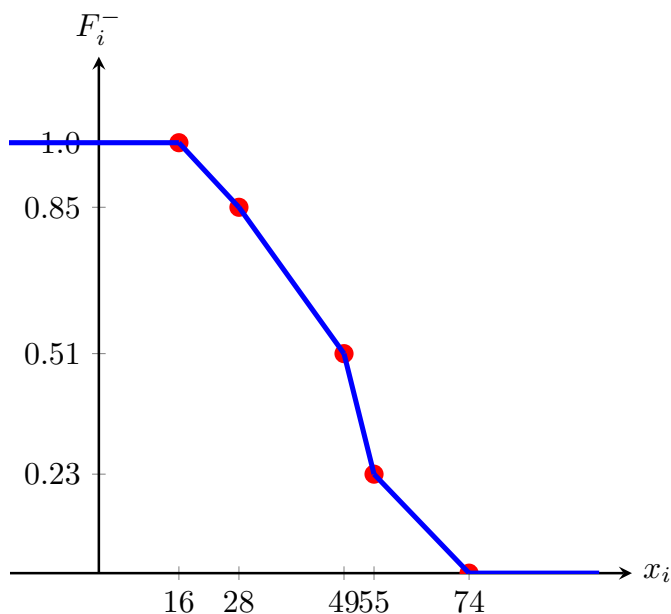
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



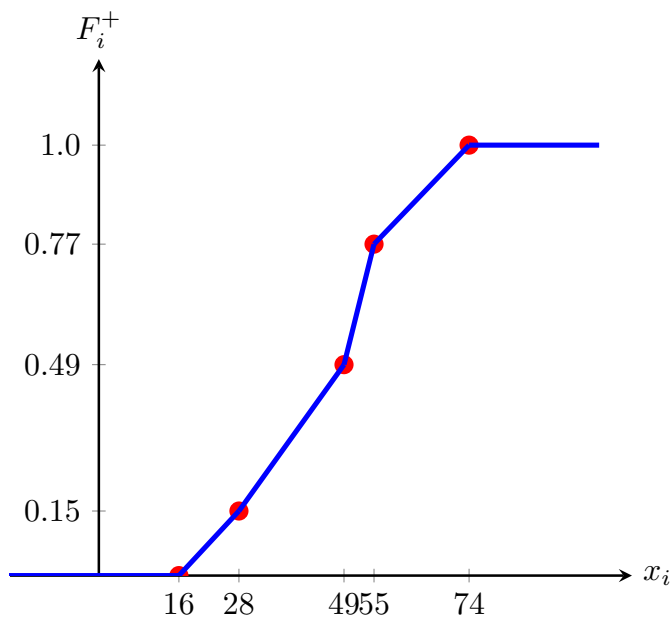
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



**Exercice 24 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  :

$[x_i^-, x_i^+[$	$[16, 27[$	$[27, 46[$	$[46, 62[$	$[62, 75[$	$[75, 85[$
$n_i$	11	26	30	20	17

1. Construire le tableau statistique de la variable  $X$  ;
2. Représenter graphiquement les effectifs  $n_i$  ;
3. Représenter graphiquement les fréquences  $f_i$  ;
4. Représenter graphiquement les effectifs cumulés décroissants  $N_i^-$  ;
5. Représenter graphiquement les effectifs cumulés croissants  $N_i^+$  ;
6. Représenter graphiquement les fréquences cumulées décroissantes  $F_i^-$  ;
7. Représenter graphiquement les fréquences cumulées croissantes  $F_i^+$ .

**Solution :**

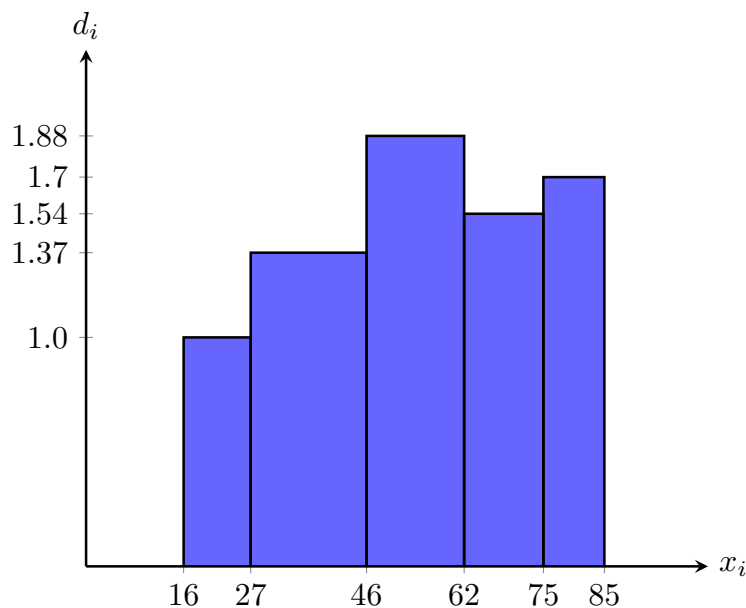
1. Tableau statistique :

$[x_i^-, x_i^+ [$	$n_i$	$f_i$	$N_i^-$	$N_i^+$	$F_i^-$	$F_i^+$
$[16, 27 [$	11	0.11	104	11	1	0.11
$[27, 46 [$	26	0.25	93	37	0.89	0.36
$[46, 62 [$	30	0.29	67	67	0.64	0.65
$[62, 75 [$	20	0.19	37	87	0.35	0.84
$[75, 85 [$	17	0.16	17	104	0.16	1
$\Sigma$	104	1	–	–	–	–

2. Représentation graphique des effectifs  $n_i$  :

Les effectifs  $n_i$  sont représentés graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités d'effectifs  $d_i$  en divisant les effectifs  $n_i$  des classes par leurs amplitudes respectives  $a_i$  :

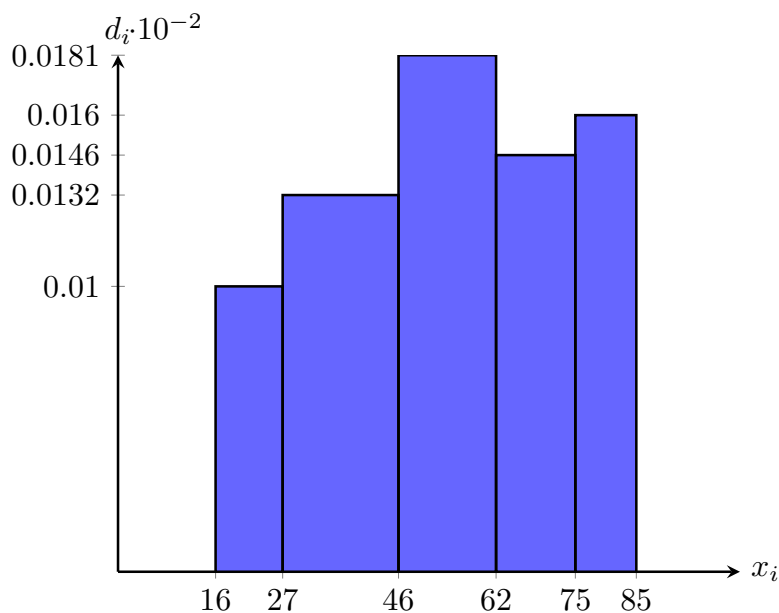
$[x_i^-, x_i^+ [$	$[16, 27 [$	$[27, 46 [$	$[46, 62 [$	$[62, 75 [$	$[75, 85 [$
$n_i$	11	26	30	20	17
$a_i$	11	19	16	13	10
$d_i$	1.0	1.37	1.88	1.54	1.7



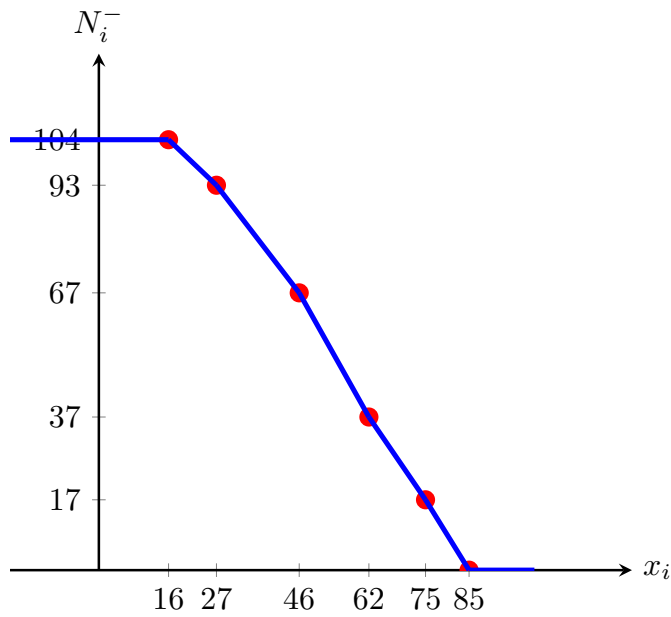
3. Représentation graphique des fréquences  $f_i$  :

Les fréquences  $f_i$  sont représentées graphiquement à l'aide d'un histogramme. Pour construire ce dernier, il convient d'abord de calculer les densités de fréquences  $d_i$  en divisant les fréquences  $f_i$  des classes par leurs amplitudes respectives  $a_i$  :

$[x_i^-, x_i^+]$	$[16, 27[$	$[27, 46[$	$[46, 62[$	$[62, 75[$	$[75, 85[$
$f_i$	0.11	0.25	0.29	0.19	0.16
$a_i$	11	19	16	13	10
$d_i$	0.01	0.0132	0.0181	0.0146	0.016

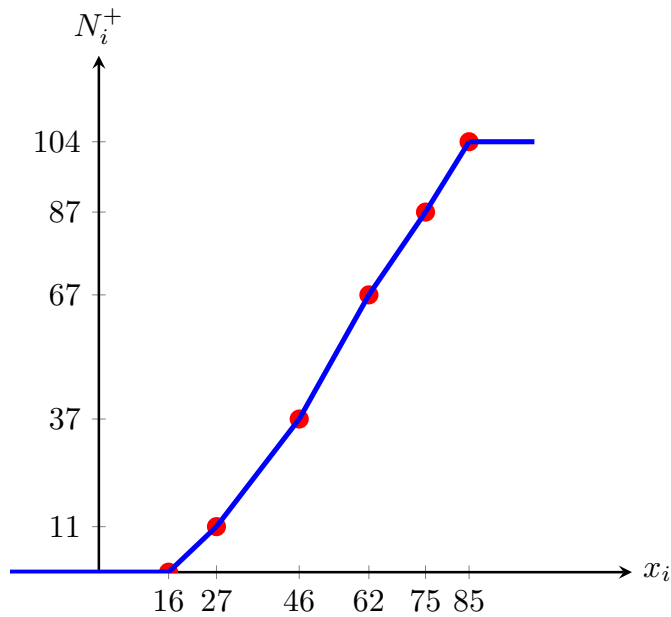
4. Représentation graphique des effectifs cumulés décroissants  $N_i^-$  :

Les effectifs cumulés décroissants  $N_i^-$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



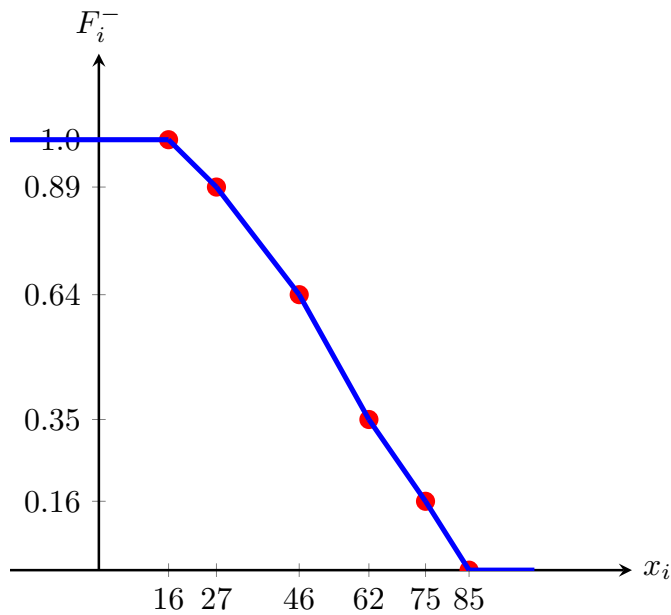
5. Représentation graphique des effectifs cumulés croissants  $N_i^+$  :

Les effectifs cumulés croissants  $N_i^+$  sont représentés graphiquement à l'aide de la courbe cumulative suivante :



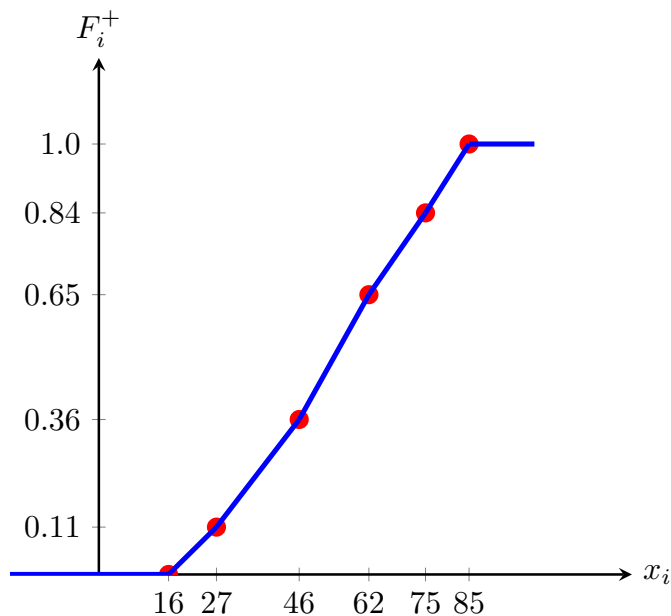
6. Représentation graphique des fréquences cumulées décroissantes  $F_i^-$  :

Les fréquences cumulées décroissantes  $F_i^-$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



7. Représentation graphique des fréquences cumulées croissantes  $F_i^+$  :

Les fréquences cumulées croissantes  $F_i^+$  sont représentées graphiquement à l'aide de la courbe cumulative suivante :



## Conclusion

Dans ce chapitre introductif à la statistique ont été résolus des exercices relatifs au tableau statistique et aux représentations graphiques. Dans le chapitre suivant, seront résolus des exercices sur des mesures numériques dans le cas d'une seule variable statistique quantitative.



# Chapitre 2

## Statistique descriptive univariée

Dans ce deuxième chapitre, on s'intéressera à des mesures numériques basées sur les moments (section 2.1) et à des mesures numériques basées sur les quantiles (section 2.2) dans le cas d'une seule variable statistique quantitative.

### 2.1 Statistique descriptive basée sur les moments

Dans cette section, seront résolus des exercices sur des mesures numériques basées sur les moments, telles que les mesures de tendance centrale, les mesures de dispersion et les mesures de forme, dans le cas d'une variable quantitative discrète (sous-section 2.1.1) et dans le cas d'une variable quantitative continue (sous-section 2.1.2).

#### 2.1.1 Variable quantitative discrète

Une variable quantitative discrète  $X$  est une variable statistique qui prend en ensemble fini et dénombrable de valeurs numériques notées  $x_i$ .

##### Exercice 25 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en jours :

$x_i$	5	14	36	$n$
$n_i$	1	26	20	47

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

##### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$x_i$	$n_i$	$n_i x_i$	$n_i  x_i - \bar{x} $	$n_i x_i^2$	$n_i (x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^3$	$n_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3$	$n_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4$
5	1	5	18.17	25	330.15	-5998.81	-4.36	7.13
14	26	364	238.42	5096	2186.31	-20048.48	-14.58	12.02
36	20	720	256.6	25920	3292.18	42238.64	30.72	35.44
$\Sigma$	47	1089	513.19	31041	5808.64	16191.35	11.78	54.59

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^3 n_i x_i \\ &= \frac{1}{n} (n_1 x_1 + n_2 x_2 + n_3 x_3) \\ &= \frac{1}{47} (1 \times 5 + 26 \times 14 + 20 \times 36) \\ &= \frac{1089}{47} \\ &= 23.17 \text{ jours}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 23.17 jours. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par son écart absolu moyen, par sa variance, par son écart-type ou par son coefficient de variation :

- L'écart absolu moyen  $EAM$  de la variable statistique  $X$  est donné par :

$$\begin{aligned}EAM &= \frac{1}{n} \sum_{i=1}^3 n_i |x_i - \bar{x}| \\ &= \frac{1}{n} [n_1 |x_1 - \bar{x}| + n_2 |x_2 - \bar{x}| + n_3 |x_3 - \bar{x}|] \\ &= \frac{1}{47} [1 |5 - 23.17| + 26 |14 - 23.17| + 20 |36 - 23.17|] \\ &= \frac{513.19}{47} \\ &= 10.92 \text{ jours}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 23.17$  jours et s'en écartent, en termes absolus, de  $EAM = 10.92$  jours en moyenne.

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \frac{1}{n} \sum_{i=1}^3 n_i (x_i - \bar{x})^2 \\ &= \frac{1}{n} [n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + n_3 (x_3 - \bar{x})^2] \\ &= \frac{1}{47} [1 (5 - 23.17)^2 + 26 (14 - 23.17)^2 + 20 (36 - 23.17)^2] \\ &= \frac{5808.64}{47} \\ &= 123.59 \text{ "jours}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens

comme suit :

$$\begin{aligned}
 \sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\
 &= \left( \frac{1}{n} \sum_{i=1}^3 n_i x_i^2 \right) - \bar{x}^2 \\
 &= \frac{1}{n} (n_1 x_1^2 + n_2 x_2^2 + n_3 x_3^2) - \bar{x}^2 \\
 &= \frac{1}{47} (1 \times 5^2 + 26 \times 14^2 + 20 \times 36^2) - 23.17^2 \\
 &= \frac{31041}{47} - 536.85 \\
 &= 123.6 \text{ "jours"}^2
 \end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}
 \sigma_x &= \sqrt{\sigma_x^2} \\
 &= \sqrt{123.6 \text{ "jours"}^2} \\
 &= 11.12 \text{ jours}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 23.17$  jours et s'en écartent de  $\sigma_x = 11.12$  jours en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}
 CV &= \frac{\sigma_x}{\bar{x}} \\
 &= \frac{11.12 \text{ jours}}{23.17 \text{ jours}} \\
 &= 0.48
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 23.17$  jours et s'en écartent, en termes relatifs, de  $CV = 0.48$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}
 \gamma_1 &= \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \\
 &= \frac{1}{n} \left[ n_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^3 + n_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^3 + n_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^3 \right] \\
 &= \frac{1}{47} \left[ 1 \left( \frac{5 - 23.17}{11.12} \right)^3 + 26 \left( \frac{14 - 23.17}{11.12} \right)^3 + 20 \left( \frac{36 - 23.17}{11.12} \right)^3 \right] \\
 &= \frac{11.78}{47} \\
 &= 0.25
 \end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures

à la moyenne arithmétique ( $\bar{x} = 23.17$  jours) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0625\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \frac{1}{n} \sum_{i=1}^3 n_i (x_i - \bar{x})^3 \\ &= \frac{1}{n} \left[ n_1 (x_1 - \bar{x})^3 + n_2 (x_2 - \bar{x})^3 + n_3 (x_3 - \bar{x})^3 \right] \\ &= \frac{1}{47} \left[ 1 (5 - 23.17)^3 + 26 (14 - 23.17)^3 + 20 (36 - 23.17)^3 \right] \\ &= \frac{16191.35}{47} \\ &= 344.5\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 23.17$  jours) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^4 + n_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^4 + n_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^4 \right] \\ &= \frac{1}{47} \left[ 1 \left( \frac{5 - 23.17}{11.12} \right)^4 + 26 \left( \frac{14 - 23.17}{11.12} \right)^4 + 20 \left( \frac{36 - 23.17}{11.12} \right)^4 \right] \\ &= \frac{54.59}{47} \\ &= 1.16\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -1.84\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### Exercice 26 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en dollars :

$x_i$	3	13	31	44	$n$
$n_i$	6	33	27	17	83

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$x_i$	$n_i$	$n_i x_i$	$n_i  x_i - \bar{x} $	$n_i x_i^2$	$n_i (x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^3$	$n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3$	$n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4$
3	6	18	128.88	54	2768.34	-59463.99	-25.11	40.45
13	33	429	378.84	5577	4349.08	-49927.48	-21.08	18.15
31	27	837	176.04	25947	1147.78	7483.53	3.16	1.55
44	17	748	331.84	32912	6477.52	126441.13	53.38	78.17
$\Sigma$	83	2032	1015.6	64490	14742.72	24533.19	10.35	138.32

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^4 n_i x_i \\ &= \frac{1}{n} (n_1 x_1 + n_2 x_2 + n_3 x_3 + n_4 x_4) \\ &= \frac{1}{83} (6 \times 3 + 33 \times 13 + 27 \times 31 + 17 \times 44) \\ &= \frac{2032}{83} \\ &= 24.48 \text{ dollars}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 24.48 dollars. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par son écart absolu moyen, par sa variance, par son écart-type ou par son coefficient de variation :

- L'écart absolu moyen  $EAM$  de la variable statistique  $X$  est donné par :

$$\begin{aligned}
 EAM &= \frac{1}{n} \sum_{i=1}^4 n_i |x_i - \bar{x}| \\
 &= \frac{1}{n} [n_1 |x_1 - \bar{x}| + n_2 |x_2 - \bar{x}| + n_3 |x_3 - \bar{x}| + n_4 |x_4 - \bar{x}|] \\
 &= \frac{1}{83} [6 |3 - 24.48| + 33 |13 - 24.48| + 27 |31 - 24.48| + 17 |44 - 24.48|] \\
 &= \frac{1015.6}{83} \\
 &= 12.24 \text{ dollars}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 24.48$  dollars et s'en écartent, en termes absolus, de  $EAM = 12.24$  dollars en moyenne.

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \sigma_x^2 &= \frac{1}{n} \sum_{i=1}^4 n_i (x_i - \bar{x})^2 \\
 &= \frac{1}{n} [n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + n_3 (x_3 - \bar{x})^2 + n_4 (x_4 - \bar{x})^2] \\
 &= \frac{1}{83} [6 (3 - 24.48)^2 + 33 (13 - 24.48)^2 + 27 (31 - 24.48)^2 + 17 (44 - 24.48)^2] \\
 &= \frac{14742.72}{83} \\
 &= 177.62 \text{ "dollars"}^2
 \end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}
 \sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\
 &= \left( \frac{1}{n} \sum_{i=1}^4 n_i x_i^2 \right) - \bar{x}^2 \\
 &= \frac{1}{n} (n_1 x_1^2 + n_2 x_2^2 + n_3 x_3^2 + n_4 x_4^2) - \bar{x}^2 \\
 &= \frac{1}{83} (6 \times 3^2 + 33 \times 13^2 + 27 \times 31^2 + 17 \times 44^2) - 24.48^2 \\
 &= \frac{64490}{83} - 599.27 \\
 &= 177.72 \text{ "dollars"}^2
 \end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}
 \sigma_x &= \sqrt{\sigma_x^2} \\
 &= \sqrt{177.72 \text{ "dollars"}^2} \\
 &= 13.33 \text{ dollars}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 24.48$  dollars et s'en écartent de  $\sigma_x = 13.33$  dollars en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned} CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{13.33 \text{ dollars}}{24.48 \text{ dollars}} \\ &= 0.54 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 24.48$  dollars et s'en écartent, en termes relatifs, de  $CV = 0.54$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned} \gamma_1 &= \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^3 + n_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^3 + n_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^3 + n_4 \left( \frac{x_4 - \bar{x}}{\sigma_x} \right)^3 \right] \\ &= \frac{10.35}{83} \\ &= 0.12 \end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 24.48$  dollars) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned} \beta_1 &= \left[ \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0144 \end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned} \mu_3 &= \frac{1}{n} \sum_{i=1}^4 n_i (x_i - \bar{x})^3 \\ &= \frac{1}{n} \left[ n_1 (x_1 - \bar{x})^3 + n_2 (x_2 - \bar{x})^3 + n_3 (x_3 - \bar{x})^3 + n_4 (x_4 - \bar{x})^3 \right] \\ &= \frac{1}{83} \left[ 6(3 - 24.48)^3 + 33(13 - 24.48)^3 + 27(31 - 24.48)^3 + 17(44 - 24.48)^3 \right] \\ &= \frac{24533.19}{83} \\ &= 295.58 \end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 24.48$  dollars) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^4 + n_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^4 + n_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^4 + n_4 \left( \frac{x_4 - \bar{x}}{\sigma_x} \right)^4 \right] \\ &= \frac{138.32}{83} \\ &= 1.67\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -1.33\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### Exercice 27 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en semaines :

$x_i$	4	18	29	41	54	$n$
$n_i$	9	39	32	28	12	120

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$x_i$	$n_i$	$n_i x_i$	$n_i  x_i - \bar{x} $	$n_i x_i^2$	$n_i (x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^3$	$n_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3$	$n_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4$
4	9	36	223.65	144	5557.7	-138108.91	-56.13	103.33
18	39	702	423.15	12636	4591.18	-49814.28	-20.25	16.27
29	32	928	4.8	26912	0.72	0.11	0.0	0.0
41	28	1148	340.2	47068	4133.43	50221.17	20.41	18.37
54	12	648	301.8	34992	7590.27	190895.29	77.59	144.54
$\Sigma$	120	3462	1293.6	121752	21873.3	53193.38	21.62	282.51

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \bar{x} &= \frac{1}{n} \sum_{i=1}^5 n_i x_i \\
 &= \frac{1}{n} (n_1 x_1 + n_2 x_2 + n_3 x_3 + n_4 x_4 + n_5 x_5) \\
 &= \frac{1}{120} (9 \times 4 + 39 \times 18 + 32 \times 29 + 28 \times 41 + 12 \times 54) \\
 &= \frac{3462}{120} \\
 &= 28.85 \text{ semaines}
 \end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 28.85 semaines. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par son écart absolu moyen, par sa variance, par son écart-type ou par son coefficient de variation :

- L'écart absolu moyen  $EAM$  de la variable statistique  $X$  est donné par :

$$\begin{aligned}
 EAM &= \frac{1}{n} \sum_{i=1}^5 n_i |x_i - \bar{x}| \\
 &= \frac{1}{n} [n_1 |x_1 - \bar{x}| + n_2 |x_2 - \bar{x}| + n_3 |x_3 - \bar{x}| + n_4 |x_4 - \bar{x}| + n_5 |x_5 - \bar{x}|] \\
 &= \frac{1293.6}{120} \\
 &= 10.78 \text{ semaines}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 28.85$  semaines et s'en écartent, en termes absolus, de  $EAM = 10.78$  semaines en moyenne.

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \sigma_x^2 &= \frac{1}{n} \sum_{i=1}^5 n_i (x_i - \bar{x})^2 \\
 &= \frac{1}{n} [n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + n_3 (x_3 - \bar{x})^2 + n_4 (x_4 - \bar{x})^2 + n_5 (x_5 - \bar{x})^2] \\
 &= \frac{21873.3}{120} \\
 &= 182.28 \text{ "semaines"}^2
 \end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\ &= \left( \frac{1}{n} \sum_{i=1}^5 n_i x_i^2 \right) - \bar{x}^2 \\ &= \frac{1}{n} (n_1 x_1^2 + n_2 x_2^2 + n_3 x_3^2 + n_4 x_4^2 + n_5 x_5^2) - \bar{x}^2 \\ &= \frac{1}{120} (9 \times 4^2 + 39 \times 18^2 + 32 \times 29^2 + 28 \times 41^2 + 12 \times 54^2) - 28.85^2 \\ &= \frac{121752}{120} - 832.32 \\ &= 182.28 \text{ "semaines"}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{182.28 \text{ "semaines"}^2} \\ &= 13.5 \text{ semaines}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 28.85$  semaines et s'en écartent de  $\sigma_x = 13.5$  semaines en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{13.5 \text{ semaines}}{28.85 \text{ semaines}} \\ &= 0.47\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 28.85$  semaines et s'en écartent, en termes relatifs, de  $CV = 0.47$  en moyenne.

### 3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \\ &= \frac{21.62}{120} \\ &= 0.18\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 28.85$  semaines) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0324\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \frac{1}{n} \sum_{i=1}^5 n_i (x_i - \bar{x})^3 \\ &= \frac{1}{n} \left[ n_1 (x_1 - \bar{x})^3 + n_2 (x_2 - \bar{x})^3 + n_3 (x_3 - \bar{x})^3 + n_4 (x_4 - \bar{x})^3 + n_5 (x_5 - \bar{x})^3 \right] \\ &= \frac{53193.38}{120} \\ &= 443.28\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 28.85$  semaines) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 \\ &= \frac{282.51}{120} \\ &= 2.35\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -0.65\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

**Exercice 28 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en gigaoctets :

$x_i$	1	17	35
$f_i$	0.15	0.7	0.15

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

**Solution :**

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$x_i$	$f_i$	$f_i x_i$	$f_i  x_i - \bar{x} $	$f_i x_i^2$	$f_i (x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^3$	$f_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3$	$f_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4$
1	0.15	0.15	2.44	0.15	39.85	-649.61	-0.8	1.4
17	0.7	11.9	0.21	202.3	0.06	-0.02	-0.0	0.0
35	0.15	5.25	2.66	183.75	46.99	831.78	1.03	1.95
$\Sigma$	1	17.3	5.31	386.2	86.9	182.15	0.23	3.35

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \bar{x} &= \sum_{i=1}^3 f_i x_i \\
 &= f_1 x_1 + f_2 x_2 + f_3 x_3 \\
 &= 0.15 \times 1 + 0.7 \times 17 + 0.15 \times 35 \\
 &= 17.3 \text{ gigaoctets}
 \end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 17.3 gigaoctets. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par son écart absolu moyen, par sa variance, par son écart-type ou par son coefficient de variation :

- L'écart absolu moyen  $EAM$  de la variable statistique  $X$  est donné par :

$$\begin{aligned}
 EAM &= \sum_{i=1}^3 f_i |x_i - \bar{x}| \\
 &= f_1 |x_1 - \bar{x}| + f_2 |x_2 - \bar{x}| + f_3 |x_3 - \bar{x}| \\
 &= 0.15 |1 - 17.3| + 0.7 |17 - 17.3| + 0.15 |35 - 17.3| \\
 &= 5.31 \text{ gigaoctets}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 17.3$  gigaoctets et s'en écartent, en termes absolus, de  $EAM = 5.31$  gigaoctets en moyenne.

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \sum_{i=1}^3 f_i (x_i - \bar{x})^2 \\ &= f_1 (x_1 - \bar{x})^2 + f_2 (x_2 - \bar{x})^2 + f_3 (x_3 - \bar{x})^2 \\ &= 0.15 (1 - 17.3)^2 + 0.7 (17 - 17.3)^2 + 0.15 (35 - 17.3)^2 \\ &= 86.9 \text{ "gigaoctets}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \left( \sum_{i=1}^3 f_i x_i^2 \right) - \bar{x}^2 \\ &= (f_1 x_1^2 + f_2 x_2^2 + f_3 x_3^2) - \bar{x}^2 \\ &= (0.15 \times 1^2 + 0.7 \times 17^2 + 0.15 \times 35^2) - 17.3^2 \\ &= 386.2 - 299.29 \\ &= 86.91 \text{ "gigaoctets}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{86.91 \text{ "gigaoctets}^2} \\ &= 9.32 \text{ gigaoctets}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 17.3$  gigaoctets et s'en écartent de  $\sigma_x = 9.32$  gigaoctets en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{9.32 \text{ gigaoctets}}{17.3 \text{ gigaoctets}} \\ &= 0.54\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 17.3$  gigaoctets et s'en écartent, en termes relatifs, de  $CV = 0.54$  en moyenne.

### 3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \sum_{i=1}^3 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \\ &= f_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^3 + f_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^3 + f_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^3 \\ &= 0.15 \left( \frac{1 - 17.3}{9.32} \right)^3 + 0.7 \left( \frac{17 - 17.3}{9.32} \right)^3 + 0.15 \left( \frac{35 - 17.3}{9.32} \right)^3 \\ &= 0.23\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 17.3$  gigaoctets) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \sum_{i=1}^3 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0529\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \sum_{i=1}^3 f_i (x_i - \bar{x})^3 \\ &= f_1 (x_1 - \bar{x})^3 + f_2 (x_2 - \bar{x})^3 + f_3 (x_3 - \bar{x})^3 \\ &= 0.15 (1 - 17.3)^3 + 0.7 (17 - 17.3)^3 + 0.15 (35 - 17.3)^3 \\ &= 182.15\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 17.3$  gigaoctets) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \sum_{i=1}^3 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 \\ &= f_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^4 + f_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^4 + f_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^4 \\ &= 0.15 \left( \frac{1 - 17.3}{9.32} \right)^4 + 0.7 \left( \frac{17 - 17.3}{9.32} \right)^4 + 0.15 \left( \frac{35 - 17.3}{9.32} \right)^4 \\ &= 3.35\end{aligned}$$

Ce coefficient étant supérieur à 3, la distribution statistique de la variable  $X$  est en conséquence leptokurtique. Cela signifie que les observations extrêmes sont excessivement nombreuses.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \sum_{i=1}^3 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= 0.35\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence leptokurtique. Cela signifie que les observations extrêmes sont excessivement nombreuses.

**Exercice 29 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en grammes :

$x_i$	9	15	23	48
$f_i$	0.12	0.39	0.41	0.08

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

**Solution :**

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$x_i$	$f_i$	$f_i x_i$	$f_i  x_i - \bar{x} $	$f_i x_i^2$	$f_i (x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^3$	$f_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3$	$f_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4$
9	0.12	1.08	1.34	9.72	15.05	-168.59	-0.2	0.23
15	0.39	5.85	2.03	87.75	10.55	-54.84	-0.06	0.03
23	0.41	9.43	1.15	216.89	3.21	9.0	0.01	0.0
48	0.08	3.84	2.22	184.32	61.83	1718.8	1.99	5.82
$\Sigma$	1	20.2	6.74	498.68	90.64	1504.37	1.74	6.08

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \bar{x} &= \sum_{i=1}^4 f_i x_i \\
 &= f_1 x_1 + f_2 x_2 + f_3 x_3 + f_4 x_4 \\
 &= 0.12 \times 9 + 0.39 \times 15 + 0.41 \times 23 + 0.08 \times 48 \\
 &= 20.2 \text{ grammes}
 \end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 20.2 grammes. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par son écart absolu moyen, par sa variance, par son écart-type ou par son coefficient de variation :
  - L'écart absolu moyen  $EAM$  de la variable statistique  $X$  est donné par :

$$\begin{aligned}
 EAM &= \sum_{i=1}^4 f_i |x_i - \bar{x}| \\
 &= f_1 |x_1 - \bar{x}| + f_2 |x_2 - \bar{x}| + f_3 |x_3 - \bar{x}| + f_4 |x_4 - \bar{x}| \\
 &= 0.12 |9 - 20.2| + 0.39 |15 - 20.2| + 0.41 |23 - 20.2| + 0.08 |48 - 20.2| \\
 &= 6.74 \text{ grammes}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 20.2$  grammes et s'en écartent, en termes absolus, de  $EAM = 6.74$  grammes en moyenne.

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \sum_{i=1}^4 f_i (x_i - \bar{x})^2 \\ &= f_1 (x_1 - \bar{x})^2 + f_2 (x_2 - \bar{x})^2 + f_3 (x_3 - \bar{x})^2 + f_4 (x_4 - \bar{x})^2 \\ &= 0.12 (9 - 20.2)^2 + 0.39 (15 - 20.2)^2 + 0.41 (23 - 20.2)^2 + 0.08 (48 - 20.2)^2 \\ &= 90.64 \text{ "grammes}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \left( \sum_{i=1}^4 f_i x_i^2 \right) - \bar{x}^2 \\ &= (f_1 x_1^2 + f_2 x_2^2 + f_3 x_3^2 + f_4 x_4^2) - \bar{x}^2 \\ &= (0.12 \times 9^2 + 0.39 \times 15^2 + 0.41 \times 23^2 + 0.08 \times 48^2) - 20.2^2 \\ &= 498.68 - 408.04 \\ &= 90.64 \text{ "grammes}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{90.64 \text{ "grammes}^2} \\ &= 9.52 \text{ grammes}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 20.2$  grammes et s'en écartent de  $\sigma_x = 9.52$  grammes en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{9.52 \text{ grammes}}{20.2 \text{ grammes}} \\ &= 0.47\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 20.2$  grammes et s'en écartent, en termes relatifs, de  $CV = 0.47$  en moyenne.

### 3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \sum_{i=1}^4 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \\ &= f_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^3 + f_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^3 + f_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^3 + f_4 \left( \frac{x_4 - \bar{x}}{\sigma_x} \right)^3 \\ &= 1.74\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 20.2$  grammes) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \sum_{i=1}^4 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 3.0276\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \sum_{i=1}^4 f_i (x_i - \bar{x})^3 \\ &= f_1 (x_1 - \bar{x})^3 + f_2 (x_2 - \bar{x})^3 + f_3 (x_3 - \bar{x})^3 + f_4 (x_4 - \bar{x})^3 \\ &= 0.12 (9 - 20.2)^3 + 0.39 (15 - 20.2)^3 + 0.41 (23 - 20.2)^3 + 0.08 (48 - 20.2)^3 \\ &= 1504.37\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 20.2$  grammes) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \sum_{i=1}^4 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 \\ &= f_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^4 + f_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^4 + f_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^4 + f_4 \left( \frac{x_4 - \bar{x}}{\sigma_x} \right)^4 \\ &= 6.08\end{aligned}$$

Ce coefficient étant supérieur à 3, la distribution statistique de la variable  $X$  est en conséquence leptokurtique. Cela signifie que les observations extrêmes sont excessivement nombreuses.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \sum_{i=1}^4 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= 3.08\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence leptokurtique. Cela signifie que les observations extrêmes sont excessivement nombreuses.

**Exercice 30 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en semaines :

$x_i$	7	15	21	47	52
$f_i$	0.11	0.32	0.2	0.35	0.02

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

**Solution :**

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$x_i$	$f_i$	$f_i x_i$	$f_i  x_i - \bar{x} $	$f_i x_i^2$	$f_i (x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^3$	$f_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3$	$f_i \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4$
7	0.11	0.77	2.23	5.39	45.15	-914.77	-0.23	0.3
15	0.32	4.8	3.92	72.0	48.1	-589.69	-0.15	0.12
21	0.2	4.2	1.25	88.2	7.84	-49.06	-0.01	0.0
47	0.35	16.45	6.91	773.15	136.38	2692.21	0.68	0.85
52	0.02	1.04	0.49	54.08	12.24	302.85	0.08	0.12
$\Sigma$	1	27.26	14.8	992.82	249.71	1441.54	0.37	1.39

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \bar{x} &= \sum_{i=1}^5 f_i x_i \\
 &= f_1 x_1 + f_2 x_2 + f_3 x_3 + f_4 x_4 + f_5 x_5 \\
 &= 0.11 \times 7 + 0.32 \times 15 + 0.2 \times 21 + 0.35 \times 47 + 0.02 \times 52 \\
 &= 27.26 \text{ semaines}
 \end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 27.26 semaines. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par son écart absolu moyen, par sa variance, par son écart-type ou par son coefficient de variation :
  - L'écart absolu moyen  $EAM$  de la variable statistique  $X$  est donné par :

$$\begin{aligned}
 EAM &= \sum_{i=1}^5 f_i |x_i - \bar{x}| \\
 &= f_1 |x_1 - \bar{x}| + f_2 |x_2 - \bar{x}| + f_3 |x_3 - \bar{x}| + f_4 |x_4 - \bar{x}| + f_5 |x_5 - \bar{x}| \\
 &= 14.8 \text{ semaines}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 27.26$  semaines et s'en écartent, en termes absolus, de  $EAM = 14.8$  semaines en moyenne.

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \sum_{i=1}^5 f_i (x_i - \bar{x})^2 \\ &= f_1 (x_1 - \bar{x})^2 + f_2 (x_2 - \bar{x})^2 + f_3 (x_3 - \bar{x})^2 + f_4 (x_4 - \bar{x})^2 + f_5 (x_5 - \bar{x})^2 \\ &= 249.71 \text{ "semaines"}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \left( \sum_{i=1}^5 f_i x_i^2 \right) - \bar{x}^2 \\ &= (f_1 x_1^2 + f_2 x_2^2 + f_3 x_3^2 + f_4 x_4^2 + f_5 x_5^2) - \bar{x}^2 \\ &= (0.11 \times 7^2 + 0.32 \times 15^2 + 0.2 \times 21^2 + 0.35 \times 47^2 + 0.02 \times 52^2) - 27.26^2 \\ &= 992.82 - 743.11 \\ &= 249.71 \text{ "semaines"}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{249.71 \text{ "semaines"}^2} \\ &= 15.8 \text{ semaines}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 27.26$  semaines et s'en écartent de  $\sigma_x = 15.8$  semaines en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{15.8 \text{ semaines}}{27.26 \text{ semaines}} \\ &= 0.58\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 27.26$  semaines et s'en écartent, en termes relatifs, de  $CV = 0.58$  en moyenne.

### 3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \sum_{i=1}^5 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \\ &= f_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^3 + f_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^3 + f_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^3 + f_4 \left( \frac{x_4 - \bar{x}}{\sigma_x} \right)^3 + f_5 \left( \frac{x_5 - \bar{x}}{\sigma_x} \right)^3 \\ &= 0.37\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 27.26$  semaines) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \sum_{i=1}^5 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.1369\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \sum_{i=1}^5 f_i (x_i - \bar{x})^3 \\ &= f_1 (x_1 - \bar{x})^3 + f_2 (x_2 - \bar{x})^3 + f_3 (x_3 - \bar{x})^3 + f_4 (x_4 - \bar{x})^3 + f_5 (x_5 - \bar{x})^3 \\ &= 1441.54\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 27.26$  semaines) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \sum_{i=1}^5 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 \\ &= f_1 \left( \frac{x_1 - \bar{x}}{\sigma_x} \right)^4 + f_2 \left( \frac{x_2 - \bar{x}}{\sigma_x} \right)^4 + f_3 \left( \frac{x_3 - \bar{x}}{\sigma_x} \right)^4 + f_4 \left( \frac{x_4 - \bar{x}}{\sigma_x} \right)^4 + f_5 \left( \frac{x_5 - \bar{x}}{\sigma_x} \right)^4 \\ &= 1.39\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \sum_{i=1}^5 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -1.61\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### 2.1.2 Variable quantitative continue

Une variable quantitative continue  $X$  est une variable statistique qui prend en ensemble infini de valeurs numériques regroupées dans des classes notées  $[x_i^-, x_i^+]$ .

#### Exercice 31 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en années :

$[x_i^-, x_i^+]$	$[6, 11[$	$[11, 30[$	$[30, 49[$	$n$
$n_i$	3	25	7	35

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

#### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$[x_i^-, x_i^+]$	$c_i$	$n_i$	$n_i c_i$	$n_i c_i^2$	$n_i (c_i - \bar{x})^2$	$n_i (c_i - \bar{x})^3$	$n_i \left(\frac{c_i - \bar{x}}{\sigma_x}\right)^3$	$n_i \left(\frac{c_i - \bar{x}}{\sigma_x}\right)^4$
$[6, 11[$	8.5	3	25.5	216.75	654.46	-9666.35	-14.33	24.13
$[11, 30[$	20.5	25	512.5	10506.25	191.82	-531.35	-0.79	0.25
$[30, 49[$	39.5	7	276.5	10921.75	1843.89	29926.34	44.37	82.11
$\Sigma$	-	35	814.5	21644.75	2690.17	19728.64	29.25	106.49

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \bar{x} &= \frac{1}{n} \sum_{i=1}^3 n_i c_i \\
 &= \frac{1}{n} (n_1 c_1 + n_2 c_2 + n_3 c_3) \\
 &= \frac{1}{35} (3 \times 8.5 + 25 \times 20.5 + 7 \times 39.5) \\
 &= \frac{814.5}{35} \\
 &= 23.27 \text{ années}
 \end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 23.27 années. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par sa variance, par son écart-type ou par son coefficient de variation :

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}
 \sigma_x^2 &= \frac{1}{3} \sum_{i=1}^3 n_i (c_i - \bar{x})^2 \\
 &= \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + n_3 (c_3 - \bar{x})^2 \right] \\
 &= \frac{1}{35} \left[ 3 (8.5 - 23.27)^2 + 25 (20.5 - 23.27)^2 + 7 (39.5 - 23.27)^2 \right] \\
 &= \frac{2690.17}{35} \\
 &= 76.86 \text{ "années}^2\text{"}
 \end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}
 \sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\
 &= \left( \frac{1}{n} \sum_{i=1}^3 n_i c_i^2 \right) - \bar{x}^2 \\
 &= \frac{1}{n} (n_1 c_1^2 + n_2 c_2^2 + n_3 c_3^2) - \bar{x}^2 \\
 &= \frac{1}{35} (3 \times 8.5^2 + 25 \times 20.5^2 + 7 \times 39.5^2) - 23.27^2 \\
 &= \frac{21644.75}{35} - 541.49 \\
 &= 76.93 \text{ "années}^2\text{"}
 \end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}
 \sigma_x &= \sqrt{\sigma_x^2} \\
 &= \sqrt{76.93 \text{ "années}^2\text{"}} \\
 &= 8.77 \text{ années}
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 23.27$  années et s'en écartent de  $\sigma_x = 8.77$  années en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}
 CV &= \frac{\sigma_x}{\bar{x}} \\
 &= \frac{8.77 \text{ années}}{23.27 \text{ années}} \\
 &= 0.38
 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 23.27$  années et s'en écartent, en termes relatifs, de  $CV = 0.38$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}
 \gamma_1 &= \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \\
 &= \frac{1}{n} \left[ n_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^3 + n_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^3 + n_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^3 \right] \\
 &= \frac{1}{35} \left[ 3 \left( \frac{8.5 - 23.27}{8.77} \right)^3 + 25 \left( \frac{20.5 - 23.27}{8.77} \right)^3 + 7 \left( \frac{39.5 - 23.27}{8.77} \right)^3 \right] \\
 &= \frac{29.25}{35} \\
 &= 0.84
 \end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 23.27$  années) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}
 \beta_1 &= \left[ \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\
 &= \gamma_1^2 \\
 &= 0.7056
 \end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}
 \mu_3 &= \frac{1}{n} \sum_{i=1}^3 n_i (c_i - \bar{x})^3 \\
 &= \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^3 + n_2 (c_2 - \bar{x})^3 + n_3 (c_3 - \bar{x})^3 \right] \\
 &= \frac{1}{35} \left[ 3 (8.5 - 23.27)^3 + 25 (20.5 - 23.27)^3 + 7 (39.5 - 23.27)^3 \right] \\
 &= \frac{19728.64}{35} \\
 &= 563.68
 \end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 23.27$  années) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^4 + n_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^4 + n_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^4 \right] \\ &= \frac{1}{35} \left[ 3 \left( \frac{8.5 - 23.27}{8.77} \right)^4 + 25 \left( \frac{20.5 - 23.27}{8.77} \right)^4 + 7 \left( \frac{39.5 - 23.27}{8.77} \right)^4 \right] \\ &= \frac{106.49}{35} \\ &= 3.04\end{aligned}$$

Ce coefficient étant supérieur à 3, la distribution statistique de la variable  $X$  est en conséquence leptokurtique. Cela signifie que les observations extrêmes sont excessivement nombreuses.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \frac{1}{n} \sum_{i=1}^3 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= 0.04\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence leptokurtique. Cela signifie que les observations extrêmes sont excessivement nombreuses.

### Exercice 32 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en grammes :

$[x_i^-, x_i^+ [$	$[9, 13[$	$[13, 33[$	$[33, 42[$	$[42, 52[$	$n$
$n_i$	3	25	13	13	54

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$[x_i^-, x_i^+ [$	$c_i$	$n_i$	$n_i c_i$	$n_i c_i^2$	$n_i (c_i - \bar{x})^2$	$n_i (c_i - \bar{x})^3$	$n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3$	$n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4$
$[9, 13[$	11.0	3	33.0	363.0	1273.08	-26225.45	-19.12	35.46
$[13, 33[$	23.0	25	575.0	13225.0	1849.0	-15901.4	-11.6	8.98
$[33, 42[$	37.5	13	487.5	18281.25	452.53	2669.93	1.95	1.03
$[42, 52[$	47.0	13	611.0	28717.0	3083.08	47479.43	34.62	47.99
$\Sigma$	-	54	1706.5	60586.25	6657.69	8022.51	5.85	93.46

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^4 n_i c_i \\ &= \frac{1}{n} (n_1 c_1 + n_2 c_2 + n_3 c_3 + n_4 c_4) \\ &= \frac{1}{54} (3 \times 11.0 + 25 \times 23.0 + 13 \times 37.5 + 13 \times 47.0) \\ &= \frac{1706.5}{54} \\ &= 31.6 \text{ grammes}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 31.6 grammes. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par sa variance, par son écart-type ou par son coefficient de variation :

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \frac{1}{n} \sum_{i=1}^4 n_i (c_i - \bar{x})^2 \\ &= \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + n_3 (c_3 - \bar{x})^2 + n_4 (c_4 - \bar{x})^2 \right] \\ &= \frac{1}{54} \left[ 3(11.0 - 31.6)^2 + 25(23.0 - 31.6)^2 + 13(37.5 - 31.6)^2 + 13(47.0 - 31.6)^2 \right] \\ &= \frac{6657.69}{54} \\ &= 123.29 \text{ "grammes}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\ &= \left( \frac{1}{n} \sum_{i=1}^4 n_i c_i^2 \right) - \bar{x}^2 \\ &= \frac{1}{n} (n_1 c_1^2 + n_2 c_2^2 + n_3 c_3^2 + n_4 c_4^2) - \bar{x}^2 \\ &= \frac{1}{54} (3 \times 11.0^2 + 25 \times 23.0^2 + 13 \times 37.5^2 + 13 \times 47.0^2) - 31.6^2 \\ &= \frac{60586.25}{54} - 998.56 \\ &= 123.41 \text{ "grammes}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{123.41 \text{ "grammes}^2} \\ &= 11.11 \text{ grammes}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 31.6$  grammes et s'en écartent de  $\sigma_x = 11.11$  grammes en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned} CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{11.11 \text{ grammes}}{31.6 \text{ grammes}} \\ &= 0.35 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 31.6$  grammes et s'en écartent, en termes relatifs, de  $CV = 0.35$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned} \gamma_1 &= \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^3 + n_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^3 + n_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^3 + n_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^3 \right] \\ &= \frac{5.85}{54} \\ &= 0.11 \end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 31.6$  grammes) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned} \beta_1 &= \left[ \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0121 \end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned} \mu_3 &= \frac{1}{n} \sum_{i=1}^4 n_i (c_i - \bar{x})^3 \\ &= \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^3 + n_2 (c_2 - \bar{x})^3 + n_3 (c_3 - \bar{x})^3 + n_4 (c_4 - \bar{x})^3 \right] \\ &= \frac{1}{54} \left[ 3 (11.0 - 31.6)^3 + 25 (23.0 - 31.6)^3 + 13 (37.5 - 31.6)^3 + 13 (47.0 - 31.6)^3 \right] \\ &= \frac{8022.51}{54} \\ &= 148.56 \end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 31.6$  grammes) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^4 + n_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^4 + n_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^4 + n_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^4 \right] \\ &= \frac{93.46}{54} \\ &= 1.73\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \frac{1}{n} \sum_{i=1}^4 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -1.27\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### Exercice 33 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en dirhams :

$[x_i^-, x_i^+[$	$[7, 18[$	$[18, 30[$	$[30, 49[$	$[49, 54[$	$[54, 65[$	$n$
$n_i$	9	15	11	11	12	58

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$[x_i^-, x_i^+[$	$c_i$	$n_i$	$n_i c_i$	$n_i c_i^2$	$n_i (c_i - \bar{x})^2$	$n_i (c_i - \bar{x})^3$	$n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3$	$n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4$
$[7, 18[$	12.5	9	112.5	1406.25	5724.44	-144370.27	-30.5	45.82
$[18, 30[$	24.0	15	360.0	8640.0	2823.58	-38739.46	-8.18	6.69
$[30, 49[$	39.5	11	434.5	17162.75	34.85	62.04	0.01	0.0
$[49, 54[$	51.5	11	566.5	29174.75	2088.77	28783.28	6.08	4.99
$[54, 65[$	59.5	12	714.0	42483.0	5692.42	123980.93	26.19	33.98
$\Sigma$	-	58	2187.5	98866.75	16364.06	-154264.41	-6.4	91.48

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^5 n_i c_i \\ &= \frac{1}{n} (n_1 c_1 + n_2 c_2 + n_3 c_3 + n_4 c_4 + n_5 c_5) \\ &= \frac{1}{58} (9 \times 12.5 + 15 \times 24.0 + 11 \times 39.5 + 11 \times 51.5 + 12 \times 59.5) \\ &= \frac{2187.5}{58} \\ &= 37.72 \text{ dirhams}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 37.72 dirhams. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par sa variance, par son écart-type ou par son coefficient de variation :

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \frac{1}{n} \sum_{i=1}^5 n_i (c_i - \bar{x})^2 \\ &= \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + n_3 (c_3 - \bar{x})^2 + n_4 (c_4 - \bar{x})^2 + n_5 (c_5 - \bar{x})^2 \right] \\ &= \frac{16364.06}{58} \\ &= 282.14 \text{ "dirhams}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\ &= \left( \frac{1}{n} \sum_{i=1}^5 n_i c_i^2 \right) - \bar{x}^2 \\ &= \frac{1}{n} (n_1 c_1^2 + n_2 c_2^2 + n_3 c_3^2 + n_4 c_4^2 + n_5 c_5^2) - \bar{x}^2 \\ &= \frac{1}{58} (9 \times 12.5^2 + 15 \times 24.0^2 + 11 \times 39.5^2 + 11 \times 51.5^2 + 12 \times 59.5^2) - 37.72^2 \\ &= \frac{98866.75}{58} - 1422.8 \\ &= 281.8 \text{ "dirhams}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{281.8 \text{ "dirhams}^2} \\ &= 16.79 \text{ dirhams}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 37.72$  dirhams et s'en écartent de  $\sigma_x = 16.79$  dirhams en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned} CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{16.79 \text{ dirhams}}{37.72 \text{ dirhams}} \\ &= 0.45 \end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 37.72$  dirhams et s'en écartent, en termes relatifs, de  $CV = 0.45$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned} \gamma_1 &= \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^3 + n_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^3 + n_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^3 + n_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^3 + n_5 \left( \frac{c_5 - \bar{x}}{\sigma_x} \right)^3 \right] \\ &= \frac{-6.4}{58} \\ &= -0.11 \end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à gauche. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 37.72$  dirhams) sont moins fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned} \beta_1 &= \left[ \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0121 \end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned} \mu_3 &= \frac{1}{n} \sum_{i=1}^5 n_i (c_i - \bar{x})^3 \\ &= \frac{1}{n} \left[ n_1 (c_1 - \bar{x})^3 + n_2 (c_2 - \bar{x})^3 + n_3 (c_3 - \bar{x})^3 + n_4 (c_4 - \bar{x})^3 + n_5 (c_5 - \bar{x})^3 \right] \\ &= \frac{-154264.41}{58} \\ &= -2659.73 \end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à gauche puisque le moment centré  $\mu_3$  est négatif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 37.72$  dirhams) sont moins fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 \\ &= \frac{1}{n} \left[ n_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^4 + n_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^4 + n_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^4 + n_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^4 + n_5 \left( \frac{c_5 - \bar{x}}{\sigma_x} \right)^4 \right] \\ &= \frac{91.48}{58} \\ &= 1.58\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \frac{1}{n} \sum_{i=1}^5 n_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -1.42\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### Exercice 34 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  mesurée en heures :

$[x_i^-, x_i^+[$	$f_i$	$[5, 18[$	$[18, 26[$	$[26, 43[$
		0.32	0.5	0.18

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$[x_i^-, x_i^+[$	$c_i$	$f_i$	$f_i c_i$	$f_i c_i^2$	$f_i (c_i - \bar{x})^2$	$f_i (c_i - \bar{x})^3$	$f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3$	$f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4$
$[5, 18[$	11.5	0.32	3.68	42.32	28.22	-264.94	-0.54	0.65
$[18, 26[$	22.0	0.5	11.0	242.0	0.62	0.68	0.0	0.0
$[26, 43[$	34.5	0.18	6.21	214.24	33.34	453.78	0.93	1.6
$\Sigma$	-	1	20.89	498.56	62.18	189.52	0.39	2.25

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \sum_{i=1}^3 f_i c_i \\ &= f_1 c_1 + f_2 c_2 + f_3 c_3 \\ &= 0.32 \times 11.5 + 0.5 \times 22.0 + 0.18 \times 34.5 \\ &= 20.89 \text{ heures}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 20.89 heures. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par sa variance, par son écart-type ou par son coefficient de variation :

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \sum_{i=1}^3 f_i (c_i - \bar{x})^2 \\ &= f_1 (c_1 - \bar{x})^2 + f_2 (c_2 - \bar{x})^2 + f_3 (c_3 - \bar{x})^2 \\ &= 0.32 (11.5 - 20.89)^2 + 0.5 (22.0 - 20.89)^2 + 0.18 (34.5 - 20.89)^2 \\ &= 62.18 \text{ "heures"}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\ &= \left( \sum_{i=1}^3 f_i c_i^2 \right) - \bar{x}^2 \\ &= (f_1 c_1^2 + f_2 c_2^2 + f_3 c_3^2) - \bar{x}^2 \\ &= (0.32 \times 11.5^2 + 0.5 \times 22.0^2 + 0.18 \times 34.5^2) - 20.89^2 \\ &= 498.565 - 436.39 \\ &= 62.17 \text{ "heures"}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{62.17 \text{ "heures"}^2} \\ &= 7.88 \text{ heures}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 20.89$  heures et s'en écartent de  $\sigma_x = 7.88$  heures en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{7.88 \text{ heures}}{20.89 \text{ heures}} \\ &= 0.38\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 20.89$  heures et s'en écartent, en termes relatifs, de  $CV = 0.38$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \sum_{i=1}^3 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \\ &= f_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^3 + f_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^3 + f_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^3 \\ &= 0.32 \left( \frac{11.5 - 20.89}{7.88} \right)^3 + 0.5 \left( \frac{22.0 - 20.89}{7.88} \right)^3 + 0.18 \left( \frac{34.5 - 20.89}{7.88} \right)^3 \\ &= 0.39\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 20.89$  heures) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \sum_{i=1}^3 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.1521\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \sum_{i=1}^3 f_i (c_i - \bar{x})^3 \\ &= f_1 (c_1 - \bar{x})^3 + f_2 (c_2 - \bar{x})^3 + f_3 (c_3 - \bar{x})^3 \\ &= 0.32 (11.5 - 20.89)^3 + 0.5 (22.0 - 20.89)^3 + 0.18 (34.5 - 20.89)^3 \\ &= 189.52\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 20.89$  heures) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \sum_{i=1}^3 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 \\ &= f_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^4 + f_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^4 + f_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^4 \\ &= 0.32 \left( \frac{11.5 - 20.89}{7.88} \right)^4 + 0.5 \left( \frac{22.0 - 20.89}{7.88} \right)^4 + 0.18 \left( \frac{34.5 - 20.89}{7.88} \right)^4 \\ &= 2.25\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \sum_{i=1}^3 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -0.75\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### Exercice 35 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  mesurée en dirhams :

$[x_i^-, x_i^+[$	$[7, 19[$	$[19, 39[$	$[39, 42[$	$[42, 51[$
$f_i$	0.12	0.32	0.35	0.21

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$[x_i^-, x_i^+[$	$c_i$	$f_i$	$f_i c_i$	$f_i c_i^2$	$f_i (c_i - \bar{x})^2$	$f_i (c_i - \bar{x})^3$	$f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3$	$f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4$
$[7, 19[$	13.0	0.12	1.56	20.28	56.92	-1239.81	-1.11	2.32
$[19, 39[$	29.0	0.32	9.28	269.12	10.69	-61.79	-0.06	0.03
$[39, 42[$	40.5	0.35	14.18	574.09	11.45	65.5	0.06	0.03
$[42, 51[$	46.5	0.21	9.76	454.07	28.85	338.07	0.3	0.34
$\Sigma$	-	1	34.78	1317.56	107.91	-898.03	-0.81	2.72

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \sum_{i=1}^4 f_i c_i \\ &= f_1 c_1 + f_2 c_2 + f_3 c_3 + f_4 c_4 \\ &= 0.12 \times 13.0 + 0.32 \times 29.0 + 0.35 \times 40.5 + 0.21 \times 46.5 \\ &= 34.78 \text{ dirhams}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 34.78 dirhams. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par sa variance, par son écart-type ou par son coefficient de variation :

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \sum_{i=1}^4 f_i (c_i - \bar{x})^2 \\ &= f_1 (c_1 - \bar{x})^2 + f_2 (c_2 - \bar{x})^2 + f_3 (c_3 - \bar{x})^2 + f_4 (c_4 - \bar{x})^2 \\ &= 0.12 (13.0 - 34.78)^2 + 0.32 (29.0 - 34.78)^2 + 0.35 (40.5 - 34.78)^2 + 0.21 (46.5 - 34.78)^2 \\ &= 107.91 \text{ "dirhams}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\ &= \left( \sum_{i=1}^4 f_i c_i^2 \right) - \bar{x}^2 \\ &= (f_1 c_1^2 + f_2 c_2^2 + f_3 c_3^2 + f_4 c_4^2) - \bar{x}^2 \\ &= (0.12 \times 13.0^2 + 0.32 \times 29.0^2 + 0.35 \times 40.5^2 + 0.21 \times 46.5^2) - 34.78^2 \\ &= 1317.56 - 1209.65 \\ &= 107.91 \text{ "dirhams}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{107.91 \text{ "dirhams}^2} \\ &= 10.39 \text{ dirhams}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 34.78$  dirhams et s'en écartent de  $\sigma_x = 10.39$  dirhams en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{10.39 \text{ dirhams}}{34.78 \text{ dirhams}} \\ &= 0.3\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 34.78$  dirhams et s'en écartent, en termes relatifs, de  $CV = 0.3$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \sum_{i=1}^4 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \\ &= f_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^3 + f_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^3 + f_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^3 + f_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^3 \\ &= -0.81\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à gauche. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 34.78$  dirhams) sont moins fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \sum_{i=1}^4 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.6561\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \sum_{i=1}^4 f_i (c_i - \bar{x})^3 \\ &= f_1 (c_1 - \bar{x})^3 + f_2 (c_2 - \bar{x})^3 + f_3 (c_3 - \bar{x})^3 + f_4 (c_4 - \bar{x})^3 \\ &= 0.12 (13.0 - 34.78)^3 + 0.32 (29.0 - 34.78)^3 + 0.35 (40.5 - 34.78)^3 + 0.21 (46.5 - 34.78)^3 \\ &= -898.03\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à gauche puisque le moment centré  $\mu_3$  est négatif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 34.78$  dirhams) sont moins fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \sum_{i=1}^4 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 \\ &= f_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^4 + f_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^4 + f_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^4 + f_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^4 \\ &= 2.72\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \sum_{i=1}^4 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -0.28\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

### Exercice 36 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  mesurée en grammes :

$[x_i^-, x_i^+]$	$f_i$	$[6, 18[$	$[18, 27[$	$[27, 44[$	$[44, 51[$	$[51, 65[$
		0.01	0.27	0.36	0.25	0.11

1. Calculer la moyenne arithmétique de la variable statistique  $X$  ;
2. Mesurer la dispersion de la distribution statistique de  $X$  autour de sa moyenne arithmétique ;
3. Mesurer l'asymétrie de la distribution statistique de  $X$  ;
4. Mesurer l'aplatissement de la distribution statistique de  $X$ .

### Solution :

Les calculs nécessaires à la résolution de l'exercice sont donnés dans le tableau suivant :

$[x_i^-, x_i^+]$	$c_i$	$f_i$	$f_i c_i$	$f_i c_i^2$	$f_i (c_i - \bar{x})^2$	$f_i (c_i - \bar{x})^3$	$f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3$	$f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4$
$[6, 18[$	12.0	0.01	0.12	1.44	6.37	-160.6	-0.1	0.21
$[18, 27[$	22.5	0.27	6.08	136.69	58.58	-862.92	-0.52	0.65
$[27, 44[$	35.5	0.36	12.78	453.69	1.08	-1.86	-0.0	0.0
$[44, 51[$	47.5	0.25	11.88	564.06	26.37	270.8	0.16	0.14
$[51, 65[$	58.0	0.11	6.38	370.04	47.45	985.6	0.6	1.05
$\Sigma$	-	1	37.23	1525.92	139.85	231.02	0.14	2.05

1. La moyenne arithmétique  $\bar{x}$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\bar{x} &= \sum_{i=1}^5 f_i c_i \\ &= f_1 c_1 + f_2 c_2 + f_3 c_3 + f_4 c_4 + f_5 c_5 \\ &= 0.01 \times 12.0 + 0.27 \times 22.5 + 0.36 \times 35.5 + 0.25 \times 47.5 + 0.11 \times 58.0 \\ &= 37.23 \text{ grammes}\end{aligned}$$

Le centre de la distribution statistique de  $X$  se situe à 37.23 grammes. Cette valeur centrale concentre toute l'information contenue dans les observations de la variable  $X$  étudiée.

2. La dispersion d'une distribution statistique autour de sa moyenne arithmétique est mesurée par sa variance, par son écart-type ou par son coefficient de variation :

- La variance  $\sigma_x^2$  de la variable statistique  $X$  est donnée par :

$$\begin{aligned}\sigma_x^2 &= \sum_{i=1}^5 f_i (c_i - \bar{x})^2 \\ &= f_1 (c_1 - \bar{x})^2 + f_2 (c_2 - \bar{x})^2 + f_3 (c_3 - \bar{x})^2 + f_4 (c_4 - \bar{x})^2 + f_5 (c_5 - \bar{x})^2 \\ &= 139.85 \text{ "grammes}^2\end{aligned}$$

La variance  $\sigma_x^2$  peut également se calculer selon la formule réduite de König-Huygens comme suit :

$$\begin{aligned}\sigma_x^2 &= \overline{x^2} - \bar{x}^2 \\ &= \left( \sum_{i=1}^5 f_i c_i^2 \right) - \bar{x}^2 \\ &= (f_1 c_1^2 + f_2 c_2^2 + f_3 c_3^2 + f_4 c_4^2 + f_5 c_5^2) - \bar{x}^2 \\ &= (0.01 \times 12.0^2 + 0.27 \times 22.5^2 + 0.36 \times 35.5^2 + 0.25 \times 47.5^2 + 0.11 \times 58.0^2) - 37.23^2 \\ &= 1525.92 - 1386.07 \\ &= 139.85 \text{ "grammes}^2\end{aligned}$$

- L'écart-type  $\sigma_x$  est égal à la racine carrée de la variance  $\sigma_x^2$  (on retiendra le résultat de la formule réduite) :

$$\begin{aligned}\sigma_x &= \sqrt{\sigma_x^2} \\ &= \sqrt{139.85 \text{ "grammes}^2} \\ &= 11.83 \text{ grammes}\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 37.23$  grammes et s'en écartent de  $\sigma_x = 11.83$  grammes en moyenne.

- Le coefficient de variation est une mesure de dispersion relative des observations autour de la moyenne arithmétique  $\bar{x}$ , il est obtenu comme suit :

$$\begin{aligned}CV &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{11.83 \text{ grammes}}{37.23 \text{ grammes}} \\ &= 0.32\end{aligned}$$

Les observations de la variable statistique  $X$  sont dispersées autour du centre  $\bar{x} = 37.23$  grammes et s'en écartent, en termes relatifs, de  $CV = 0.32$  en moyenne.

3. L'asymétrie d'une distribution statistique est mesurée par un coefficient d'asymétrie :

- Le coefficient d'asymétrie (ou de *skewness*) de Fisher  $\gamma_1$  est donné par :

$$\begin{aligned}\gamma_1 &= \sum_{i=1}^5 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \\ &= f_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^3 + f_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^3 + f_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^3 + f_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^3 + f_5 \left( \frac{c_5 - \bar{x}}{\sigma_x} \right)^3 \\ &= 0.14\end{aligned}$$

Ce coefficient étant positif, la distribution statistique de la variable  $X$  est en conséquence asymétrique à droite. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 37.23$  grammes) sont plus fréquentes que celles qui lui sont inférieures.

- Le coefficient d'asymétrie (ou de *skewness normalisé*) de Pearson  $\beta_1$  est donné par :

$$\begin{aligned}\beta_1 &= \left[ \sum_{i=1}^5 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 \right]^2 \\ &= \gamma_1^2 \\ &= 0.0196\end{aligned}$$

Ce coefficient étant non nul signifie que la distribution statistique de  $X$  est asymétrique. Le sens de l'asymétrie est donné par le signe du moment centré  $\mu_3$  donné par :

$$\begin{aligned}\mu_3 &= \sum_{i=1}^5 f_i (c_i - \bar{x})^3 \\ &= f_1 (c_1 - \bar{x})^3 + f_2 (c_2 - \bar{x})^3 + f_3 (c_3 - \bar{x})^3 + f_4 (c_4 - \bar{x})^3 + f_5 (c_5 - \bar{x})^3 \\ &= 231.02\end{aligned}$$

La distribution statistique de la variable  $X$  est asymétrique à droite puisque le moment centré  $\mu_3$  est positif. Cela signifie que les observations qui sont supérieures à la moyenne arithmétique ( $\bar{x} = 37.23$  grammes) sont plus fréquentes que celles qui lui sont inférieures.

4. L'aplatissement d'une distribution statistique est mesurée par un coefficient d'aplatissement :

- Le coefficient d'aplatissement (ou de *kurtosis*) de Pearson  $\beta_2$  est donné par :

$$\begin{aligned}\beta_2 &= \sum_{i=1}^5 f_i \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 \\ &= f_1 \left( \frac{c_1 - \bar{x}}{\sigma_x} \right)^4 + f_2 \left( \frac{c_2 - \bar{x}}{\sigma_x} \right)^4 + f_3 \left( \frac{c_3 - \bar{x}}{\sigma_x} \right)^4 + f_4 \left( \frac{c_4 - \bar{x}}{\sigma_x} \right)^4 + f_5 \left( \frac{c_5 - \bar{x}}{\sigma_x} \right)^4 \\ &= 2.05\end{aligned}$$

Ce coefficient étant inférieur à 3, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

- Le coefficient d'aplatissement (ou d'*excès de kurtosis*) de Fisher  $\gamma_2$  est donné par :

$$\begin{aligned}\gamma_2 &= \sum_{i=1}^5 f_i \left( \frac{x_i - \bar{x}}{\sigma_x} \right)^4 - 3 \\ &= \beta_2 - 3 \\ &= -0.95\end{aligned}$$

Ce coefficient étant négatif, la distribution statistique de la variable  $X$  est en conséquence platykurtique. Cela signifie que les observations extrêmes sont excessivement rares.

## Conclusion

Dans cette section, ont été résolus des exercices sur des mesures numériques basées sur les moments, telles que les mesures de tendance centrale, les mesures de dispersion et les mesures de forme, dans le cas d'une seule variable quantitative. Dans la section suivante seront résolus des exercices sur des mesures numériques basées sur les quantiles, telles que la médiane, les quantiles et les écarts interquantiles, toujours dans le cas d'une seule variable quantitative.

## 2.2 Statistique descriptive basée sur les quantiles

Dans cette section, seront résolus des exercices sur des mesures numériques basées sur les quantiles, telles que la médiane, les quantiles et les écarts interquantiles, dans le cas d'une variable quantitative discrète (sous-section 2.2.1) et dans le cas d'une variable quantitative continue (sous-section 2.2.2).

### 2.2.1 Variable quantitative discrète

Une variable quantitative discrète  $X$  est une variable statistique qui prend en ensemble fini et dénombrable de valeurs numériques notées  $x_i$ .

#### Exercice 37 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en motos :

$x_i$	5	18	26
$n_i$	6	39	32

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.32 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

#### Solution :

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$x_i$	$n_i$	$f_i$	$F_i^+$
5	6	0.08	0.08
18	39	0.51	0.59
26	32	0.41	1
$\Sigma$	77	1	-

## 1. La médiane :

La médiane, notée  $M_e$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à une fréquence cumulée croissante de 0.5. On note qu'aucune fréquence cumulée croissante n'est égale à 0.5, la médiane est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. La médiane est donc égale à  $M_e = 18$  motos. Cela signifie que 50% des observations environ sont inférieures à 18 motos et 50% des observations environ sont supérieures à cette valeur.

## 2. Le premier et le dernier quartiles :

Le premier quartile, noté  $Q_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 25% et 75% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.25. On note qu'aucune fréquence cumulée croissante n'est égale à 0.25, le premier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier quartile est donc égal à  $Q_1 = 18$  motos. Cela signifie que 25% des observations environ sont inférieures à 18 motos et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, noté  $Q_3$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 75% et 25% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.75. On note qu'aucune fréquence cumulée croissante n'est égale à 0.75, le dernier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier quartile est donc égal à  $Q_3 = 26$  motos. Cela signifie que 75% des observations environ sont inférieures à 26 motos et 25% des observations environ sont supérieures à cette valeur.

## 3. Le premier et le dernier déciles :

Le premier décile, noté  $D_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 10% et 90% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.10. On note qu'aucune fréquence cumulée croissante n'est égale à 0.10, le premier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier décile est donc égal à  $D_1 = 18$  motos. Cela signifie que 10% des observations environ sont inférieures à 18 motos et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, noté  $D_9$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 90% et 10% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.90. On note qu'aucune fréquence cumulée croissante n'est égale à 0.90, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier décile est donc égal à  $D_9 = 26$  motos. Cela signifie que 90% des observations environ sont inférieures à 26 motos et 10% des observations environ sont supérieures à cette valeur.

## 4. Le premier et le dernier centiles :

Le premier centile, noté  $C_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 1% et 99% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.01. On note qu'aucune fréquence cumulée croissante n'est égale à 0.01, le premier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier centile est donc égal à  $C_1 = 5$  motos. Cela signifie que 1% des observations environ sont inférieures à 5 motos et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, noté  $C_{99}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 99% et 1% des observations environ. Il est en

conséquence associé à une fréquence cumulée croissante de 0.99. On note qu'aucune fréquence cumulée croissante n'est égale à 0.99, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier centile est donc égal à  $C_{99} = 26$  motos. Cela signifie que 99% des observations environ sont inférieures à 26 motos et 1% des observations environ sont supérieures à cette valeur.

5. Le quantile d'ordre 0.32 :

Le quantile d'ordre 0.32, noté  $q_{0.32}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.32 et une proportion 0.68 des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.32. On note qu'aucune fréquence cumulée croissante n'est égale à 0.32, le quantile d'ordre 0.32 est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le quantile d'ordre 0.32 est donc égal à  $q_{0.32} = 18$  motos.

6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [5, 26]$ ). L'étendue est donc égale à  $E = 26 - 5 = 21$  motos.

7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles ( i.e.  $IIQ = [18, 26]$ ). L'écart interquartile est donc égal à  $EIQ = 26 - 18 = 8$  motos.

8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [18, 26]$ ). L'écart interdécile est donc égal à  $EID = 26 - 18 = 8$  motos.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée  $EIC$ , est l'amplitude de son intervalle intercentile  $IIC$  borné par son premier et son dernier centiles ( i.e.  $IIC = [5, 26]$ ). L'écart intercentile est donc égal à  $EIC = 26 - 5 = 21$  motos.

### Exercice 38 :

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en ouvriers :

$x_i$	1	19	23	42
$n_i$	7	28	36	14

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.29 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;

8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$x_i$	$n_i$	$f_i$	$F_i^+$
1	7	0.08	0.08
19	28	0.33	0.41
23	36	0.42	0.83
42	14	0.17	1
$\Sigma$	85	1	–

1. La médiane :

La médiane, notée  $M_e$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à une fréquence cumulée croissante de 0.5. On note qu'aucune fréquence cumulée croissante n'est égale à 0.5, la médiane est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. La médiane est donc égale à  $M_e = 23$  ouvriers. Cela signifie que 50% des observations environ sont inférieures à 23 ouvriers et 50% des observations environ sont supérieures à cette valeur.

2. Le premier et le dernier quartiles :

Le premier quartile, noté  $Q_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 25% et 75% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.25. On note qu'aucune fréquence cumulée croissante n'est égale à 0.25, le premier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier quartile est donc égal à  $Q_1 = 19$  ouvriers. Cela signifie que 25% des observations environ sont inférieures à 19 ouvriers et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, noté  $Q_3$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 75% et 25% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.75. On note qu'aucune fréquence cumulée croissante n'est égale à 0.75, le dernier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier quartile est donc égal à  $Q_3 = 23$  ouvriers. Cela signifie que 75% des observations environ sont inférieures à 23 ouvriers et 25% des observations environ sont supérieures à cette valeur.

3. Le premier et le dernier déciles :

Le premier décile, noté  $D_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 10% et 90% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.10. On note qu'aucune fréquence cumulée croissante n'est égale à 0.10, le premier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier décile est donc égal à  $D_1 = 19$  ouvriers. Cela signifie que 10% des observations environ sont inférieures à 19 ouvriers et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, noté  $D_9$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 90% et 10% des observations environ. Il est

en conséquence associé à une fréquence cumulée croissante de 0.90. On note qu'aucune fréquence cumulée croissante n'est égale à 0.90, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier décile est donc égal à  $D_9 = 42$  ouvriers. Cela signifie que 90% des observations environ sont inférieures à 42 ouvriers et 10% des observations environ sont supérieures à cette valeur.

4. Le premier et le dernier centiles :

Le premier centile, noté  $C_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 1% et 99% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.01. On note qu'aucune fréquence cumulée croissante n'est égale à 0.01, le premier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier centile est donc égal à  $C_1 = 1$  ouvriers. Cela signifie que 1% des observations environ sont inférieures à 1 ouvriers et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, noté  $C_{99}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 99% et 1% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.99. On note qu'aucune fréquence cumulée croissante n'est égale à 0.99, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier centile est donc égal à  $C_{99} = 42$  ouvriers. Cela signifie que 99% des observations environ sont inférieures à 42 ouvriers et 1% des observations environ sont supérieures à cette valeur.

5. Le quantile d'ordre 0.29 :

Le quantile d'ordre 0.29, noté  $q_{0.29}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.29 et une proportion 0.71 des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.29. On note qu'aucune fréquence cumulée croissante n'est égale à 0.29, le quantile d'ordre 0.29 est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le quantile d'ordre 0.29 est donc égal à  $q_{0.29} = 19$  ouvriers.

6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [1, 42]$ ). L'étendue est donc égale à  $E = 42 - 1 = 41$  ouvriers.

7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles ( i.e.  $IIQ = [19, 23]$ ). L'écart interquartile est donc égal à  $EIQ = 23 - 19 = 4$  ouvriers.

8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [19, 42]$ ). L'écart interdécile est donc égal à  $EID = 42 - 19 = 23$  ouvriers.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée  $EIC$ , est l'amplitude de son intervalle intercentile  $IIC$  borné par son premier et son dernier centiles ( i.e.  $IIC = [1, 42]$ ). L'écart intercentile est donc égal à  $EIC = 42 - 1 = 41$  ouvriers.

**Exercice 39 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en voitures :

$x_i$	7	17	26	41	56
$n_i$	4	21	20	11	8

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.24 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$x_i$	$n_i$	$f_i$	$F_i^+$
7	4	0.06	0.06
17	21	0.33	0.39
26	20	0.31	0.7
41	11	0.17	0.87
56	8	0.13	1
$\Sigma$	64	1	—

1. La médiane :

La médiane, notée  $M_e$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à une fréquence cumulée croissante de 0.5. On note qu'aucune fréquence cumulée croissante n'est égale à 0.5, la médiane est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. La médiane est donc égale à  $M_e = 26$  voitures. Cela signifie que 50% des observations environ sont inférieures à 26 voitures et 50% des observations environ sont supérieures à cette valeur.

2. Le premier et le dernier quartiles :

Le premier quartile, noté  $Q_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 25% et 75% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.25. On note qu'aucune fréquence cumulée croissante n'est égale à 0.25, le premier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier quartile est donc égal à  $Q_1 = 17$  voitures. Cela signifie que 25% des observations environ sont inférieures à 17 voitures et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, noté  $Q_3$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 75% et 25% des observations environ. Il est

en conséquence associé à une fréquence cumulée croissante de 0.75. On note qu'aucune fréquence cumulée croissante n'est égale à 0.75, le dernier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier quartile est donc égal à  $Q_3 = 41$  voitures. Cela signifie que 75% des observations environ sont inférieures à 41 voitures et 25% des observations environ sont supérieures à cette valeur.

3. Le premier et le dernier déciles :

Le premier décile, noté  $D_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 10% et 90% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.10. On note qu'aucune fréquence cumulée croissante n'est égale à 0.10, le premier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier décile est donc égal à  $D_1 = 17$  voitures. Cela signifie que 10% des observations environ sont inférieures à 17 voitures et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, noté  $D_9$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 90% et 10% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.90. On note qu'aucune fréquence cumulée croissante n'est égale à 0.90, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier décile est donc égal à  $D_9 = 56$  voitures. Cela signifie que 90% des observations environ sont inférieures à 56 voitures et 10% des observations environ sont supérieures à cette valeur.

4. Le premier et le dernier centiles :

Le premier centile, noté  $C_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 1% et 99% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.01. On note qu'aucune fréquence cumulée croissante n'est égale à 0.01, le premier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier centile est donc égal à  $C_1 = 7$  voitures. Cela signifie que 1% des observations environ sont inférieures à 7 voitures et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, noté  $C_{99}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 99% et 1% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.99. On note qu'aucune fréquence cumulée croissante n'est égale à 0.99, le dernier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier centile est donc égal à  $C_{99} = 56$  voitures. Cela signifie que 99% des observations environ sont inférieures à 56 voitures et 1% des observations environ sont supérieures à cette valeur.

5. Le quantile d'ordre 0.24 :

Le quantile d'ordre 0.24, noté  $q_{0.24}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.24 et une proportion 0.76 des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.24. On note qu'aucune fréquence cumulée croissante n'est égale à 0.24, le quantile d'ordre 0.24 est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le quantile d'ordre 0.24 est donc égal à  $q_{0.24} = 17$  voitures.

6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [7, 56]$ ). L'étendue est donc égale à  $E = 56 - 7 = 49$  voitures.

## 7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée EIQ, est l'amplitude de son intervalle interquartile IIQ borné par son premier et son dernier quartiles ( i.e.  $\text{IIQ} = [17, 41]$ ). L'écart interquartile est donc égal à  $\text{EIQ} = 41 - 17 = 24$  voitures.

## 8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée EID, est l'amplitude de son intervalle interdécile IID borné par son premier et son dernier déciles ( i.e.  $\text{IID} = [17, 56]$ ). L'écart interdécile est donc égal à  $\text{EID} = 56 - 17 = 39$  voitures.

## 9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $\text{IIC} = [7, 56]$ ). L'écart intercentile est donc égal à  $\text{EIC} = 56 - 7 = 49$  voitures.

**Exercice 40 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en voitures :

$x_i$	7	11	39
$f_i$	0.06	0.49	0.45

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.08 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$x_i$	$f_i$	$F_i^+$
7	0.06	0.06
11	0.49	0.55
39	0.45	1
$\Sigma$	1	-

## 1. La médiane :

La médiane, notée  $M_e$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à une fréquence cumulée croissante de 0.5. On note qu'aucune fréquence cumulée croissante n'est égale à 0.5, la médiane est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. La médiane est donc égale à  $M_e = 11$  voitures. Cela signifie que 50% des observations environ sont inférieures à 11 voitures et 50% des observations environ sont supérieures à cette valeur.

## 2. Le premier et le dernier quartiles :

Le premier quartile, noté  $Q_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 25% et 75% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.25. On note qu'aucune fréquence cumulée croissante n'est égale à 0.25, le premier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier quartile est donc égal à  $Q_1 = 11$  voitures. Cela signifie que 25% des observations environ sont inférieures à 11 voitures et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, noté  $Q_3$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 75% et 25% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.75. On note qu'aucune fréquence cumulée croissante n'est égale à 0.75, le dernier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier quartile est donc égal à  $Q_3 = 39$  voitures. Cela signifie que 75% des observations environ sont inférieures à 39 voitures et 25% des observations environ sont supérieures à cette valeur.

## 3. Le premier et le dernier déciles :

Le premier décile, noté  $D_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 10% et 90% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.10. On note qu'aucune fréquence cumulée croissante n'est égale à 0.10, le premier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier décile est donc égal à  $D_1 = 11$  voitures. Cela signifie que 10% des observations environ sont inférieures à 11 voitures et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, noté  $D_9$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 90% et 10% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.90. On note qu'aucune fréquence cumulée croissante n'est égale à 0.90, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier décile est donc égal à  $D_9 = 39$  voitures. Cela signifie que 90% des observations environ sont inférieures à 39 voitures et 10% des observations environ sont supérieures à cette valeur.

## 4. Le premier et le dernier centiles :

Le premier centile, noté  $C_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 1% et 99% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.01. On note qu'aucune fréquence cumulée croissante n'est égale à 0.01, le premier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier centile est donc égal à  $C_1 = 7$  voitures. Cela signifie que 1% des observations environ sont inférieures à 7 voitures et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, noté  $C_{99}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 99% et 1% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.99. On note qu'aucune fréquence cumulée croissante n'est égale à 0.99, le dernier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier centile est donc égal à  $C_{99} = 39$  voitures. Cela signifie que 99% des observations environ sont inférieures à 39 voitures et 1% des observations environ sont supérieures à cette valeur.

## 5. Le quantile d'ordre 0.08 :

Le quantile d'ordre 0.08, noté  $q_{0.08}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.08 et une proportion 0.92 des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.08. On note qu'aucune fréquence cumulée croissante n'est égale à 0.08, le quantile d'ordre 0.08 est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le quantile d'ordre 0.08 est donc égal à  $q_{0.08} = 11$  voitures.

6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [7, 39]$ ). L'étendue est donc égale à  $E = 39 - 7 = 32$  voitures.

7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles ( i.e.  $IIQ = [11, 39]$ ). L'écart interquartile est donc égal à  $EIQ = 39 - 11 = 28$  voitures.

8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [11, 39]$ ). L'écart interdécile est donc égal à  $EID = 39 - 11 = 28$  voitures.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée  $EIC$ , est l'amplitude de son intervalle intercentile  $IIC$  borné par son premier et son dernier centiles ( i.e.  $IIC = [7, 39]$ ). L'écart intercentile est donc égal à  $EIC = 39 - 7 = 32$  voitures.

**Exercice 41 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en motos :

$x_i$	2	15	22	46
$f_i$	0.08	0.41	0.3	0.21

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.27 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$x_i$	$f_i$	$F_i^+$
2	0.08	0.08
15	0.41	0.49
22	0.3	0.79
46	0.21	1
$\Sigma$	1	–

## 1. La médiane :

La médiane, notée  $M_e$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à une fréquence cumulée croissante de 0.5. On note qu'aucune fréquence cumulée croissante n'est égale à 0.5, la médiane est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. La médiane est donc égale à  $M_e = 22$  motos. Cela signifie que 50% des observations environ sont inférieures à 22 motos et 50% des observations environ sont supérieures à cette valeur.

## 2. Le premier et le dernier quartiles :

Le premier quartile, noté  $Q_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 25% et 75% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.25. On note qu'aucune fréquence cumulée croissante n'est égale à 0.25, le premier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier quartile est donc égal à  $Q_1 = 15$  motos. Cela signifie que 25% des observations environ sont inférieures à 15 motos et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, noté  $Q_3$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 75% et 25% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.75. On note qu'aucune fréquence cumulée croissante n'est égale à 0.75, le dernier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier quartile est donc égal à  $Q_3 = 22$  motos. Cela signifie que 75% des observations environ sont inférieures à 22 motos et 25% des observations environ sont supérieures à cette valeur.

## 3. Le premier et le dernier déciles :

Le premier décile, noté  $D_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 10% et 90% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.10. On note qu'aucune fréquence cumulée croissante n'est égale à 0.10, le premier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier décile est donc égal à  $D_1 = 15$  motos. Cela signifie que 10% des observations environ sont inférieures à 15 motos et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, noté  $D_9$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 90% et 10% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.90. On note qu'aucune fréquence cumulée croissante n'est égale à 0.90, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier décile est donc égal à  $D_9 = 46$  motos. Cela signifie que 90% des observations environ sont inférieures à 46 motos et 10% des observations environ sont supérieures à cette valeur.

## 4. Le premier et le dernier centiles :

Le premier centile, noté  $C_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 1% et 99% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.01. On note qu'aucune fréquence

cumulée croissante n'est égale à 0.01, le premier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier centile est donc égal à  $C_1 = 2$  motos. Cela signifie que 1% des observations environ sont inférieures à 2 motos et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, noté  $C_{99}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 99% et 1% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.99. On note qu'aucune fréquence cumulée croissante n'est égale à 0.99, le dernier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier centile est donc égal à  $C_{99} = 46$  motos. Cela signifie que 99% des observations environ sont inférieures à 46 motos et 1% des observations environ sont supérieures à cette valeur.

5. Le quantile d'ordre 0.27 :

Le quantile d'ordre 0.27, noté  $q_{0.27}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.27 et une proportion 0.73 des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.27. On note qu'aucune fréquence cumulée croissante n'est égale à 0.27, le quantile d'ordre 0.27 est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le quantile d'ordre 0.27 est donc égal à  $q_{0.27} = 15$  motos.

6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [2, 46]$ ). L'étendue est donc égale à  $E = 46 - 2 = 44$  motos.

7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles ( i.e.  $IIQ = [15, 22]$ ). L'écart interquartile est donc égal à  $EIQ = 22 - 15 = 7$  motos.

8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [15, 46]$ ). L'écart interdécile est donc égal à  $EID = 46 - 15 = 31$  motos.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée  $EIC$ , est l'amplitude de son intervalle intercentile  $IIC$  borné par son premier et son dernier centiles ( i.e.  $IIC = [2, 46]$ ). L'écart intercentile est donc égal à  $EIC = 46 - 2 = 44$  motos.

**Exercice 42 :**

Le tableau suivant résume la distribution statistique d'une variable statistique discrète  $X$  mesurée en actions :

$x_i$	2	13	21	42	51
$f_i$	0.02	0.37	0.36	0.15	0.1

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;

4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.81 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$x_i$	$f_i$	$F_i^+$
2	0.02	0.02
13	0.37	0.39
21	0.36	0.75
42	0.15	0.9
51	0.1	1
$\Sigma$	1	-

1. La médiane :

La médiane, notée  $M_e$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à une fréquence cumulée croissante de 0.5. On note qu'aucune fréquence cumulée croissante n'est égale à 0.5, la médiane est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. La médiane est donc égale à  $M_e = 21$  actions. Cela signifie que 50% des observations environ sont inférieures à 21 actions et 50% des observations environ sont supérieures à cette valeur.

2. Le premier et le dernier quartiles :

Le premier quartile, noté  $Q_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 25% et 75% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.25. On note qu'aucune fréquence cumulée croissante n'est égale à 0.25, le premier quartile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier quartile est donc égal à  $Q_1 = 13$  actions. Cela signifie que 25% des observations environ sont inférieures à 13 actions et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, noté  $Q_3$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 75% et 25% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.75. On note qu'il y a une fréquence cumulée croissante égale à 0.75, on a donc un intervalle quartile [21, 42]. Par convention, on considère que la valeur du dernier quartile est le centre de cet intervalle. Le dernier quartile est donc égal à  $Q_3 = 31.5$  actions. Cela signifie que 75% des observations environ sont inférieures à 31.5 actions et 25% des observations environ sont supérieures à cette valeur.

3. Le premier et le dernier déciles :

Le premier décile, noté  $D_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 10% et 90% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.10. On note qu'aucune

fréquence cumulée croissante n'est égale à 0.10, le premier décile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier décile est donc égal à  $D_1 = 13$  actions. Cela signifie que 10% des observations environ sont inférieures à 13 actions et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, noté  $D_9$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 90% et 10% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.90. On note qu'il y a une fréquence cumulée croissante égale à 0.90, on a donc un intervalle décile  $[42, 51]$ . Par convention, on considère que la valeur du dernier décile est le centre de cet intervalle. Le dernier décile est donc égal à  $D_9 = 46.5$  actions. Cela signifie que 90% des observations environ sont inférieures à 46.5 actions et 10% des observations environ sont supérieures à cette valeur.

4. Le premier et le dernier centiles :

Le premier centile, noté  $C_1$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 1% et 99% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.01. On note qu'aucune fréquence cumulée croissante n'est égale à 0.01, le premier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le premier centile est donc égal à  $C_1 = 2$  actions. Cela signifie que 1% des observations environ sont inférieures à 2 actions et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, noté  $C_{99}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement 99% et 1% des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.99. On note qu'aucune fréquence cumulée croissante n'est égale à 0.99, le dernier centile est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le dernier centile est donc égal à  $C_{99} = 51$  actions. Cela signifie que 99% des observations environ sont inférieures à 51 actions et 1% des observations environ sont supérieures à cette valeur.

5. Le quantile d'ordre 0.81 :

Le quantile d'ordre 0.81, noté  $q_{0.81}$ , est la modalité  $x_i$  qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.81 et une proportion 0.19 des observations environ. Il est en conséquence associé à une fréquence cumulée croissante de 0.81. On note qu'aucune fréquence cumulée croissante n'est égale à 0.81, le quantile d'ordre 0.81 est donc la modalité ayant la fréquence cumulée croissante immédiatement supérieure à cette valeur. Le quantile d'ordre 0.81 est donc égal à  $q_{0.81} = 42$  actions.

6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [2, 51]$ ). L'étendue est donc égale à  $E = 51 - 2 = 49$  actions.

7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles ( i.e.  $IIQ = [13, 31.5]$ ). L'écart interquartile est donc égal à  $EIQ = 31.5 - 13 = 18.5$  actions.

8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [13, 46.5]$ ). L'écart interdécile est donc égal à  $EID = 46.5 - 13 = 33.5$  actions.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $IIC = [2, 51]$ ). L'écart intercentile est donc égal à  $EIC = 51 - 2 = 49$  actions.

## 2.2.2 Variable quantitative continue

Une variable quantitative continue  $X$  est une variable statistique qui prend en ensemble infini de valeurs numériques regroupées dans des classes notées  $[x_i^-, x_i^+]$ .

### Exercice 43 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  mesurée en gigaoctets :

$[x_i^-, x_i^+]$	$[8, 15[$	$[15, 37[$	$[37, 43[$
$n_i$	2	16	23

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.91 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

### Solution :

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$[x_i^-, x_i^+]$	$n_i$	$f_i$	$F_i^+$
$[8, 15[$	2	0.05	0.05
$[15, 37[$	16	0.39	0.44
$[37, 43[$	23	0.56	1
$\Sigma$	41	1	–

1. La médiane :

La médiane, notée  $M_e$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à la fréquence cumulée croissante 0.5. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul de la médiane se fait en deux étapes :

- Identification de la classe médiane, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.5. Il s'agit dans cet exercice de la classe  $[37, 43[$  ;

- Calcul de la médiane par interpolation linéaire selon la formule suivante :

$$M_e = 37 + \frac{0.5 - 0.44}{1.0 - 0.44} \times (43 - 37) = 37.64 \text{ gigaoctets}$$

Cela signifie que 50% des observations environ sont inférieures à 37.64 gigaoctets et 50% des observations environ sont supérieures à cette valeur.

2. Le premier et le dernier quartiles :

Le premier quartile, notée  $Q_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 25% et une proportion 75% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.25. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier quartile se fait en deux étapes :

- Identification de la classe contenant le premier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.25. Il s'agit dans cet exercice de la classe  $[15, 37[$ ;
- Calcul du premier quartile par interpolation linéaire selon la formule suivante :

$$Q_1 = 15 + \frac{0.25 - 0.05}{0.44 - 0.05} \times (37 - 15) = 26.28 \text{ gigaoctets}$$

Cela signifie que 25% des observations environ sont inférieures à 26.28 gigaoctets et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, notée  $Q_3$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 75% et une proportion 25% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.75. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier quartile se fait en deux étapes :

- Identification de la classe contenant le dernier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.75. Il s'agit dans cet exercice de la classe  $[37, 43[$ ;
- Calcul du dernier quartile par interpolation linéaire selon la formule suivante :

$$Q_3 = 37 + \frac{0.75 - 0.44}{1.0 - 0.44} \times (43 - 37) = 40.32 \text{ gigaoctets}$$

Cela signifie que 75% des observations environ sont inférieures à 40.32 gigaoctets et 25% des observations environ sont supérieures à cette valeur.

3. Le premier et le dernier déciles :

Le premier décile, notée  $D_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 10% et une proportion 90% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.1. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier décile se fait en deux étapes :

- Identification de la classe contenant le premier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.1. Il s'agit dans cet exercice de la classe  $[15, 37[$ ;
- Calcul du premier décile par interpolation linéaire selon la formule suivante :

$$D_1 = 15 + \frac{0.1 - 0.05}{0.44 - 0.05} \times (37 - 15) = 17.82 \text{ gigaoctets}$$

Cela signifie que 10% des observations environ sont inférieures à 17.82 gigaoctets et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, notée  $D_9$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 90% et une proportion 10% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.9. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier décile se fait en deux étapes :

- Identification de la classe contenant le dernier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.9. Il s'agit dans cet exercice de la classe  $[37, 43[$ ;
- Calcul du dernier décile par interpolation linéaire selon la formule suivante :

$$D_9 = 37 + \frac{0.9 - 0.44}{1.0 - 0.44} \times (43 - 37) = 41.93 \text{ gigaoctets}$$

Cela signifie que 90% des observations environ sont inférieures à 41.93 gigaoctets et 10% des observations environ sont supérieures à cette valeur.

#### 4. Le premier et le dernier centiles :

Le premier centile, notée  $C_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 1% et une proportion 99% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.01. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier centile se fait en deux étapes :

- Identification de la classe contenant le premier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.01. Il s'agit dans cet exercice de la classe  $[8, 15[$ ;
- Calcul du premier centile par interpolation linéaire selon la formule suivante :

$$C_1 = 8 + \frac{0.01 - 0}{0.05 - 0} \times (15 - 8) = 9.4 \text{ gigaoctets}$$

Cela signifie que 1% des observations environ sont inférieures à 9.4 gigaoctets et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, notée  $C_{99}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 99% et une proportion 1% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.99. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier centile se fait en deux étapes :

- Identification de la classe contenant le dernier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.99. Il s'agit dans cet exercice de la classe  $[37, 43[$ ;
- Calcul du dernier centile par interpolation linéaire selon la formule suivante :

$$C_{99} = 37 + \frac{0.99 - 0.44}{1.0 - 0.44} \times (43 - 37) = 42.89 \text{ gigaoctets}$$

Cela signifie que 99% des observations environ sont inférieures à 42.89 gigaoctets et 1% des observations environ sont supérieures à cette valeur.

## 5. Le quantile d'ordre 0.91 :

Le quantile d'ordre 0.91, notée  $q_{0.91}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.91 et une proportion 0.09 des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.91. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du quantile d'ordre 0.91 se fait en deux étapes :

- Identification de la classe contenant le quantile d'ordre 0.91, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.91. Il s'agit dans cet exercice de la classe  $[37, 43[$  ;
- Calcul du quantile d'ordre 0.91 par interpolation linéaire selon la formule suivante :

$$q_{0.91} = 37 + \frac{0.91 - 0.44}{1.0 - 0.44} \times (43 - 37) = 42.04 \text{ gigaoctets}$$

Cela signifie qu'une proportion 0.91 des observations environ sont inférieures à 42.04 gigaoctets et une proportion 0.09 des observations environ sont supérieures à cette valeur.

## 6. L'étendue :

L'étendue d'une distribution statistique, notée E, est l'amplitude de son intervalle de variation IV borné par ses valeurs minimale et maximale ( i.e.  $IV = [8, 43]$ ). L'étendue est donc égale à  $E = 43 - 8 = 35$  gigaoctets.

## 7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée EIQ, est l'amplitude de son intervalle interquartile IIQ borné par son premier et son dernier quartiles (i.e.  $IIQ = [26.28, 40.32]$ ). L'écart interquartile est donc égal à  $EIQ = 40.32 - 26.28 = 14.04$  gigaoctets.

## 8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée EID, est l'amplitude de son intervalle interdécile IID borné par son premier et son dernier déciles ( i.e.  $IID = [17.82, 41.93]$ ). L'écart interdécile est donc égal à  $EID = 41.93 - 17.82 = 24.11$  gigaoctets.

## 9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $IIC = [9.4, 42.89]$ ). L'écart intercentile est donc égal à  $EIC = 42.89 - 9.4 = 33.49$  gigaoctets.

**Exercice 44 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue X mesurée en semaines :

$[x_i^-, x_i^+ [$	$[9, 13[$	$[13, 29[$	$[29, 42[$	$[42, 51[$
$n_i$	3	39	26	13

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.74 ;

6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$F_i^+$
$[9, 13[$	3	0.04	0.04
$[13, 29[$	39	0.48	0.52
$[29, 42[$	26	0.32	0.84
$[42, 51[$	13	0.16	1
$\Sigma$	81	1	–

1. La médiane :

La médiane, notée  $M_e$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à la fréquence cumulée croissante 0.5. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul de la médiane se fait en deux étapes :

- Identification de la classe médiane, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.5. Il s'agit dans cet exercice de la classe  $[13, 29[$  ;
- Calcul de la médiane par interpolation linéaire selon la formule suivante :

$$M_e = 13 + \frac{0.5 - 0.04}{0.52 - 0.04} \times (29 - 13) = 28.33 \text{ semaines}$$

Cela signifie que 50% des observations environ sont inférieures à 28.33 semaines et 50% des observations environ sont supérieures à cette valeur.

2. Le premier et le dernier quartiles :

Le premier quartile, notée  $Q_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 25% et une proportion 75% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.25. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier quartile se fait en deux étapes :

- Identification de la classe contenant le premier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.25. Il s'agit dans cet exercice de la classe  $[13, 29[$  ;
- Calcul du premier quartile par interpolation linéaire selon la formule suivante :

$$Q_1 = 13 + \frac{0.25 - 0.04}{0.52 - 0.04} \times (29 - 13) = 20.0 \text{ semaines}$$

Cela signifie que 25% des observations environ sont inférieures à 20.0 semaines et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, notée  $Q_3$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 75% et une proportion 25% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.75. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier quartile se fait en deux étapes :

- Identification de la classe contenant le dernier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.75. Il s'agit dans cet exercice de la classe  $[29, 42[$ ;
- Calcul du dernier quartile par interpolation linéaire selon la formule suivante :

$$Q_3 = 29 + \frac{0.75 - 0.52}{0.84 - 0.52} \times (42 - 29) = 38.34 \text{ semaines}$$

Cela signifie que 75% des observations environ sont inférieures à 38.34 semaines et 25% des observations environ sont supérieures à cette valeur.

### 3. Le premier et le dernier déciles :

Le premier décile, notée  $D_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 10% et une proportion 90% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.1. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du premier décile se fait en deux étapes :

- Identification de la classe contenant le premier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.1. Il s'agit dans cet exercice de la classe  $[13, 29[$ ;
- Calcul du premier décile par interpolation linéaire selon la formule suivante :

$$D_1 = 13 + \frac{0.1 - 0.04}{0.52 - 0.04} \times (29 - 13) = 15.0 \text{ semaines}$$

Cela signifie que 10% des observations environ sont inférieures à 15.0 semaines et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, notée  $D_9$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 90% et une proportion 10% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.9. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du dernier décile se fait en deux étapes :

- Identification de la classe contenant le dernier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.9. Il s'agit dans cet exercice de la classe  $[42, 51[$ ;
- Calcul du dernier décile par interpolation linéaire selon la formule suivante :

$$D_9 = 42 + \frac{0.9 - 0.84}{1.0 - 0.84} \times (51 - 42) = 45.38 \text{ semaines}$$

Cela signifie que 90% des observations environ sont inférieures à 45.38 semaines et 10% des observations environ sont supérieures à cette valeur.

### 4. Le premier et le dernier centiles :

Le premier centile, notée  $C_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 1% et une proportion 99% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.01. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du premier centile se fait en deux étapes :

- Identification de la classe contenant le premier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.01. Il s'agit dans cet exercice de la classe  $[9, 13[$ ;

- Calcul du premier centile par interpolation linéaire selon la formule suivante :

$$C_1 = 9 + \frac{0.01 - 0}{0.04 - 0} \times (13 - 9) = 10.0 \text{ semaines}$$

Cela signifie que 1% des observations environ sont inférieures à 10.0 semaines et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, notée  $C_{99}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 99% et une proportion 1% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.99. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier centile se fait en deux étapes :

- Identification de la classe contenant le dernier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.99. Il s'agit dans cet exercice de la classe  $[42, 51[$ ;
- Calcul du dernier centile par interpolation linéaire selon la formule suivante :

$$C_{99} = 42 + \frac{0.99 - 0.84}{1.0 - 0.84} \times (51 - 42) = 50.44 \text{ semaines}$$

Cela signifie que 99% des observations environ sont inférieures à 50.44 semaines et 1% des observations environ sont supérieures à cette valeur.

#### 5. Le quantile d'ordre 0.74 :

Le quantile d'ordre 0.74, notée  $q_{0.74}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.74 et une proportion 0.26 des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.74. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du quantile d'ordre 0.74 se fait en deux étapes :

- Identification de la classe contenant le quantile d'ordre 0.74, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.74. Il s'agit dans cet exercice de la classe  $[29, 42[$ ;
- Calcul du quantile d'ordre 0.74 par interpolation linéaire selon la formule suivante :

$$q_{0.74} = 29 + \frac{0.74 - 0.52}{0.84 - 0.52} \times (42 - 29) = 37.94 \text{ semaines}$$

Cela signifie qu'une proportion 0.74 des observations environ sont inférieures à 37.94 semaines et une proportion 0.26 des observations environ sont supérieures à cette valeur.

#### 6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [9, 51]$ ). L'étendue est donc égale à  $E = 51 - 9 = 42$  semaines.

#### 7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles (i.e.  $IIQ = [20.0, 38.34]$ ). L'écart interquartile est donc égal à  $EIQ = 38.34 - 20.0 = 18.34$  semaines.

#### 8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [15.0, 45.38]$ ). L'écart interdécile est donc égal à  $EID = 45.38 - 15.0 = 30.38$  semaines.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $IIC = [10.0, 50.44]$ ). L'écart intercentile est donc égal à  $EIC = 50.44 - 10.0 = 40.44$  semaines.

### Exercice 45 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  mesurée en mois :

$[x_i^-, x_i^+[$	$[1, 19[$	$[19, 25[$	$[25, 41[$	$[41, 53[$	$[53, 78[$
$n_i$	6	32	25	19	9

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.55 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

### Solution :

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$[x_i^-, x_i^+[$	$n_i$	$f_i$	$F_i^+$
$[1, 19[$	6	0.07	0.07
$[19, 25[$	32	0.35	0.42
$[25, 41[$	25	0.27	0.69
$[41, 53[$	19	0.21	0.9
$[53, 78[$	9	0.1	1
$\Sigma$	91	1	–

1. La médiane :

La médiane, notée  $M_e$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à la fréquence cumulée croissante 0.5. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul de la médiane se fait en deux étapes :

- Identification de la classe médiane, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.5. Il s'agit dans cet exercice de la classe  $[25, 41[$  ;
- Calcul de la médiane par interpolation linéaire selon la formule suivante :

$$M_e = 25 + \frac{0.5 - 0.42}{0.69 - 0.42} \times (41 - 25) = 29.74 \text{ mois}$$

Cela signifie que 50% des observations environ sont inférieures à 29.74 mois et 50% des observations environ sont supérieures à cette valeur.

## 2. Le premier et le dernier quartiles :

Le premier quartile, notée  $Q_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 25% et une proportion 75% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.25. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier quartile se fait en deux étapes :

- Identification de la classe contenant le premier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.25. Il s'agit dans cet exercice de la classe  $[19, 25[$ ;
- Calcul du premier quartile par interpolation linéaire selon la formule suivante :

$$Q_1 = 19 + \frac{0.25 - 0.07}{0.42 - 0.07} \times (25 - 19) = 22.09 \text{ mois}$$

Cela signifie que 25% des observations environ sont inférieures à 22.09 mois et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, notée  $Q_3$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 75% et une proportion 25% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.75. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier quartile se fait en deux étapes :

- Identification de la classe contenant le dernier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.75. Il s'agit dans cet exercice de la classe  $[41, 53[$ ;
- Calcul du dernier quartile par interpolation linéaire selon la formule suivante :

$$Q_3 = 41 + \frac{0.75 - 0.69}{0.9 - 0.69} \times (53 - 41) = 44.43 \text{ mois}$$

Cela signifie que 75% des observations environ sont inférieures à 44.43 mois et 25% des observations environ sont supérieures à cette valeur.

## 3. Le premier et le dernier déciles :

Le premier décile, notée  $D_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 10% et une proportion 90% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.1. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier décile se fait en deux étapes :

- Identification de la classe contenant le premier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.1. Il s'agit dans cet exercice de la classe  $[19, 25[$ ;
- Calcul du premier décile par interpolation linéaire selon la formule suivante :

$$D_1 = 19 + \frac{0.1 - 0.07}{0.42 - 0.07} \times (25 - 19) = 19.51 \text{ mois}$$

Cela signifie que 10% des observations environ sont inférieures à 19.51 mois et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, notée  $D_9$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 90% et une proportion 10% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.9. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier décile se fait en deux étapes :

- Identification de la classe contenant le dernier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.9. Il s'agit dans cet exercice de la classe  $[41, 53[$ ;
- Comme il y a une fréquence cumulée croissante précisément égale à 0.9, alors le dernier décile est simplement la borne supérieure de la classe disposant de cette fréquence cumulée croissante, en l'occurrence la classe  $[41, 53[$ . Le dernier décile est donc  $D_9 = 53$  mois.

Cela signifie que 90% des observations environ sont inférieures à 53 mois et 10% des observations environ sont supérieures à cette valeur.

#### 4. Le premier et le dernier centiles :

Le premier centile, notée  $C_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 1% et une proportion 99% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.01. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier centile se fait en deux étapes :

- Identification de la classe contenant le premier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.01. Il s'agit dans cet exercice de la classe  $[1, 19[$ ;
- Calcul du premier centile par interpolation linéaire selon la formule suivante :

$$C_1 = 1 + \frac{0.01 - 0}{0.07 - 0} \times (19 - 1) = 3.5714 \text{ mois}$$

Cela signifie que 1% des observations environ sont inférieures à 3.5714 mois et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, notée  $C_{99}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 99% et une proportion 1% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.99. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier centile se fait en deux étapes :

- Identification de la classe contenant le dernier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.99. Il s'agit dans cet exercice de la classe  $[53, 78[$ ;
- Calcul du dernier centile par interpolation linéaire selon la formule suivante :

$$C_{99} = 53 + \frac{0.99 - 0.9}{1.0 - 0.9} \times (78 - 53) = 75.5 \text{ mois}$$

Cela signifie que 99% des observations environ sont inférieures à 75.5 mois et 1% des observations environ sont supérieures à cette valeur.

## 5. Le quantile d'ordre 0.55 :

Le quantile d'ordre 0.55, notée  $q_{0.55}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.55 et une proportion 0.45 des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.55. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du quantile d'ordre 0.55 se fait en deux étapes :

- Identification de la classe contenant le quantile d'ordre 0.55, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.55. Il s'agit dans cet exercice de la classe  $[25, 41[$  ;
- Calcul du quantile d'ordre 0.55 par interpolation linéaire selon la formule suivante :

$$q_{0.55} = 25 + \frac{0.55 - 0.42}{0.69 - 0.42} \times (41 - 25) = 32.7 \text{ mois}$$

Cela signifie qu'une proportion 0.55 des observations environ sont inférieures à 32.7 mois et une proportion 0.45 des observations environ sont supérieures à cette valeur.

## 6. L'étendue :

L'étendue d'une distribution statistique, notée E, est l'amplitude de son intervalle de variation IV borné par ses valeurs minimale et maximale ( i.e.  $IV = [1, 78]$ ). L'étendue est donc égale à  $E = 78 - 1 = 77$  mois.

## 7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée EIQ, est l'amplitude de son intervalle interquartile IIQ borné par son premier et son dernier quartiles (i.e.  $IIQ = [22.09, 44.43]$ ). L'écart interquartile est donc égal à  $EIQ = 44.43 - 22.09 = 22.34$  mois.

## 8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée EID, est l'amplitude de son intervalle interdécile IID borné par son premier et son dernier déciles ( i.e.  $IID = [19.51, 53]$ ). L'écart interdécile est donc égal à  $EID = 53 - 19.51 = 33.49$  mois.

## 9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $IIC = [3.5714, 75.5]$ ). L'écart intercentile est donc égal à  $EIC = 75.5 - 3.5714 = 71.93$  mois.

**Exercice 46 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue X mesurée en dollars :

$[x_i^-, x_i^+[$	$[2, 13[$	$[13, 28[$	$[28, 47[$
$f_i$	0.12	0.51	0.37

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.16 ;

6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$[x_i^-, x_i^+ [$	$f_i$	$F_i^+$
$[2, 13[$	0.12	0.12
$[13, 28[$	0.51	0.63
$[28, 47[$	0.37	1
$\Sigma$	1	–

1. La médiane :

La médiane, notée  $M_e$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à la fréquence cumulée croissante 0.5. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul de la médiane se fait en deux étapes :

- Identification de la classe médiane, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.5. Il s'agit dans cet exercice de la classe  $[13, 28[$  ;
- Calcul de la médiane par interpolation linéaire selon la formule suivante :

$$M_e = 13 + \frac{0.5 - 0.12}{0.63 - 0.12} \times (28 - 13) = 24.18 \text{ dollars}$$

Cela signifie que 50% des observations environ sont inférieures à 24.18 dollars et 50% des observations environ sont supérieures à cette valeur.

2. Le premier et le dernier quartiles :

Le premier quartile, notée  $Q_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 25% et une proportion 75% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.25. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du premier quartile se fait en deux étapes :

- Identification de la classe contenant le premier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.25. Il s'agit dans cet exercice de la classe  $[13, 28[$  ;
- Calcul du premier quartile par interpolation linéaire selon la formule suivante :

$$Q_1 = 13 + \frac{0.25 - 0.12}{0.63 - 0.12} \times (28 - 13) = 16.82 \text{ dollars}$$

Cela signifie que 25% des observations environ sont inférieures à 16.82 dollars et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, notée  $Q_3$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 75% et une proportion 25% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.75. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du dernier quartile se fait en deux étapes :

- Identification de la classe contenant le dernier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.75. Il s'agit dans cet exercice de la classe  $[28, 47[$ ;
- Calcul du dernier quartile par interpolation linéaire selon la formule suivante :

$$Q_3 = 28 + \frac{0.75 - 0.63}{1.0 - 0.63} \times (47 - 28) = 34.16 \text{ dollars}$$

Cela signifie que 75% des observations environ sont inférieures à 34.16 dollars et 25% des observations environ sont supérieures à cette valeur.

### 3. Le premier et le dernier déciles :

Le premier décile, notée  $D_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 10% et une proportion 90% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.1. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du premier décile se fait en deux étapes :

- Identification de la classe contenant le premier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.1. Il s'agit dans cet exercice de la classe  $[2, 13[$ ;
- Calcul du premier décile par interpolation linéaire selon la formule suivante :

$$D_1 = 2 + \frac{0.1 - 0}{0.12 - 0} \times (13 - 2) = 11.1667 \text{ dollars}$$

Cela signifie que 10% des observations environ sont inférieures à 11.1667 dollars et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, notée  $D_9$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 90% et une proportion 10% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.9. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du dernier décile se fait en deux étapes :

- Identification de la classe contenant le dernier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.9. Il s'agit dans cet exercice de la classe  $[28, 47[$ ;
- Calcul du dernier décile par interpolation linéaire selon la formule suivante :

$$D_9 = 28 + \frac{0.9 - 0.63}{1.0 - 0.63} \times (47 - 28) = 41.86 \text{ dollars}$$

Cela signifie que 90% des observations environ sont inférieures à 41.86 dollars et 10% des observations environ sont supérieures à cette valeur.

### 4. Le premier et le dernier centiles :

Le premier centile, notée  $C_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 1% et une proportion 99% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.01. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul du premier centile se fait en deux étapes :

- Identification de la classe contenant le premier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.01. Il s'agit dans cet exercice de la classe  $[2, 13[$ ;

- Calcul du premier centile par interpolation linéaire selon la formule suivante :

$$C_1 = 2 + \frac{0.01 - 0}{0.12 - 0} \times (13 - 2) = 2.9167 \text{ dollars}$$

Cela signifie que 1% des observations environ sont inférieures à 2.9167 dollars et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, notée  $C_{99}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 99% et une proportion 1% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.99. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier centile se fait en deux étapes :

- Identification de la classe contenant le dernier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.99. Il s'agit dans cet exercice de la classe [28, 47[;
- Calcul du dernier centile par interpolation linéaire selon la formule suivante :

$$C_{99} = 28 + \frac{0.99 - 0.63}{1.0 - 0.63} \times (47 - 28) = 46.49 \text{ dollars}$$

Cela signifie que 99% des observations environ sont inférieures à 46.49 dollars et 1% des observations environ sont supérieures à cette valeur.

#### 5. Le quantile d'ordre 0.16 :

Le quantile d'ordre 0.16, notée  $q_{0.16}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.16 et une proportion 0.84 des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.16. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du quantile d'ordre 0.16 se fait en deux étapes :

- Identification de la classe contenant le quantile d'ordre 0.16, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.16. Il s'agit dans cet exercice de la classe [13, 28[;
- Calcul du quantile d'ordre 0.16 par interpolation linéaire selon la formule suivante :

$$q_{0.16} = 13 + \frac{0.16 - 0.12}{0.63 - 0.12} \times (28 - 13) = 14.18 \text{ dollars}$$

Cela signifie qu'une proportion 0.16 des observations environ sont inférieures à 14.18 dollars et une proportion 0.84 des observations environ sont supérieures à cette valeur.

#### 6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [2, 47]$ ). L'étendue est donc égale à  $E = 47 - 2 = 45$  dollars.

#### 7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles (i.e.  $IIQ = [16.82, 34.16]$ ). L'écart interquartile est donc égal à  $EIQ = 34.16 - 16.82 = 17.34$  dollars.

#### 8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [11.1667, 41.86]$ ). L'écart interdécile est donc égal à  $EID = 41.86 - 11.1667 = 30.69$  dollars.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $IIC = [2.9167, 46.49]$ ). L'écart intercentile est donc égal à  $EIC = 46.49 - 2.9167 = 43.57$  dollars.

### Exercice 47 :

Le tableau suivant résume la distribution statistique d'une variable statistique continue  $X$  mesurée en gigaoctets :

$[x_i^-, x_i^+[$	$[9, 15[$	$[15, 37[$	$[37, 45[$	$[45, 55[$
$f_i$	0.05	0.19	0.49	0.27

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.32 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

### Solution :

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$[x_i^-, x_i^+[$	$f_i$	$F_i^+$
$[9, 15[$	0.05	0.05
$[15, 37[$	0.19	0.24
$[37, 45[$	0.49	0.73
$[45, 55[$	0.27	1
$\Sigma$	1	–

1. La médiane :

La médiane, notée  $M_e$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à la fréquence cumulée croissante 0.5. Dans le cas d'une variable quantitative continue, comme c'est la cas dans cet exercice, le calcul de la médiane se fait en deux étapes :

- Identification de la classe médiane, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.5. Il s'agit dans cet exercice de la classe  $[37, 45[$  ;
- Calcul de la médiane par interpolation linéaire selon la formule suivante :

$$M_e = 37 + \frac{0.5 - 0.24}{0.73 - 0.24} \times (45 - 37) = 41.24 \text{ gigaoctets}$$

Cela signifie que 50% des observations environ sont inférieures à 41.24 gigaoctets et 50% des observations environ sont supérieures à cette valeur.

## 2. Le premier et le dernier quartiles :

Le premier quartile, notée  $Q_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 25% et une proportion 75% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.25. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier quartile se fait en deux étapes :

- Identification de la classe contenant le premier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.25. Il s'agit dans cet exercice de la classe  $[37, 45[$ ;
- Calcul du premier quartile par interpolation linéaire selon la formule suivante :

$$Q_1 = 37 + \frac{0.25 - 0.24}{0.73 - 0.24} \times (45 - 37) = 37.16 \text{ gigaoctets}$$

Cela signifie que 25% des observations environ sont inférieures à 37.16 gigaoctets et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, notée  $Q_3$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 75% et une proportion 25% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.75. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier quartile se fait en deux étapes :

- Identification de la classe contenant le dernier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.75. Il s'agit dans cet exercice de la classe  $[45, 55[$ ;
- Calcul du dernier quartile par interpolation linéaire selon la formule suivante :

$$Q_3 = 45 + \frac{0.75 - 0.73}{1.0 - 0.73} \times (55 - 45) = 45.74 \text{ gigaoctets}$$

Cela signifie que 75% des observations environ sont inférieures à 45.74 gigaoctets et 25% des observations environ sont supérieures à cette valeur.

## 3. Le premier et le dernier déciles :

Le premier décile, notée  $D_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 10% et une proportion 90% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.1. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier décile se fait en deux étapes :

- Identification de la classe contenant le premier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.1. Il s'agit dans cet exercice de la classe  $[15, 37[$ ;
- Calcul du premier décile par interpolation linéaire selon la formule suivante :

$$D_1 = 15 + \frac{0.1 - 0.05}{0.24 - 0.05} \times (37 - 15) = 20.79 \text{ gigaoctets}$$

Cela signifie que 10% des observations environ sont inférieures à 20.79 gigaoctets et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, notée  $D_9$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 90% et une proportion 10% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.9. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier décile se fait en deux étapes :

- Identification de la classe contenant le dernier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.9. Il s'agit dans cet exercice de la classe  $[45, 55[$ ;
- Calcul du dernier décile par interpolation linéaire selon la formule suivante :

$$D_9 = 45 + \frac{0.9 - 0.73}{1.0 - 0.73} \times (55 - 45) = 51.3 \text{ gigaoctets}$$

Cela signifie que 90% des observations environ sont inférieures à 51.3 gigaoctets et 10% des observations environ sont supérieures à cette valeur.

#### 4. Le premier et le dernier centiles :

Le premier centile, notée  $C_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 1% et une proportion 99% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.01. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier centile se fait en deux étapes :

- Identification de la classe contenant le premier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.01. Il s'agit dans cet exercice de la classe  $[9, 15[$ ;
- Calcul du premier centile par interpolation linéaire selon la formule suivante :

$$C_1 = 9 + \frac{0.01 - 0}{0.05 - 0} \times (15 - 9) = 10.2 \text{ gigaoctets}$$

Cela signifie que 1% des observations environ sont inférieures à 10.2 gigaoctets et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, notée  $C_{99}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 99% et une proportion 1% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.99. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier centile se fait en deux étapes :

- Identification de la classe contenant le dernier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.99. Il s'agit dans cet exercice de la classe  $[45, 55[$ ;
- Calcul du dernier centile par interpolation linéaire selon la formule suivante :

$$C_{99} = 45 + \frac{0.99 - 0.73}{1.0 - 0.73} \times (55 - 45) = 54.63 \text{ gigaoctets}$$

Cela signifie que 99% des observations environ sont inférieures à 54.63 gigaoctets et 1% des observations environ sont supérieures à cette valeur.

#### 5. Le quantile d'ordre 0.32 :

Le quantile d'ordre 0.32, notée  $q_{0.32}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.32 et une proportion 0.68 des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.32. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du quantile d'ordre 0.32 se fait en deux étapes :

- Identification de la classe contenant le quantile d'ordre 0.32, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.32. Il s'agit dans cet exercice de la classe  $[37, 45[$ ;

- Calcul du quantile d'ordre 0.32 par interpolation linéaire selon la formule suivante :

$$q_{0.32} = 37 + \frac{0.32 - 0.24}{0.73 - 0.24} \times (45 - 37) = 38.31 \text{ gigaoctets}$$

Cela signifie qu'une proportion 0.32 des observations environ sont inférieures à 38.31 gigaoctets et une proportion 0.68 des observations environ sont supérieures à cette valeur.

6. L'étendue :

L'étendue d'une distribution statistique, notée E, est l'amplitude de son intervalle de variation IV borné par ses valeurs minimale et maximale ( i.e. IV = [9, 55]). L'étendue est donc égale à E = 55 - 9 = 46 gigaoctets.

7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée EIQ, est l'amplitude de son intervalle interquartile IIQ borné par son premier et son dernier quartiles (i.e. IIQ = [37.16, 45.74]). L'écart interquartile est donc égal à EIQ = 45.74 - 37.16 = 8.58 gigaoctets.

8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée EID, est l'amplitude de son intervalle interdécile IID borné par son premier et son dernier déciles ( i.e. IID = [20.79, 51.3]). L'écart interdécile est donc égal à EID = 51.3 - 20.79 = 30.51 gigaoctets.

9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e. IIC = [10.2, 54.63]). L'écart intercentile est donc égal à EIC = 54.63 - 10.2 = 44.43 gigaoctets.

**Exercice 48 :**

Le tableau suivant résume la distribution statistique d'une variable statistique continue X mesurée en euros :

$[x_i^-, x_i^+]$	[1, 11[	[11, 29[	[29, 42[	[42, 52[	[52, 73[
$f_i$	0.05	0.36	0.31	0.26	0.02

1. Calculer la médiane ;
2. Calculer le premier et le dernier quartiles ;
3. Calculer le premier et le dernier déciles ;
4. Calculer le premier et le dernier centiles ;
5. Calculer le quantile d'ordre 0.76 ;
6. Calculer l'étendue ;
7. Calculer l'écart interquartile ;
8. Calculer l'écart interdécile ;
9. Calculer l'écart intercentile.

**Solution :**

Le calcul des quantiles est fondé, entre autres, sur les fréquences cumulées croissantes  $F_i^+$  qu'il convient de calculer dans un premier temps :

$[x_i^-, x_i^+ [$	$f_i$	$F_i^+$
$[1, 11 [$	0.05	0.05
$[11, 29 [$	0.36	0.41
$[29, 42 [$	0.31	0.72
$[42, 52 [$	0.26	0.98
$[52, 73 [$	0.02	1
$\Sigma$	1	–

## 1. La médiane :

La médiane, notée  $M_e$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant chacune la moitié des observations environ. Elle est en conséquence associée à la fréquence cumulée croissante 0.5. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul de la médiane se fait en deux étapes :

- Identification de la classe médiane, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.5. Il s'agit dans cet exercice de la classe  $[29, 42[$ ;
- Calcul de la médiane par interpolation linéaire selon la formule suivante :

$$M_e = 29 + \frac{0.5 - 0.41}{0.72 - 0.41} \times (42 - 29) = 32.77 \text{ euros}$$

Cela signifie que 50% des observations environ sont inférieures à 32.77 euros et 50% des observations environ sont supérieures à cette valeur.

## 2. Le premier et le dernier quartiles :

Le premier quartile, notée  $Q_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 25% et une proportion 75% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.25. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier quartile se fait en deux étapes :

- Identification de la classe contenant le premier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.25. Il s'agit dans cet exercice de la classe  $[11, 29[$ ;
- Calcul du premier quartile par interpolation linéaire selon la formule suivante :

$$Q_1 = 11 + \frac{0.25 - 0.05}{0.41 - 0.05} \times (29 - 11) = 21.0 \text{ euros}$$

Cela signifie que 25% des observations environ sont inférieures à 21.0 euros et 75% des observations environ sont supérieures à cette valeur.

Le dernier quartile, notée  $Q_3$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 75% et une proportion 25% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.75. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier quartile se fait en deux étapes :

- Identification de la classe contenant le dernier quartile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.75. Il s'agit dans cet exercice de la classe  $[42, 52[$ ;
- Calcul du dernier quartile par interpolation linéaire selon la formule suivante :

$$Q_3 = 42 + \frac{0.75 - 0.72}{0.98 - 0.72} \times (52 - 42) = 43.15 \text{ euros}$$

Cela signifie que 75% des observations environ sont inférieures à 43.15 euros et 25% des observations environ sont supérieures à cette valeur.

### 3. Le premier et le dernier déciles :

Le premier décile, notée  $D_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 10% et une proportion 90% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.1. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier décile se fait en deux étapes :

- Identification de la classe contenant le premier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.1. Il s'agit dans cet exercice de la classe  $[11, 29[$ ;
- Calcul du premier décile par interpolation linéaire selon la formule suivante :

$$D_1 = 11 + \frac{0.1 - 0.05}{0.41 - 0.05} \times (29 - 11) = 13.5 \text{ euros}$$

Cela signifie que 10% des observations environ sont inférieures à 13.5 euros et 90% des observations environ sont supérieures à cette valeur.

Le dernier décile, notée  $D_9$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 90% et une proportion 10% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.9. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier décile se fait en deux étapes :

- Identification de la classe contenant le dernier décile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.9. Il s'agit dans cet exercice de la classe  $[42, 52[$ ;
- Calcul du dernier décile par interpolation linéaire selon la formule suivante :

$$D_9 = 42 + \frac{0.9 - 0.72}{0.98 - 0.72} \times (52 - 42) = 48.92 \text{ euros}$$

Cela signifie que 90% des observations environ sont inférieures à 48.92 euros et 10% des observations environ sont supérieures à cette valeur.

### 4. Le premier et le dernier centiles :

Le premier centile, notée  $C_1$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 1% et une proportion 99% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.01. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du premier centile se fait en deux étapes :

- Identification de la classe contenant le premier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.01. Il s'agit dans cet exercice de la classe  $[1, 11[$ ;
- Calcul du premier centile par interpolation linéaire selon la formule suivante :

$$C_1 = 1 + \frac{0.01 - 0}{0.05 - 0} \times (11 - 1) = 3.0 \text{ euros}$$

Cela signifie que 1% des observations environ sont inférieures à 3.0 euros et 99% des observations environ sont supérieures à cette valeur.

Le dernier centile, notée  $C_{99}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 99% et une proportion 1% des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.99. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du dernier centile se fait en deux étapes :

- Identification de la classe contenant le dernier centile, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.99. Il s'agit dans cet exercice de la classe  $[52, 73[$ ;
- Calcul du dernier centile par interpolation linéaire selon la formule suivante :

$$C_{99} = 52 + \frac{0.99 - 0.98}{1.0 - 0.98} \times (73 - 52) = 62.5 \text{ euros}$$

Cela signifie que 99% des observations environ sont inférieures à 62.5 euros et 1% des observations environ sont supérieures à cette valeur.

#### 5. Le quantile d'ordre 0.76 :

Le quantile d'ordre 0.76, notée  $q_{0.76}$ , est la valeur qui partage la série statistique ordonnée en deux parties contenant respectivement une proportion 0.76 et une proportion 0.24 des observations environ. Il est en conséquence associé à la fréquence cumulée croissante 0.76. Dans le cas d'une variable quantitative continue, comme c'est le cas dans cet exercice, le calcul du quantile d'ordre 0.76 se fait en deux étapes :

- Identification de la classe contenant le quantile d'ordre 0.76, celle qui a la fréquence cumulée croissante immédiatement supérieure ou égale à 0.76. Il s'agit dans cet exercice de la classe  $[42, 52[$ ;
- Calcul du quantile d'ordre 0.76 par interpolation linéaire selon la formule suivante :

$$q_{0.76} = 42 + \frac{0.76 - 0.72}{0.98 - 0.72} \times (52 - 42) = 43.54 \text{ euros}$$

Cela signifie qu'une proportion 0.76 des observations environ sont inférieures à 43.54 euros et une proportion 0.24 des observations environ sont supérieures à cette valeur.

#### 6. L'étendue :

L'étendue d'une distribution statistique, notée  $E$ , est l'amplitude de son intervalle de variation  $IV$  borné par ses valeurs minimale et maximale ( i.e.  $IV = [1, 73]$ ). L'étendue est donc égale à  $E = 73 - 1 = 72$  euros.

#### 7. L'écart interquartile :

L'écart interquartile d'une distribution statistique, notée  $EIQ$ , est l'amplitude de son intervalle interquartile  $IIQ$  borné par son premier et son dernier quartiles (i.e.  $IIQ = [21.0, 43.15]$ ). L'écart interquartile est donc égal à  $EIQ = 43.15 - 21.0 = 22.15$  euros.

#### 8. L'écart interdécile :

L'écart interdécile d'une distribution statistique, notée  $EID$ , est l'amplitude de son intervalle interdécile  $IID$  borné par son premier et son dernier déciles ( i.e.  $IID = [13.5, 48.92]$ ). L'écart interdécile est donc égal à  $EID = 48.92 - 13.5 = 35.42$  euros.

## 9. L'écart intercentile :

L'écart intercentile d'une distribution statistique, notée EIC, est l'amplitude de son intervalle intercentile IIC borné par son premier et son dernier centiles ( i.e.  $IIC = [3.0, 62.5]$ ). L'écart intercentile est donc égal à  $EIC = 62.5 - 3.0 = 59.5$  euros.

## Conclusion

Dans ce deuxième chapitre, ont été traités des exercices sur des mesures numériques basées sur les moments, telles que les mesures de tendance centrale, les mesures de dispersion et les mesures de forme, et des mesures numériques basées sur les quantiles, telles que la médiane, les quantiles et les écarts interquantiles, dans le cas d'une seule variable statistique.

# Conclusion

Dans le chapitre introductif à la statistique ont été résolus des exercices relatifs au tableau statistique et aux représentations graphiques. Dans le chapitre suivant, seront résolus des exercices sur des mesures de tendance centrale, des mesures de dispersion et des mesures de forme.

Dans le deuxième chapitre, ont été résolus des exercices sur des mesures numériques basées sur les moments, telles que les mesures de tendance centrale, les mesures de dispersion et les mesures de forme, et des mesures numériques basées sur les quantiles, telles que la médiane, les quantiles et les écarts interquantiles, dans le cas d'une seule variable statistique.



# Bibliographie

- [1] D.R. Anderson, D.J. Sweeney, T.A. Williams, J.D. Camm, and J.J. Cochran. *Essentials of Statistics for Business and Economics*. Cengage Learning, 2017.
- [2] S. Baggio, S. Rothen, and S. Deline. *Statistique descriptive : en 20 fiches*. Sup en poche psycho. De Boeck supérieur, 2022.
- [3] J.I. Barragués, A. Morais, and J. Guisasola. *Probability and Statistics : A Didactic Introduction*. CRC Press, 2016.
- [4] D. Beaven. *Probability and Statistics for Engineering and Sciences*. NY RESEARCH Press, 2022.
- [5] R. Boutti. *Statistique descriptive décisionnelle : 300 QCM et exercices corrigés*. Best practices. R. Boutti, 2007.
- [6] F. Conway. *Descriptive Statistics*. Creative Media Partners, LLC, 2021.
- [7] B. Delmas. *Statistique descriptive pour l'économie et la gestion*. Guides pratiques. Presses Universitaires du Septentrion, 2009.
- [8] P.G. Dixit. *Descriptive Statistics, Probability And Probability Distributions - I*. Nirali Prakashan, 2016.
- [9] G. Dufrenot, F. Bertrand, C. Derquenne, F. Jawadi, and M. Maumy. *Statistiques pour l'économie et la gestion*. De Boeck Supérieur, 2021.
- [10] J.F. Healey and C. Donoghue. *Statistics : A Tool for Social Research and Data Analysis*. Cengage Learning, 2020.
- [11] Z.C. Holcomb. *Fundamentals of Descriptive Statistics*. Pyrczak Pub., 1998.
- [12] J. Hubler. *Statistique descriptive appliquée à la gestion et à l'économie*. Lexifac. Économie. Bréal, 2007.
- [13] C. Hurlin and V. Mignon. *Statistique et probabilités en économie-gestion*. Dunod, 2018.
- [14] M.E. Johnson. *Statistics Fundamentals*. Amazon Digital Services LLC - Kdp, 2020.
- [15] J.S. Jones. *Exploratory and Descriptive Statistics*. The SAGE Quantitative Research Kit. SAGE Publications, 2022.
- [16] M.O. Lacort. *Descriptive and Inferential Statistics - Summaries of theory and Exercises solved*. Lulu.com, 2014.
- [17] V. Leboucher and É. Cépaduès. *Introduction À La Statistique Descriptive, 3E Édition*. Cépaduès, 2015.

- [18] B. Legros. *Mini manuel de Statistiques et probabilités en économie-gestion - 2e éd.* Mini Manuel. Dunod, 2016.
- [19] M. Lethielleux. *Statistique descriptive - 8e éd. : en 27 fiches.* Express. Dunod, 2016.
- [20] D.A. Lind, W.G. Marchal, S.A. Wathen, and C.A. Waite. *Statistiques pour la gestion.* McGraw-Hill Education, 2017.
- [21] C.P. Lopez. *Descriptive statistics with R.* Amazon Italia Logistica, 2022.
- [22] N. Maxwell. *Introductory Statistics.* WILLFORD Press, 2022.
- [23] J.L. Monino. *Statistique descriptive : QCM et exercices corrigés, 4 sujets d'examen corrigés, avec rappels de cours.* TD (Paris). Dunod, 2017.
- [24] A. Nurbel and I. Ahamada. *Les méthodes de la statistique descriptive : 65 exercices corrigés. Tendances centrale, dispersion et forme, concentration, ajustement, régression et corrélation.* Number vol. 1 in EPU, Éditions Publibook université. Publibook/Société des écrivains, 2007.
- [25] B. Py. *Statistique descriptive : nouvelle méthode pour bien comprendre et réussir.* Economica, 2007.
- [26] R. Rivera. *Principles of Managerial Statistics and Data Science.* Wiley, 2020.
- [27] D. Selvamuthu and D. Das. *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control.* Springer Nature Singapore, 2018.
- [28] H. Singh. *Statistics for Machine Learning : Implement Statistical methods used in Machine Learning using Python (English Edition).* ITpro collection. BPB Publications, 2021.
- [29] R. Stinerock. *Statistics with R : A Beginner s Guide.* SAGE Publications, 2022.
- [30] C. Thrane. *Doing Statistical Analysis : A Student's Guide to Quantitative Research.* Taylor & Francis, 2022.
- [31] O.W. Winkler. *Interpreting Economic and Social Data : A Foundation of Descriptive Statistics.* Mathematics and Statistics. Springer Berlin Heidelberg, 2009.

# STATISTIQUE DESCRIPTIVE UNIVARIÉE

## Exercices corrigés

Ce manuel de Statistique descriptive univariée est une collection d'exercices corrigés de statistique descriptive destinée aux étudiants de l'enseignement supérieur poursuivant leurs études dans des disciplines scientifiques (mathématiques, physique, biologie, sciences économiques, gestion...). L'objectif étant d'appliquer les concepts de la statistique sur différents jeux de données. En particulier, dans le chapitre introductif à la statistique sont résolus des exercices relatifs à deux outils indispensables dans la description et la synthèse des données statistiques, à savoir le tableau statistique et les représentations graphiques. Dans le chapitre 2, sont traités des exercices sur des mesures numériques basées sur les moments (moyenne arithmétique, variance, écart-type, coefficient de variation, coefficient d'asymétrie et coefficient d'aplatissement) et des mesures numériques basées sur les quantiles (médiane, quartiles, déciles, centiles, écart interquartile, écart interdécile et écart intercentile) dans le cas d'une seule variable statistique. Le cas bivarié et le cas multivarié feront l'objet d'éditions futures.



**Cherif El Msiyah** est docteur en Actuariat et Finance de marchés de l'université d'Orléans et enseignant chercheur à l'École Nationale de Commerce et de Gestion à Meknès. Il enseigne l'économie et la finance depuis une douzaine d'années. Il a publié un certain nombre d'articles sur l'application de la statistique en économie, en finance et en assurance.



**Jaouad Madkour** est docteur en Sciences économiques de l'université d'Orléans et enseignant chercheur à la faculté des sciences juridiques, économiques et sociales de Tanger dans laquelle il enseigne les méthodes quantitatives depuis une dizaine d'années. Il a publié un certain nombre d'articles sur l'application de la statistique en économie et en finance.

ISBN 978-9920-42-627-5

