3. Tableaux et représentation graphique des données

3.1. Terminologie

- Caractère ou variable statistique

C’est la caractéristique commune à l’ensemble des individus d’une étude, la valeur de celle-ci varie d’un individu à un autre.

Exemples : Le poids, la taille. Des étudiants qui ont réussi leur master peuvent être décrits selon les notes qu’ils ont obtenues, la mention, l’établissement d’enseignement supérieur, l’âge, etc.

- Modalités d’un caractère ou d’une variable statistique

Ce sont les différentes valeurs que peut prendre une variable statistique. Les différentes modalités du caractère doivent être définies sans ambiguïté de telle façon qu’une observation corresponde à une modalité et à une seule.

Exemples : les modalités de la variable « statut matrimonial » : célibataire, marié, veuf, divorcé ; les modalités de la variable sexe : féminin, masculin ; les modalités de la variable « mention » : passable, assez bien, bien, très bien, etc.

- Caractère quantitatif

Un caractère est quantitatif lorsque ses modalités sont mesurées par un nombre.

Exemples : le chiffre d’affaires d’une entreprise, le nombre d’enfants par ménage, le poids, la vitesse des automobiles, etc.

Il existe deux types de caractères quantitatifs.

- Les caractères quantitatifs discrets ou discontinus

Un caractère quantitatif est discret s’il ne prend que des valeurs isolées et des valeurs numériques entières.

Exemple : le nombre d’enfants par femme ne peut être que 0 ou 1 ou 2 ou 3 ou 4, … mais ne peut jamais être une valeur strictement comprise entre 0 et 1, ou 1 et 2, ou 2 et 3… le nombre de pièces d’un logement ne peut être que de 1 ou 2 ou 3, …

C’est pourquoi on dit que la valeur de la variable discrète est exacte.

- Les caractères quantitatifs continus

Un caractère est continu s’il peut prendre n’importe quelle valeur dans un intervalle. Les caractères quantitatifs continus sont des caractères numériques ayant un nombre infini de valeurs entre deux valeurs.

Exemple : la taille d’un individu peut être comprise entre 1,72 et 1,73 m sachant que ces valeurs ont été arrondies.

- Caractère qualitatif

Un caractère est qualitatif lorsque ses modalités sont désignées par des noms. Exemples : le sexe (masculin, féminin), la couleur des yeux (marron, brun, bleu, vert, etc.)

Il est possible de transformer un caractère quantitatif en caractère qualitatif. Considérons les notes sur 20 à une épreuve de statistique, on pourrait adopter la correspondance suivante :

A =] 16 à 20] : très bon B =] 12 à 16] : bon C =] 8 à 12] : moyen D =] 4 à 8] : faible

E=] 0 à 4] : très faible

- Caractère ordinal

C'est un caractère dans lequel les modalités peuvent être classés dans un ordre spécifique. Celles présentent une hiérarchie entre elles, permettant de les classer de la plus petite à la plus grande ou l'inverse.

Exemple : très satisfait, satisfait, assez satisfait, peu satisfait, pas du tout satisfait

- Caractère qualitatif binaire

C'est un caractère qui ne compte que deux modalités possibles.

Exemple : sexe (homme ou femme)

Application : Indiquer dans chaque cas si la variable est qualitative nominale, qualitative ordinale, quantitative discrète ou quantitative continue.

1. Le nombre d'enfants dans une famille
2. La couleur des yeux
3. Le revenu annuel en FCFA
4. La marque de voiture
5. L'âge d'une personne
6. La note sur 20 d'un examen
7. Le groupe sanguin
8. La température en degré Celsius
9. Le nombre de livre dans une librairie
10. Le niveau de satisfaction (très satisfait, satisfait, peu satisfait, pas du tout satisfait)

- Effectif

L'effectif total **n** d'une population est le nombre d'individus qui composent cette population.

L'effectif associé à une valeur de la variable **ni** représente le nombre de fois où la valeur apparaît.

L'effectif est alors noté **Σni**.

- Fréquence

La fréquence relative **fi** d'une modalité **i** du caractère **X** est la proportion d'individus de la population qui présentent la modalité **i**, ainsi : **fi = ni/n**

La somme des fréquences relatives est égale à **1** ou à **100%**.

Pour des raisons de commodité, les **fi** sont souvent donnés en pourcentage (100%), ainsi fi = ni/n × 100%

3.2. Tableau de données

- Cas d'une variable qualitative

Une enquête effectuée auprès de 150 étudiants de l'UGB sur les prestations des restaurants universitaires a donné les résultats suivants :

Modalités

xi = variable = niveau de satisfaction	**ni = effectif**	**fi%= fréquence**
Très satisfait	21	14%
Satisfait	35	23,3%
Peu satisfait	46	30,7%
Pas satisfait	48	32%
	n ou Σni = 150	**Σfi = 100%**

- Cas d’une variable quantitative discrète

Répartition de 100 femmes selon le nombre d’enfants.

Modalités

xi = variable = nombre d’enfants	**ni**
1	10
2	15
3	15
4	30
5	10
6	12
7	5
8	3

- Classement des observations et regroupement par classes

Les observations sont généralement regroupées par classes, à chaque classe on associe n_i, nombre d'individus de la classe i. Lorsqu'une variable est continue, chaque classe est représentée par un intervalle.

Application : La distribution de la variable X « temps de trajet en minutes du domicile au travail » d'une population de 600 individus pourrait être représentée par le tableau suivant :

Modalités

X_i = variable = temps de trajet en minute	**n_i = effectif**	**f_i% = fréquence**
[0 - 20[	165	27,5
[20 - 40[	278	46,33
[40 - 50[	93	15,5
[50 - 60[	64	10,67
	n ou Σn_i = 600	**Σf_i = 100%**

- Pour les caractères qualitatifs : le diagramme le plus utilisé est en secteurs circulaires et désigné souvent par « camembert ». Le principe consiste à représenter des données par des diagrammes dont les différentes parties ont des aires proportionnelles aux effectifs.

Exemple : Répartition des élèves d'un lycée selon la filière où ils sont inscrits.

Tableau

xi	S	G	L
fi%	50	20	30

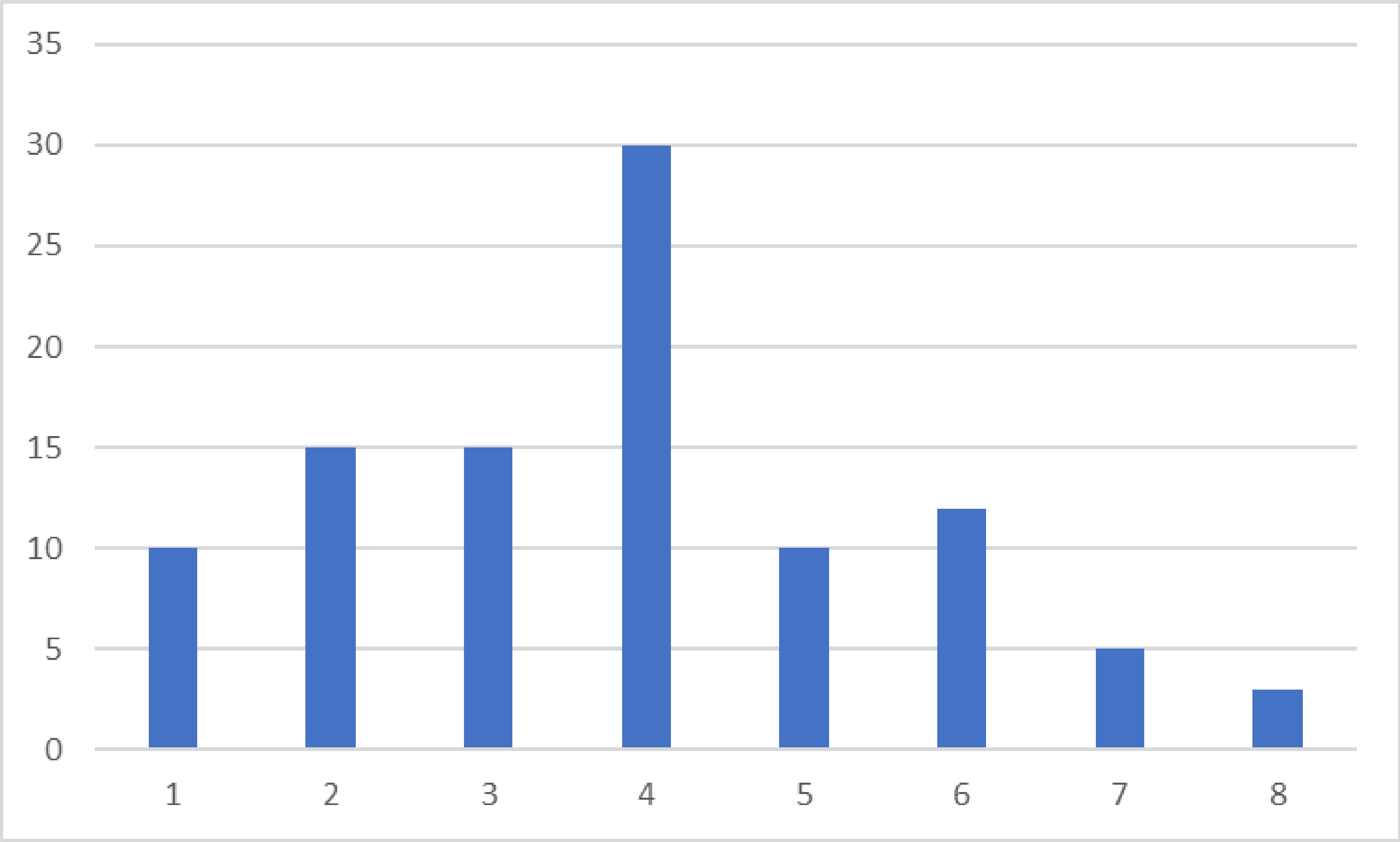
Graphique

- Pour les caractères quantitatifs discrets **:** le diagramme le plus approprié est le diagramme en bâtons**.** On porte en abscisse les différentes modalités du caractère X et en ordonnée les effectifs.

Exemple : la distribution de 100 femmes selon le nombre d’enfants

xi	1	2	3	4	5	6	7	8	Total
ni	10	15	15	30	10	12	5	3	100

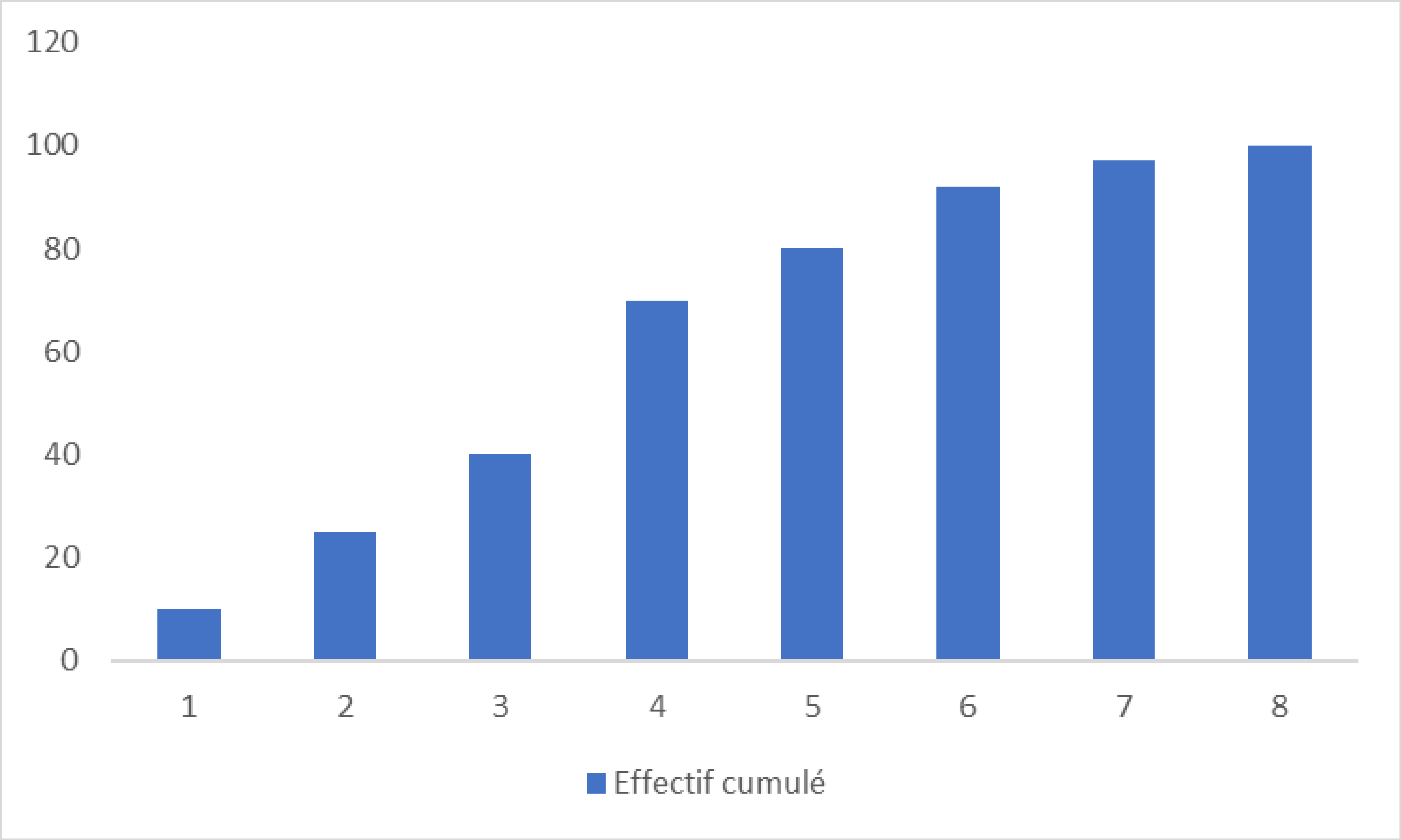
Diagramme en bâton



Le diagramme des effectifs cumulés ou diagramme en escalier permet de représenter les effectifs cumulés.

xi	ni	Effectif cumulé N(x)
1	10 →	10
2	15	+15 = 25
3	15	+15 = 40
4	30	+30 = 70
5	10	+10 = 80
6	12	+12 = 92
7	5	+5 = 97
8	3	+3 = 100

Diagramme des effectifs cumulés ou diagramme en escalier



- Pour les caractères quantitatifs continus **:** le diagramme le plus approprié est l'histogramme. Si toutes les classes ont la même amplitude ai, on porte directement en ordonnée les fi ou les ni. Si les classes n'ont pas la même amplitude, on corrige d'abord les effectifs en calculant la surface du rectangle :

$hi = (ni/ai) \times ai$

ni = effectif partiel de la classe à corriger
ai = amplitude la plus fréquente ou amplitude la plus petite
ai = amplitude de la classe à corriger

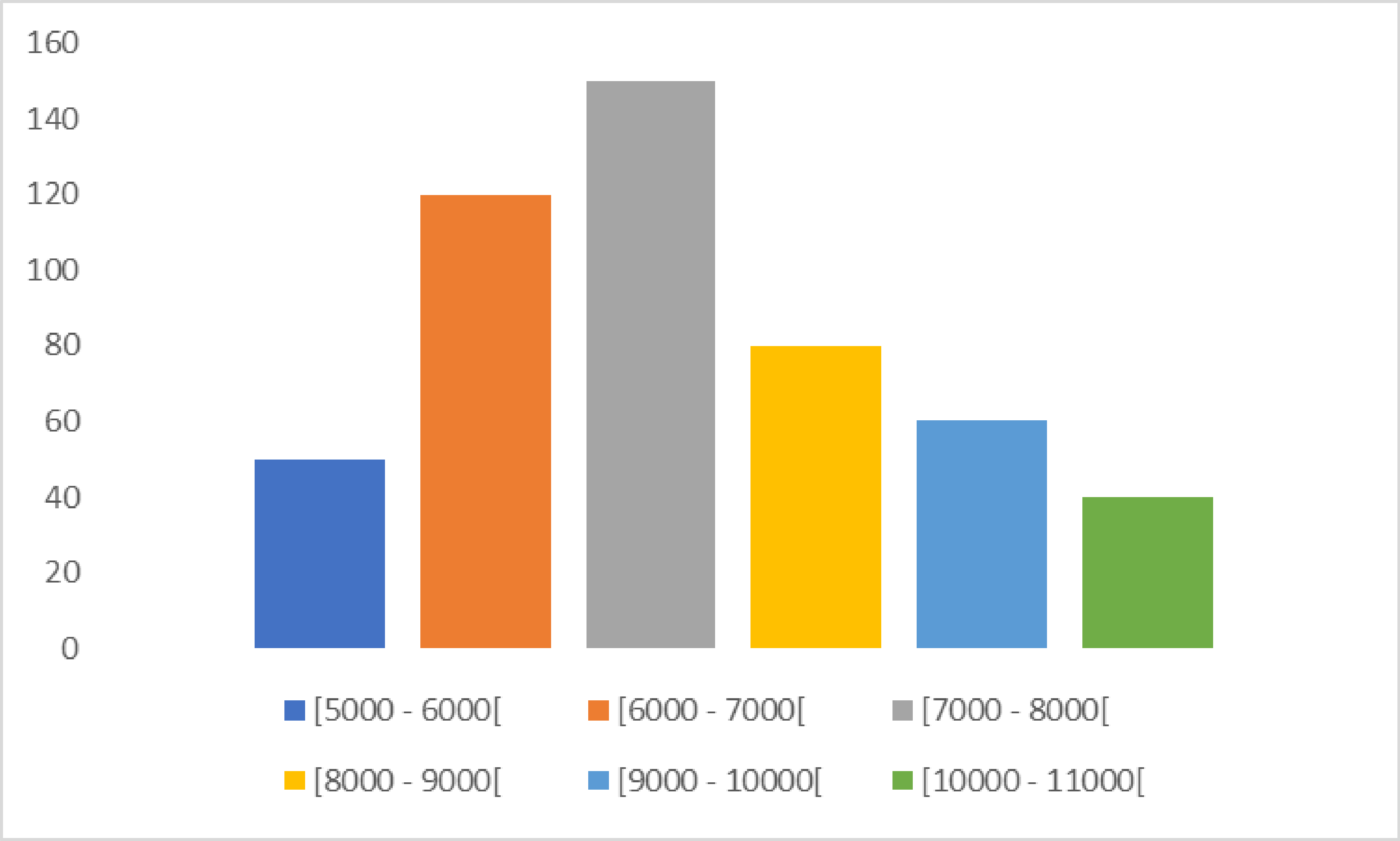
NB : L'amplitude est la différence entre l'extrémité et l'origine de la classe.

Application : Soit la répartition du loyer annuel en euros de 500 individus.

xi	Effectif	**Amplitude**
[5000 - 6000[	50	1000
[6000 - 7000[	120	1000
[7000 - 8000[	150	1000
[9000 - 10000[	80	1000
[10000 - 11000[	60	1000
[11000 - 12000[	40	1000

1^{e} cas : les classes ont la même amplitude

Histogramme



2e cas : les classes n'ont pas la même amplitude

Loyer annuel en euros	Effectif	Amplitude
[5000 - 6000[	50	**1000**
[6000 – 8000[	~~120~~ 240	**2000**
[8000 - 11000[	~~150~~ 450	**3000**
[11000 - 15000[	~~80~~	**4000**
[15000 - 20000[	~~60~~	**5000**
[20000 - 30000[	~~40~~	**10000**

Dans ce cas, on repère d'abord les classes qui ont la même amplitude et on corrige les effectifs des autres classes en utilisant la formule hi = (ni/ai) × ai

hi [6000 - 8000[= (120 / **1000**) × 2000 = 240

Hi [8000 - 11000[= (150 / **1000**) × 3000 = 450

La courbe des fréquences cumulées

xi	ni	fi %	Fréquences cumulées croissantes F(x)	Fréquences cumulées décroissantes (F(x)
[5000 - 6000[	50	10 %	10	100
[6000 - 7000[	120	24 %	+24 = 34	- 10 = 90
[7000 - 9000[	150	30 %	+30 = 64	- 24 = 66
[9000 - 11000[	80	16 %	+16 = 80	- 30 = 36
[11000 - 13000[	60	12 %	+12 = 92	- 16 = 20
[13000 - 15000[	40	8 %	+8 = 100	- 12 = 8
	n = 500			

Rappel : fi% = ni (effectif partiel) / n (effectif total) x 100

Application : classe [5000 - 6000[: fi% = (50/500) x 100 = 10%

Application : soit la distribution statistique suivante

1. Calculez fi%
2. Calculez les effectifs cumulés
3. Calculez les fréquences cumulées croissantes
4. Calculez les fréquences cumulées décroissantes
5. Construire la courbe des fréquences cumulées

xi	ni
50 - 65	23
65 – 105	15
105 - 115	19
115 – 125	47
125 - 135	9

xi	ni	fi%	N(x) (effectif cumulé)	Fréq. cumulée croissante	Fréq. cumulée décroissante
50 – 65	23	20	23	20	100
65 – 105	15	13	38	33	80
105 - 115	19	17	57	50	67
115 – 125	47	42	104	92	50
125 - 135	9	08	113	100	08

n =113

Pour tracer la courbe des fréquences cumulées décroissantes, l’origine des classes sera prise comme référence en abscisse.

Pour tracer la courbe des fréquences cumulées croissantes l’extrémité des classes sera prise comme référence en abscisse.