Cours des Statistques Semestre 1

Préparer par Karim Economiste

Plan du cours :

- Les termes utilisés aux statistiques
- Les fréquences
  - Les fréquences absolues
  - Les fréquences relatives
- Les paramètres de la tendance centrale
  - Le mode
  - La médiane
  - Les moyennes
- Les paramètres de la dispersion
  - L'étendu
  - L'écart interquartile
  - L'écart absolu moyen
  - Le coefficient de variation
- La représentation graphique

Résumé de Statistique par Karim Economiste

* Les termes utilisés en statistique

* La population: est l'ensemble des individus étudiée
exemple : 200 élèves; 1000 femmes; 250 Entreprises, 400 étudiants...
* L'unité statistique: est un individu du population
exemple : Un élève; une femme; une entreprise; un étudiant...
* Le carctère X_i : La caractéristique d'un individu ou ensemble des individus
exemple: Les étudiant peuvent être caractérisés par le caractère type de Bac
Les femmes peuvent caractérisées par le caractère nombre des enfants
* La modalité : chaque caractère possède deux ou plusieurs modalités, Ce sont les différentes situation où les individus peuvent se trouver à l'égard du caractère:
exemple: le caractère type de Bac peuvent avoir comme modalités; SUT; Lettre, PC, ECO ..
le caractère nombre des enfants peuvent avoir comme modalités; 0 ; 1 ; 2; 3; ..

* La nature du caractère

→ Caractère qualitatif:
c'est un caractère non mesurable, la modalité ne peuvent pas être des nombres.
Exemple : le caractère type de Bac (SUT, ECO, PC, lettre..)

→ Caractère quantitatif:
c'est un caractère mesurable, chaque modalité correspond à un nombre
Le caractère quantitatif est divisée en deux:
1. Caractère quantitatif discrète:
chaque modalité correspond à un nombre isolé
exemple : nombre des enfants (0 ; 1 ; 2 ; 3 ; ..)
2. Caractère quantitatif continue:
chaque modalité correpond à un classe des nombres (intervalle des nombres)
Exemple : L'âge ([0, 10[; [10, 20[; [20, 30[, [30, 40[; ...)

Exemple 1
Soit la distribution de 200 étudiants selon le type de Bac

Type de Bac X_i	nombre des étudiants n_i
SUT	40
ECO	80
PC	60
Lettre	20
Total	200

$n = 200$
$n = \sum_{i=1}^{4} n_i$
$n_1 = 40$
$n_2 = 80$
$n_3 = 60$
$n_4 = 20$
$n = n_1 + n_2 + n_3 + n_4$

* La population : 200 étudiant
* L'unité statistique : un étudiant
* Le caractère : Type de Bac
* La nature du caractère : caractère qualitatif
* Les différentes modalités du caractères : SUT, ECO, PC Lettre.

$n_1 = 40$; signifie que 40 étudiants ayant un bac SUT

$n_2 = 80$; signifie que 80 étudiants ayant un bac ECO

$n_3 = 60$; signifie que 60 étudiants ayant un Bac PC

$n_4 = 20$; signifie que 20 étudiants ayant un Bac Lettre

Exemple ②

Soit la distribution de 400 familles selon le nombre des enfants

nombre des enfants X_i	nombre des familles n_i
0	20
1	120
2	180
3	50
4	30
Total	400

* La population : 400 familles
* L'unité statistique : une familles
* Le caractère : nombre des enfants
* La nature du caractère : Caractère quantitatif discrète
* Les différentes modalités : 0 ; 1 ; 2 ; 3 ; 4

$n_1 = 20$; signifie que 20 familles ayant zéro enfant

$n_2 = 120$, signifie que 120 familles ayant un enfants

$n_3 = 180$, signifie que 180 familles ayant deux enfants

Exemple ③

Soit la distribution de 100 Entreprises selon le montant de capital (en million de dhs)

Montant de capital X_i	nombre des E/ses n_i
[10, 20[	10
[20, 30[	50
[30, 40[	20
[40, 50]	20
Total	100

* La population : 100 Entrepris
* L'unité statistique : Une E/se
* Le caractère : le Montant de capital
* La nature du caractère : Caractère quantitatif continue
* Les différentes modalités [10, 20[; [20, 30[; [30, 40[; [40, 50]

$n_1 = 10$; signifie que 10 E/ses ayant un montant de capital entre 10 millions et 20 millions

$n_2 = 50$, signifie que 50 E/ses ayant un montant de capital entre 20 millions et 30 millions

* Les frèquences

* Les frèquences Absolues

la frèquence Absolue est l'effectif n_i sont les même

frèquence Absolue = n_i

on trouve deux types de frèquence Absolue

① frèquence Absolue cumulée croissante ou l'effectif cumulée croissante (ECC ou N_{i+} ou $N_i\uparrow$ ou N_iCC)

② frèquence Absolue cumulée décroissante ou l'effectif cumulée décroissante (ECD ou N_{i-} ou $N_i\downarrow$ ou N_iCD)

X_i	l'effectif n_i	ECC	ECD
X_1	n_1	n_1	n
X_2	n_2	n_1+n_2	$n-n_1$
X_3	n_3	$n_1+n_2+n_3$	$n-n_1-n_2$
X_4	n_4	$n_1+n_2+n_3+n_4$	$n-n_1-n_2-n_3$
X_5	n_5	$n_1+n_2+n_3+n_4+n_5=n$	$n-n_1-n_2-n_3-n_4=n_5$
Total	n		

Exemple

X_i	n_i	ECC	ECD
1	20	20	300
2	60	80	280
3	120	200	220
4	80	280	100
5	20	300	20
Total	300		

* Les frèquences relatives

① frèquence relative simple f_i

$$\boxed{f_i = \frac{n_i}{n}}$$

② frèquence relative cumulée croissante (FCC ou F_{i+} ou $F_i\uparrow$)

③ frèquence relative cumulée décroissante (FCD ou F_{i-} ou $F_i\downarrow$)

X_i	n_i	f_i	FCC	FCD
X1	n_1	$\frac{n_1}{n}=f_1$	f_1	1
X2	n_2	$\frac{n_2}{n}=f_2$	f_1+f_2	$1-f_1$
X3	n_3	$\frac{n_3}{n}=f_3$	$f_1+f_2+f_3$	$1-f_1-f_2$
X4	n_4	$\frac{n_4}{n}=f_4$	$f_1+f_2+f_3+f_4$	$1-f_1-f_2-f_3$
X5	n_5	$\frac{n_5}{n}=f_5$	$f_1+f_2+f_3+f_4+f_5=1$	$1-f_1-f_2-f_3-f_4=f_5$
Total	n	1		

Exemple

X_i	n_i	f_i	FCC	FCD
1	20	0,07	0,07	1
2	60	0,2	0,27	0,93
3	120	0,4	0,67	0,73
4	80	0,26	0,93	0,33
5	20	0,07	1	0,07
Total	300	1		

$$f_1 = \frac{n_1}{n} = \frac{20}{300} = 0,066 \simeq 0,07$$

$$f_2 = \frac{n_2}{n} = \frac{60}{300} = 0,2$$

$$f_3 = \frac{n_3}{n} = \frac{120}{300} = 0,4$$

$$f_4 = \frac{n_4}{n} = \frac{80}{300} = 0,266 \simeq 0,26$$

$$f_5 = \frac{n_5}{n} = \frac{20}{300} = 0,066 \simeq 0,07$$

* Les Paramètres de la tendence centrale

① Le Mode:

Le Mode est la valeur qui correspond à l'effectif maximal.

* Cas d'un caractère discret

Exemple

X_i	n_i
1	20
2	60
3	80
4	40
Total	200

⇐ l'effectif maximal

On a 80 est l'effectif maximal
Donc la modalité qui correspond à cet effectif est $X_3 = 3$
Donc le Mode est 3

$$\boxed{M_o = 3}$$

* Cas d'un caractère continue

Pour le cas continue avant de calculer le Mode il faut d'Abord calculer l'amplitude a_i pour chaque classe

$$\boxed{a_i = \text{born sup} - \text{born inf}}$$

classe : [born inf ; born sup[

Pour le Mode de cas continue on a deux cas possible

① Cas d'Amplitude égale c'est à dire tous les classe ont le même amplitude (a_i).

$$M_o = \text{borninf} + a_i \times \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

n_i = l'effectif de classe modale

Exemple

X_i	n_i	a_i
[0, 5[	20	5
[5, 10[	60	5
[10, 15[	80	5
[15, 20]	40	5
Total	200	

$a_1 = 5 - 0 = 5$
$a_2 = 10 - 5 = 5$
$a_3 = 15 - 10 = 5$
$a_4 = 20 - 15 = 5$

Tous les classe ont le même amplitude

$$M_o = \text{borninf} + a_i \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

- Classe modale est la classe qui correspond à l'effectif maximale
on a l'effectif maximale est 80 donc le classe correspond à cet effectif est [10, 15[c'est le classe modale

$$M_o = 10 + 5 \times \frac{80 - 60}{(80 - 60) + (80 - 40)}$$

$$\boxed{M_o = 11{,}66}$$

② Cas d'Amplitude inégale ; c'est à dire les classes ont des amplitudes differents

Dans ce cas on a besoin de calculer la densité (d_i) pour chaque classe

$$\boxed{d_i = \frac{n_i}{a_i}}$$

$$M_o = \text{borninf} + a_i \times \frac{d_i - d_{i-1}}{(d_i - d_{i-1}) + (d_i - d_{i+1})}$$

d_i: densité de classe modale

Exemple

X_i	n_i	a_i	d_i
[0, 5[	20	5	4
[5, 15[	60	10	6
[15, 20[	80	5	16
[20, 30]	40	10	4
Total	200		

$d_1 = \frac{20}{5} = 4$
$d_2 = \frac{60}{10} = 6$
$d_3 = \frac{80}{5} = 16$
$d_4 = \frac{40}{10} = 4$

$$M_o = \text{borninf} + a_i \frac{d_i - d_{i-1}}{|d_i - d_{i-1}| + |d_i - d_{i+1}|}$$

- le classe modale est le classe correspond à la densité maximale ; d_i maximale est 16 donc classe modale est [15, 20[

$$M_o = 15 + 5 \cdot \frac{16 - 6}{(16 - 6) + (16 - 4)}$$

$$\boxed{M_o = 17{,}27}$$

② La mèdiane :

La mèdiane est la valeur qui pairtage la serie statistique en deux parties ègales

* Cas d'un caractère discrèt :

Pour le cas discret la mèdiane est déterminée par l'observation

Exemple

X_i	n_i	ECC
1	40	40
2	80	120
3	60	180
4	20	200
Total	200	

$\Rightarrow 120 > 100$

Pour déterminer la mèdiane on diviser total $n/2$

$$n/2 = \frac{200}{2} = 100$$

on observe ECC dans lequel est la valeur est superieur de $\frac{n}{2} = 100$, cette valeur doit être le plus moins valeur de ECC qu'est supèrieur de $\frac{n}{2} = 100$, Donc dans cet exemple est il est 120

donc la mèdiane est la valeur de X_i correspond à ce ECC = 120 qu'est 2

Donc on note

$$\boxed{Me = 2}$$

* Cas d'un caractère continue

Pour le cas continue la mèdiane est calculée par la relation suivante :

$$Me = \text{born inf} + a_i \, \frac{n/2 - N_{i-1}}{n_i}$$

$n/2$: total divisé sur 2

N_{i-1} : ECC du classe avant le classe mèdiane

n_i : l'effectif du classe mèdiane

Exemple

X_i	n_i	ECC	a_i
[0,5[	40	40	5
[5,10[	80	120	5
[10,15[	60	180	5
[15,20]	20	200	5
Total	200		

$$Me = \text{borninf} + a_i \, \frac{n/2 - N_{i-1}}{n_i}$$

- classe mèdiane : pour déterminer la classe mèdiane on divise n sur 2 ; $n/2 = \frac{200}{2} = 100$

on observe ECC où 100 se situe, Donc on a ECC = 120 > 100, donc la classe mèdiane est la classe correspond à cette valeur de ECC = 120

Donc la classe mèdiane est la classe [5,10[

$$Me = 5 + 5 \times \frac{100 - 40}{80}$$

$$\boxed{Me = 8{,}75}$$

Donc la mèdiane est 8,75

③ Les moyennes

Ⓐ La moyenne arithmétique $\overline{X}$

La moyenne arithmétique est égale à la somme des valeurs observées divisées par le nombre d'observations

* Cas discrèt

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} f_i x_i$$

Exemple

x_i	n_i	$n_i x_i$
1	40	40
2	80	160
3	60	180
4	20	80
Total	200	460

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{4} n_i x_i = \frac{\sum_{i=1}^{4} n_i x_i}{n}$$

$$\overline{X} = \frac{460}{200}$$

$$\boxed{\overline{X} = 2,3}$$

* Cas continue

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} n_i C_i = \sum_{i=1}^{k} f_i C_i$$

C_i : centre de classe

$$\boxed{C_i = \frac{\text{born sup} + \text{born inf}}{2}}$$

Exemple

On a un classe $[10,20[$

$$C_i = \frac{10+20}{2} = 15$$

Exemple

x_i	n_i	C_i	$n_i C_i$
[0,10[	60	5	300
[10,20[	120	15	1800
[20,30[	140	25	3500
[30,40]	80	35	2800
Total	400		8400

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{4} n_i C_i = \frac{\sum_{i=1}^{4} n_i C_i}{n}$$

$$\overline{X} = \frac{8400}{400}$$

$$\boxed{\overline{X} = 21}$$

Ⓑ La moyenne géométrique

$$G = e^{\log G}$$

$$\log G = \frac{1}{n} \sum_{i=1}^{k} n_i \log(x_i)$$

* Cas discrèt

$$\log G = \frac{1}{n} \sum_{i=1}^{k} n_i \log(x_i)$$

Exemple

x_i	n_i	$\log(x_i)$	$n_i \log(x_i)$
1	40	0	0
2	80	0,3	24
3	60	0,47	28,2
4	20	0,6	12
Total	200		64,2

$$\log G = \frac{1}{n} \sum_{i=1}^{4} n_i \log(x_i)$$

$$\log G = \frac{64,2}{200} \Longleftrightarrow \boxed{\log G = 0,321}$$

$$G = e^{\log G}$$

$$G = e^{0,321} \Longleftrightarrow \boxed{G = 1,37}$$

* Cas continue

$$\text{Log } G = \frac{1}{n} \sum_{i=1}^{k} n_i \log(c_i)$$

Exemple

X_i	n_i	c_i	$\log(c_i)$	$n_i \log(c_i)$
0-5	70	2,5	0,397	27,79
5-10	110	7,5	0,875	96,25
10-15	80	12,5	1,096	87,68
15-20	40	17,5	1,243	49,72
Total	300			261,44

$$\text{Log } G = \frac{1}{n} \sum_{i=1}^{4} n_i \log(c_i)$$

$$\text{Log } G = \frac{261,44}{300}$$

$$\boxed{\text{Log } G = 0,871}$$

$$G = e^{\log G}$$

$$G = e^{0,871}$$

$$\boxed{G = 2,367}$$

© La Moyenne harmonique

La moyenne harmonique se définit comme l'inverse de la moyenne arithmétique des inverses

$$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{X_i}} \quad \text{(cas discrèt)}$$

$$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{c_i}} \quad \text{(Cas continue)}$$

* Cas discrèt

X_i	n_i	n_i/X_i
1	40	40
2	80	40
3	60	20
4	20	5
Total	200	105

$$H = \frac{n}{\sum_{i=1}^{4} \frac{n_i}{X_i}}$$

$$H = \frac{200}{105}$$

$$\boxed{H = 1,904}$$

* Cas continue

X_i	n_i	c_i	n_i/c_i
0-5	70	2,5	28
5-10	110	7,5	14,66
10-15	80	12,5	6,4
15-20	40	17,5	2,28
Total	300		51,34

$$H = \frac{n}{\sum_{i=1}^{n} \frac{n_i}{c_i}}$$

$$H = \frac{300}{51,34}$$

$$\boxed{H = 5,843}$$

① La moyenne Quadratique

C'est la racine carré de la moyenne arithmétique des carrés.

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{k} n_i X_i^2} \quad \text{(cas discrét)}$$

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{k} n_i C_i^2} \quad \text{(cas continue)}$$

* Cas discrèt

X_i	n_i	X_i^2	$n_i X_i^2$
1	40	1	40
2	80	4	320
3	60	9	540
4	20	16	320
Total	200		1220

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{4} n_i X_i^2}$$

$$Q = \sqrt{\frac{1}{200} \times 1220}$$

$$Q = \sqrt{\frac{1220}{200}}$$

$$\boxed{Q = 2,469}$$

* Cas continue

X_i	n_i	C_i	C_i^2	$n_i C_i^2$
0-5	70	2,5	6,25	437,5
5-10	110	7,5	56,25	6187,5
10-15	80	12,5	156,25	12500
15-20	40	17,5	306,25	12250
Total	300			31375

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{4} n_i C_i^2}$$

$$Q = \sqrt{\frac{31375}{300}}$$

$$\boxed{Q = 104,583}$$

* Les paramètres de la dispersion :

① L'Etendu

L'étendu est la différence entre la plus grande valeur et la plus petite valeur du caractère

$$e = X_{max} - X_{min}$$

* Cas discrèt

X_i	n_i
1	20
2	80
3	60
4	40
5	20
Total	220

$$e = X_{max} - X_{min}$$

$$e = 5 - 1$$

$$\boxed{e = 4}$$

* 5 est la plus grande valeur du caractère X_i

* 1 est la plus petite valeur du caractère X_i

* Cas continue

X_i	n_i
0-5	110
5-15	200
15-20	190
20-40	100
40-50	40
Total	640

$e = X_{max} - X_{min}$

$e = 50 - 0 = 50$

$\boxed{e = 50}$

* 50 est la valeur la plus grande du caractère X_i

* 0 est la valeur la plus petite du caractère X_i

② L'écart interquartiles

Ⓐ Les quartiles

On a 3 quartiles Q_1, Q_2, Q_3

Q_1 : est la valeur qui partage la serie statistique en deux parties, 1er partie 25% et la 2ème 75%.

Q_2 est la médiane, c'est à dire la valeur qui partage la serie statistique en deux parties égales

Q_3 est la valeur qui partage la serie statistique en deux parties, 1er partie 75% et la 2ème 25%.

* cas discrèt

X_i	n_i	ECC
1	40	40
2	80	120
3	60	180
4	20	200
Total	200	

Q_1 : Pour calculer Q_1 on divise $\frac{n}{4}$; $\frac{200}{4} = 50$

on cherche dans ECC la valeur qui est grande de 50 ; c'est 120 donc la valeur X_i correspond à 120 est Q_1

$\boxed{Q_1 = 2}$

Q_2 : Pour calculer Q_2 on divise $\frac{n}{2}$; $\frac{200}{2} = 100$

on cherche dans ECC la valeur qui est grande de 100 ; c'est 120 don la valeur X_i correspond à 120 est Q_2

$\boxed{Q_2 = 2}$

Q_3 : Pour calculer Q_3 on divise $\frac{3n}{4}$; $\frac{3 \times 200}{2} = 150$;

on cherche dans ECC la valeur qui est grande de 150 ; c'est 180 donc la valeur X_i correspond à 180 est Q_3

$\boxed{Q_3 = 3}$

* Cas continue

Pour le cas continue ; les Quartiles sont calculées par une méthode différente du cas discrèt

X_i	n_i	ECC
0-5	70	70
5-10	110	180
10-15	80	260
15-20	40	300
Total	300	

$Q_1 = \text{borninf} + a_i \dfrac{n/4 - N_{i-1}}{n_i}$

$\dfrac{n}{4} : \dfrac{\text{Total}}{4}$

N_{i-1} : ECC du classe avant la classe de Q_1

n_i : effectif ni du classe de Q_1

⇒ classe de Q_1 : Pour determiner la classe de Q_1 on calcule $\dfrac{n}{4} = \dfrac{300}{4} = 75$, on cherche dans ECC la valeur qu'est superieur de 75, cette valeur doit être la plus petite superieur de 75, dans ce cas 180, donc la classe qui correspond à 180 qu'est 5-10 est la classe de Q_1

$Q_1 = 5 + 5 \times \dfrac{75 - 70}{110}$

$\boxed{Q_1 = 5{,}22}$

$Q_2 = \text{borninf} + a_i \dfrac{n/2 - N_{i-1}}{n_i}$

⇒ classe de Q_2. Pour determiner la classe de Q_2 (classe médiane) on calcule $\dfrac{n}{2} = \dfrac{300}{2} = 150$, on cherche dans ECC la valeur qu'est superieur de 150, cette valeur doit être la petite superieur de 150 dans ce cas c'est 180, donc la classe de Q_2 est 5-10

$Q_2 = 5 + 5 \times \dfrac{150 - 70}{110}$

$\boxed{Q_2 = 8{,}63}$

$Q_3 = \text{borninf} + a_i \dfrac{\frac{3n}{4} - N_{i-1}}{n_i}$

⇒ classe de Q_3 : pour déterminer la classe de Q_3 on calcule $\dfrac{3n}{4} = \dfrac{3 \times 300}{4} = 225$, on cherche dans ECC : la valeur la plus petite qu'est superieure de 225, c'est 260, donc la classe de Q_3, est la classe correspond à 260, qu'est 10-15

$Q_3 = 10 + 5 \times \dfrac{225 - 180}{80}$

$\boxed{Q_3 = 12{,}81}$

Ⓑ L'écart interquartiles

$EI = Q_3 - Q_1$

Ⓒ Coefficient d'écart interquartile

$C.E.I = \dfrac{Q_3 - Q_1}{Q_2}$

interprétation

- si $0 < CEI < 0{,}5$ ⇒ dispersion faibl
- si $0{,}5 \leq CEI < 0{,}8$ ⇒ dispersion moyenne
- si $0{,}8 \leq CEI < 1$ ⇒ dispersion forte

Ⓓ intervalle des interquartiles

$[Q_1 , Q_3]$

③ Les écarts absolus moyen

Ⓐ L'écart absolu moyen par rapport au Mode $E.AM_{Mo}$

$E.A.M_{Mo} = \dfrac{1}{n} \sum_{i=1}^{n} n_i |X_i - Mo|$

$|X_i - Mo|$ est la valeur absolue

* Cas discrèt

X_i	n_i	$\lvert X_i - M_o \rvert$	$n_i \lvert X_i - M_o \rvert$
1	40	2	80
2	60	1	60
3	80	0	0
4	20	1	20
Total	200		160

$$E.A.M_{M_o} = \frac{1}{n} \sum_{i=1}^{4} n_i \lvert X_i - M_o \rvert$$

Le Mode $\boxed{M_o = 3}$

$$E.A.M_{M_o} = \frac{160}{200}$$

$$\boxed{E.A.M_{M_o} = 0{,}8}$$

* Cas continue

X_i	n_i	a_i	c_i	$\lvert c_i - M_o \rvert$	$n_i \lvert c_i - M_o \rvert$
0-5	40	5	2,5	8,75	350
5-10	60	5	7,5	3,75	225
10-15	80	5	12,5	1,25	100
15-20	20	5	17,5	6,25	125
Total	200				800

$$E.A.M_{M_o} = \frac{1}{n} \sum_{i=1}^{4} n_i \lvert c_i - M_o \rvert$$

Le Mode

$$M_o = borninf + a_i \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

classe modale $[10, 15[$

$$M_o = 10 + 5 \times \frac{80-60}{(80-60)+(80-20)}$$

$$\boxed{M_o = 11{,}25}$$

$$E.A.M_{M_o} = \frac{800}{200}$$

$$\boxed{E.A.M_{M_o} = 4}$$

(B) L'écart absolu moyen par rapport à la Médiane $E.A.M_{Me}$

$$E.A.M_{Me} = \frac{1}{n} \sum_{i=1}^{n} n_i \lvert X_i - Me \rvert$$

* Cas discret

X_i	n_i	ECC	$\lvert X_i - Me \rvert$	$n_i \lvert X_i - Me \rvert$
1	40	40	1	40
2	80	120	0	0
3	60	180	1	60
4	20	200	2	40
Total	200			140

$$E.A.M_{Me} = \frac{1}{n} \sum_{i=1}^{4} n_i \lvert X_i - Me \rvert$$

La médiane

$\boxed{Me = 2}$

$$E.A.M_{M_o} = \frac{140}{200}$$

$$\boxed{E.A.M_e = 0{,}7}$$

* Cas continue

X_i	n_i	ECC	c_i	$\lvert c_i - Me \rvert$	$n_i \lvert c_i - Me \rvert$
0-5	40	40	2,5	6,25	250
5-10	80	120	7,5	1,25	100
10-15	60	180	12,5	3,75	225
15-20	20	200	17,5	8,75	175
Total	200				750

$$E.A.M_{Me} = \frac{1}{n} \sum_{i=1}^{n} n_i \lvert c_i - Me \rvert$$

La médiane

$$Me = borninf + a_i \frac{n/2 - N_{i-1}}{n_i}$$

classe mediane 5-10

$$Me = 5 + 5 \times \frac{100 - 40}{80}$$

$$\boxed{Me = 8{,}75}$$

$$E.A.M_{Me} = \frac{750}{200}$$

$$\boxed{E.A.M_{Me} = 3{,}75}$$

© L'écart absolu moyen par rapport à la moyenne arithmétique

$$E.A.M_{\overline{X}} = \frac{1}{n} \sum_{i=1}^{n} n_i \, |X_i - \overline{X}|$$

* Cas discret

X_i	n_i	$n_i x_i$	$\lvert X_i - \overline{X} \rvert$	$n_i \lvert x_i - \overline{X} \rvert$
1	40	40	1,4	56
2	60	120	0,4	24
3	80	240	0,6	48
4	20	80	1,6	32
Total	200	480		160

$$E.A.M_{\overline{X}} = \frac{1}{n} \sum_{i=1}^{4} n_i \, |x_i - \overline{X}|$$

- La moyenne arithmétique

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{4} n_i x_i$$

$$\overline{X} = \frac{480}{200} \iff \boxed{\overline{X} = 2{,}4}$$

$$E.A.M_{\overline{X}} = \frac{160}{200} \iff \boxed{E.A.M_{\overline{X}} = 0{,}8}$$

* Cas continue

X_i	n_i	c_i	$n_i c_i$	$\lvert c_i - \overline{X} \rvert$	$n_i \lvert c_i - \overline{X} \rvert$
0-5	40	2,5	100	7	280
5-10	60	7,5	450	2	120
10-15	80	12,5	1000	3	240
15-20	20	17,5	350	8	160
Total	200		1900		800

$$E.A.M_{\overline{X}} = \frac{1}{n} \sum_{i=1}^{4} n_i \, |c_i - \overline{X}|$$

- La moyenne arithmétique

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{4} n_i c_i$$

$$\overline{X} = \frac{1900}{200} \iff \boxed{\overline{X} = 9{,}5}$$

$$E.A.M_{\overline{X}} = \frac{800}{200}$$

$$\boxed{E.A.M_{\overline{X}} = 4}$$

④ Coefficien de variation C.V

① La variance

$$Var(x) = \frac{1}{n} \sum_{i=1}^{n} n_i (X_i - \overline{X})^2$$

② L'écart type

$\begin{cases} X_i = \text{cas discret} \\ C_i = \text{cas continue} \end{cases}$

$$\sigma x = \sqrt{Var(x)}$$

③ Coefficient de variation

$$C.V = \frac{\sigma x}{\overline{X}}$$

Exemple

X_i	n_i	c_i	$n_i c_i$	$(c_i - \overline{X})^2$	$n_i (c_i - \overline{X})^2$
0-5	40	2,5	100	49	1960
5-10	60	7,5	450	4	240
10-15	80	12,5	1000	9	760
15-20	20	17,5	350	64	1280
Total	200		1900		4200

* La moyenne arithmétique

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{4} n_i c_i$$

$$\overline{X} = \frac{1900}{200} \iff \boxed{\overline{X} = 9{,}5}$$

* La variance

$$Var(x) = \frac{1}{n} \sum_{i=1}^{4} n_i (c_i - \overline{X})^2$$

$$Var(x) = \frac{4200}{200}$$

$$\boxed{Var(x) = 21}$$

* L'écart type :

$$\sigma x = \sqrt{Var x}$$

$$\sigma x = \sqrt{21} \iff \boxed{\sigma x = 4{,}58}$$

* Coefficien de variation

$$C.V = \frac{\sigma x}{\overline{X}}$$

$$C.V = \frac{4{,}58}{9{,}5} \iff \boxed{C.V = 0{,}48}$$

$C.V = 0{,}48 < 0{,}5$ donc la dispersion est faible

Graphiques statistiques de base

La représentation graphique des fréquences simples ou cumulées relatives à une variable statistique, donne lieux aux distinctions entre graphiques de distribution appelée diagrammes différentiels et graphiques de répartition appelées diagramme intégral ou cumulatif.

Graphiques de fréquences simples, diagrammes différentiels

Le diagramme différentiel prend plusieurs formes, selon qu'on traite une variable qualitative, quantitative discrète ou quantitative continue

Cas d'une variable qualitative

Pour représenter une variable qualitative, généralement on utilise 2 types de représentations graphiques :

Diagramme en tuyaux d'orgues

Pour construire un diagramme en tuyaux d'orgues, on doit porter sur l'axe des abscisses les modalités xi et sur l'axe des ordonnées les effectifs ni, puis on représente les modalités par des rectangles juxtaposées dont la hauteur est proportionnelle à l'effectif correspondant et dont la base est une constante.

Les modalités	les effectifs
X1	20
X2	30
X3	15
X4	10

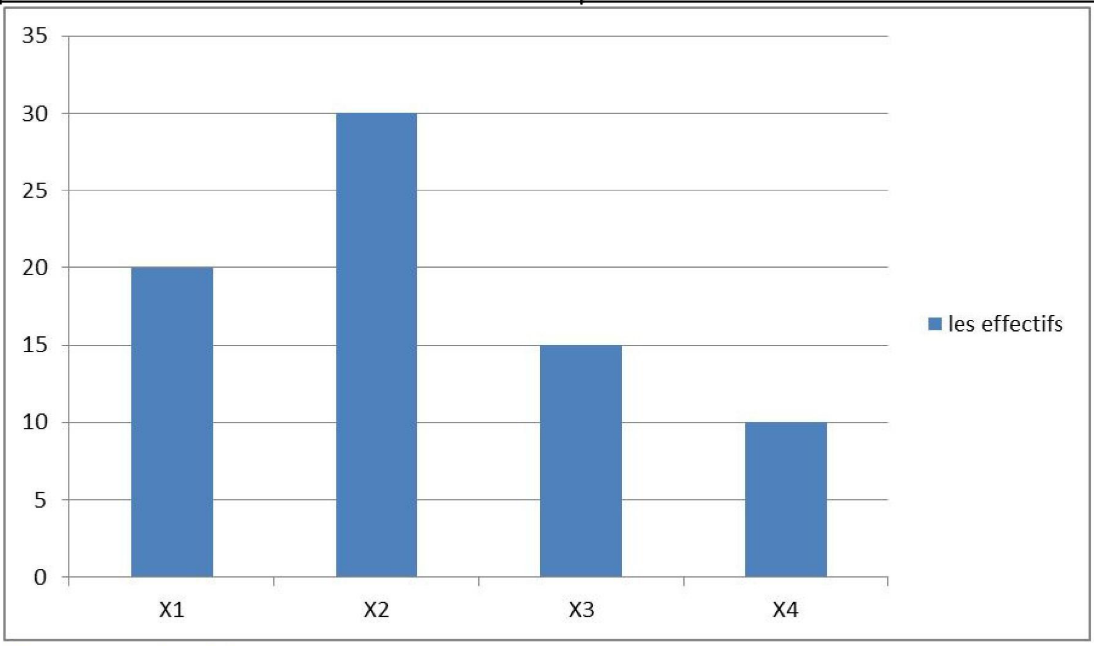


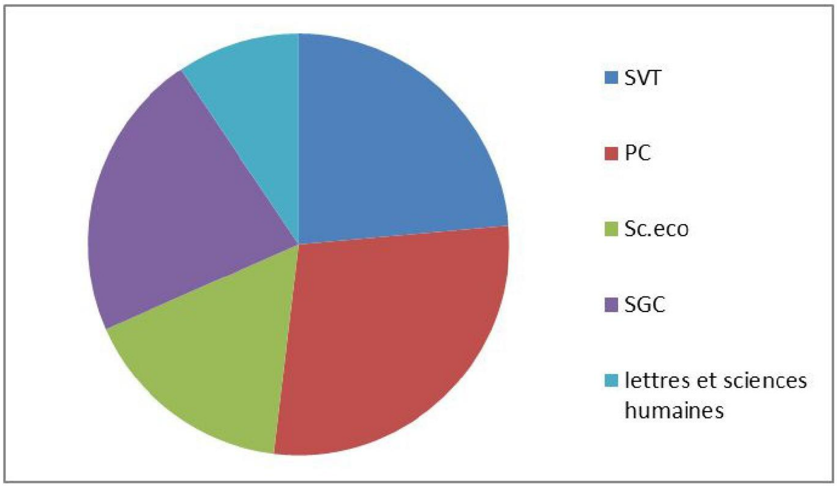
Diagramme circulaire

L'élaboration de ce diagramme consiste à répartir un angle de 360° entre les différentes modalités selon le poids de chacune dans l'ensemble de la population.
Pour obtenir le secteur correspondant à chaque modalité, on va se baser sur la règle de 3, qui va nous la chose suivante :

N 360
ni αi Donc $\alpha_i = 360*n_i/N$

Exemple

Type du baccalauréat	Nombre des étudiants
SVT	250
PC	300
Sc.Eco	175
SGC	235
Lettres et sciences humaines	100
Total	1060



SVT= 360*250 /1060 = 84,90
PC=360*300/1060= 101,88
Sc.Eco= 360*175/1060=59,43
SGC= 360*235/1060=79,81
L.SC. H = 360*100/1060=33,98

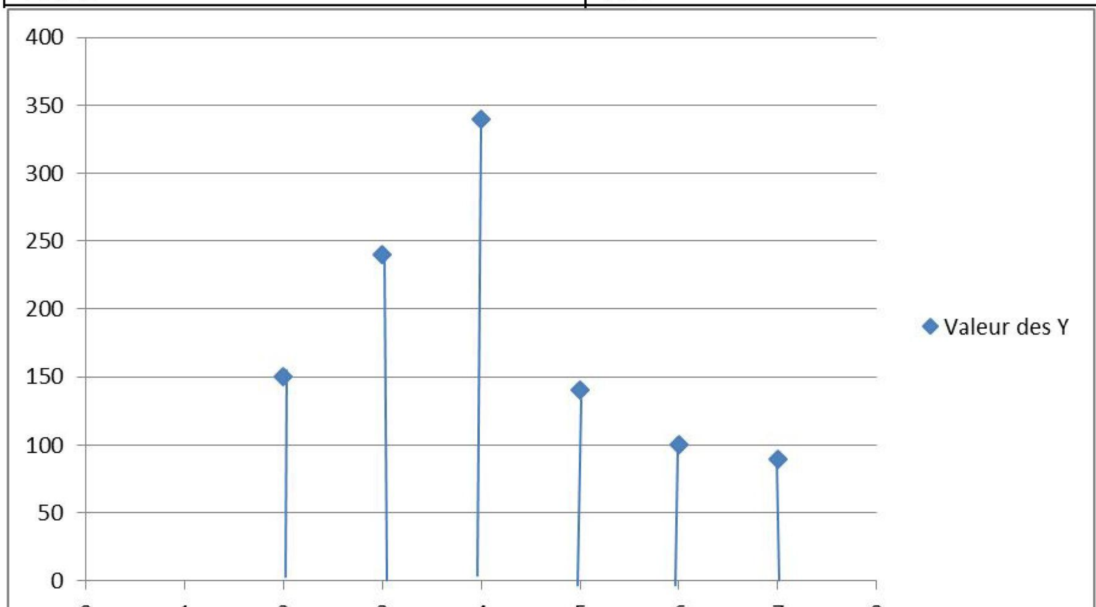
Cas d'une variable quantitative discrète

La représentation graphique des fréquences simples d'une variable discrète peut s'effectuer sous la forme de graphique en bâton. La valeur observée de la variable est portée sur l'axe des abscisses, et la fréquence simple correspondante est portée sur l'axe des ordonnées, cette dernière peut être exprimé en valeur relative ou en valeur absolue
Le bâton est un segment de droite parallèle à l'axe des ordonnées, de longueur proportionnelle à la fréquence correspondante.

Exemple

La distribution des étudiants en fonction de nombre de personnes dans la famille

Xi	Ni
2	150
3	240
4	340
5	140
6	100
Plus de 6	90
TOTAL	1060



Cas d'une variable statistique continue

La représentation graphique des fréquences simples d'une variable continue peut s'effectuer sous la forme d'un histogramme, en portant sur l'axe des abscisses les valeurs des classes du caractères, et en leur donnant les fréquences correspondantes

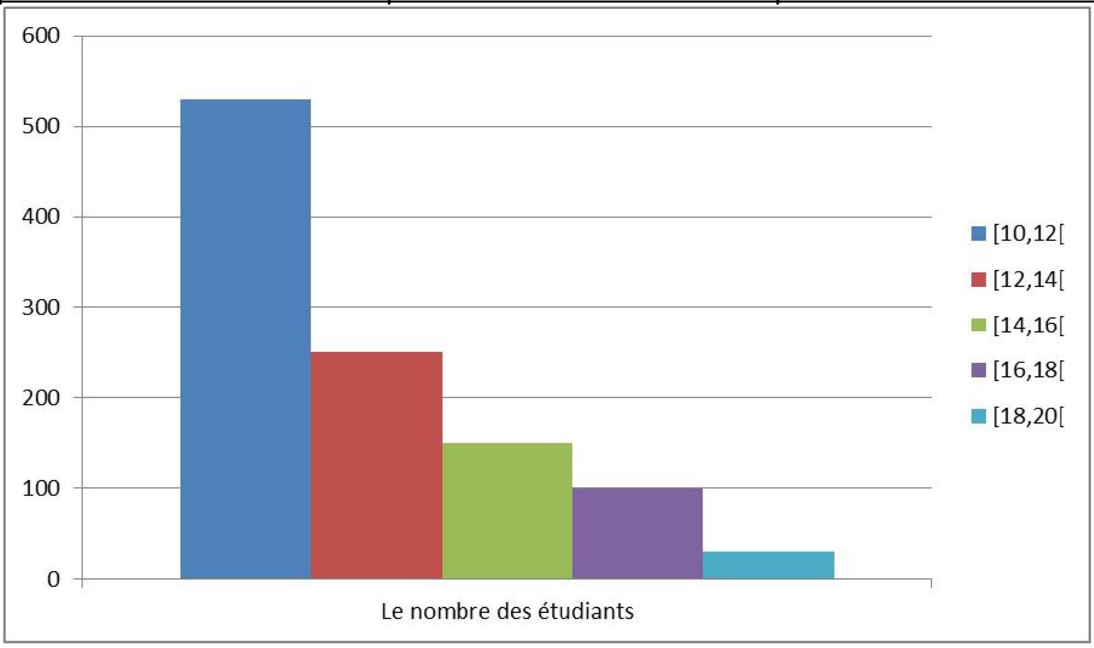
Principe de construction de l'histogramme

Pour chaque classe, on élève un rectangle ayant une base proportionnelle à l'intervalle de classe, donc ce sont les surfaces des rectangles qui ont une proportionnalité à l'effectif et non pas les hauteurs.
Dans la pratique deux cas peuvent se présenter
Cas d'amplitude égale :

Exemple

La distribution des étudiants en fonction de la note du bac

Les notes	Le nombre des étudiants	Amplitudes
[10,12[	530	2
[12,14[	250	2
[14,16[	150	2
[16,18[	100	2
[18,20[	30	2
TOTAL	1060	



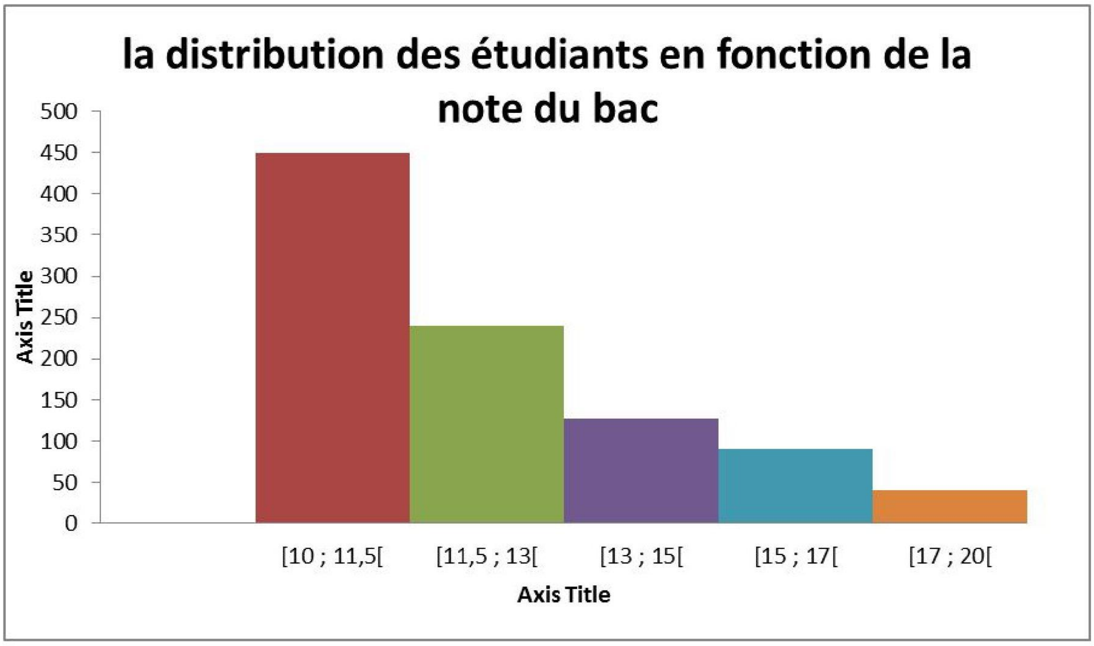
Cas des amplitudes inégales

Dans ce cas on respecte la proportionnalité des surfaces, il faut rectifier en conséquence les hauteurs.
Supposant que la distribution précédente se présente de la façon suivante

Les notes	Le nombre des étudiants	Amplitudes
[10 ;11,5[	450	1,5
[11,5 ;13[	240	1,5
[13 ; 15[	170	2
[15 ; 17[	120	2
[17 ; 20[	80	3
TOTAL	1060	

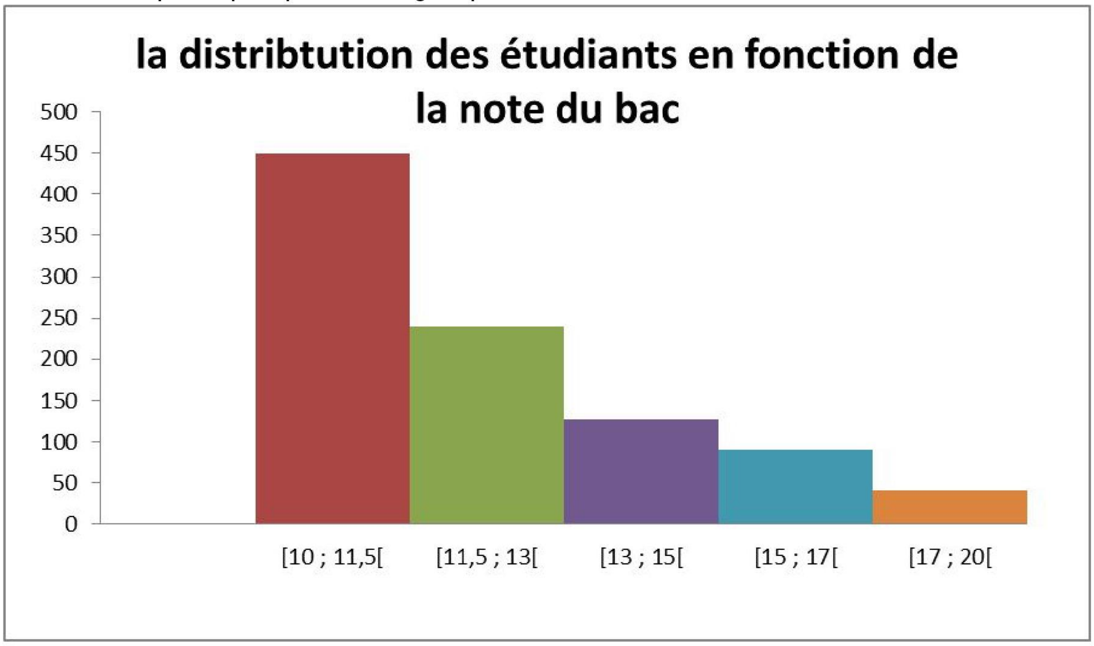
Alors, il est nécessaire de corriger les effectifs, en utilisant la formule suivante
ni corrigé=ni*an/ai
Avec : an est l'amplitude normale, c'est-à-dire celle qui s'est répétée le plus
Pour notre exemple on va considérer l'amplitude normale est 1,5

Les notes du bac	Le nombre des étudiants	amplitudes	Les effectifs corrigés
[10 ; 11,5[	450	1,5	450*1,5/1,5=450
[11,5 ; 13[	240	1,5	240*1,5/1,5=240
[13 ; 15[	170	2	170*1,5/2=127,5
[15 ; 17[	120	2	120*1,5/2=90
[17 ; 20[	80	3	80*1,5/3=40
Total	1060		



Le polygone des fréquences

Le polygone des fréquences est obtenus en joignant par des segments de droites au milieu des bases supérieures des rectangles permet de rendre compte de la continuité de la variable
A titre d'exemple on peut prendre la figure précédente



Graphiques de fréquences cumulées, diagramme intégral

Cas d'une variable statistique discrète

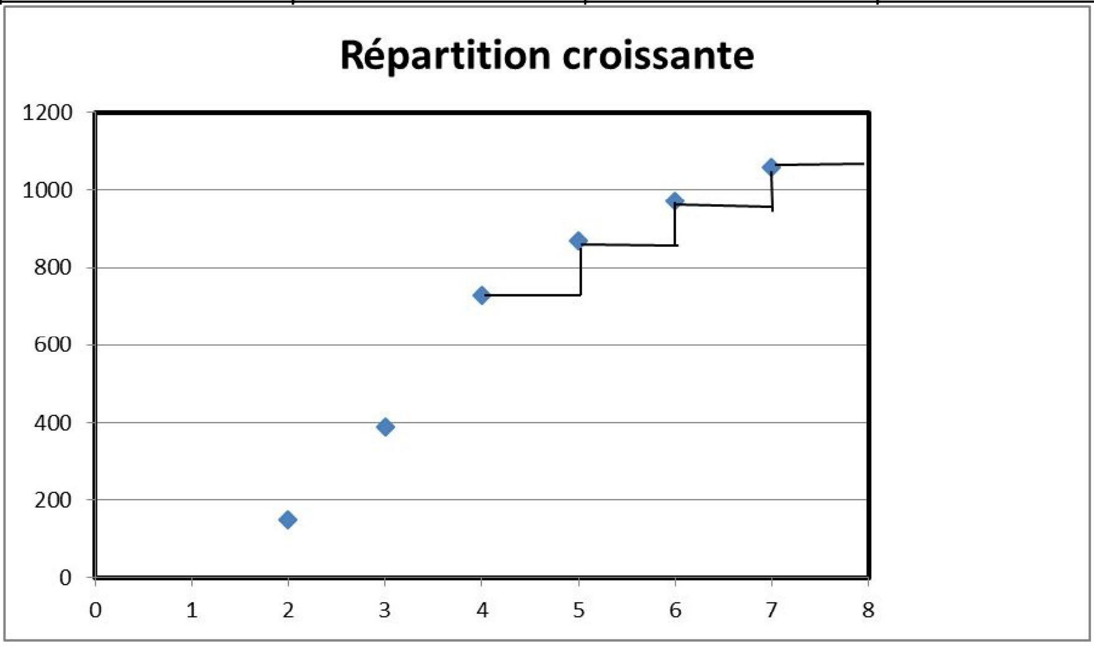
La représentation graphique des fréquences cumulées d'une variable discrète s'effectue sous la forme de graphique en escalier. Les sauts correspondent aux valeurs possibles de la variable et ils sont égaux aux fréquences cumulées croissantes ou décroissantes

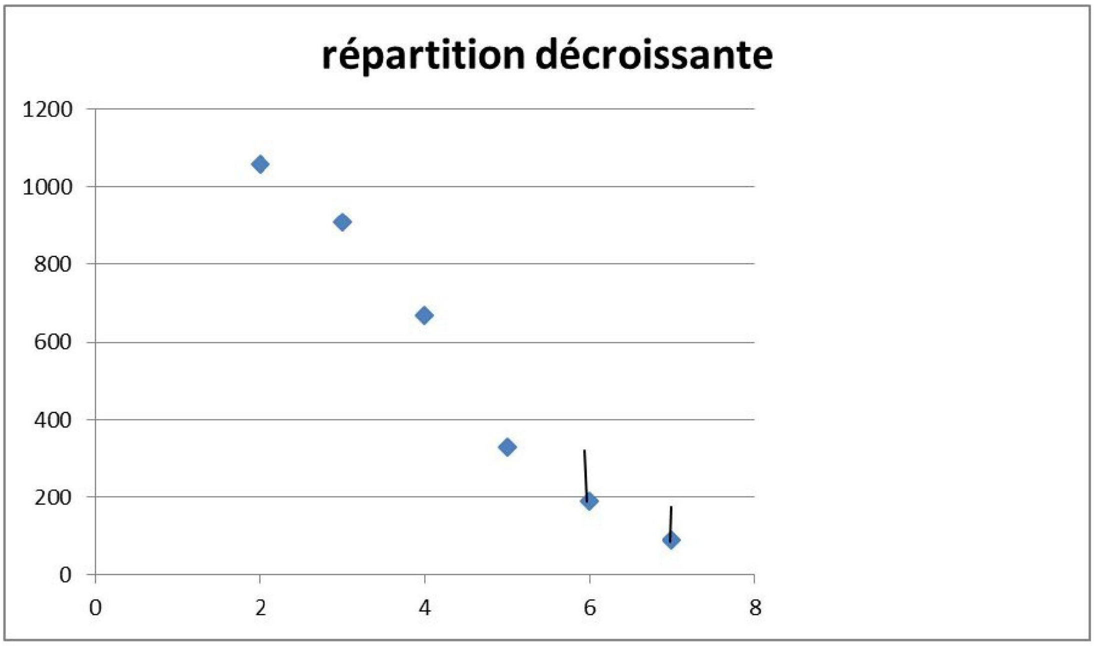
Exemple

La répartition des étudiants en fonction de nombre de personnes dans la famille

Le nombre de personnes dans la famille	Le nombre des étudiants	Effectifs cumulés croissants	Effectifs cumulés décroissant
2	150	150	1060
3	240	390	910
4	340	730	670

5	140	870	330
6	100	970	190
PLUS DE 6	90	1060	90
TOTAL	1060		





Cas d'une variable statistique continue

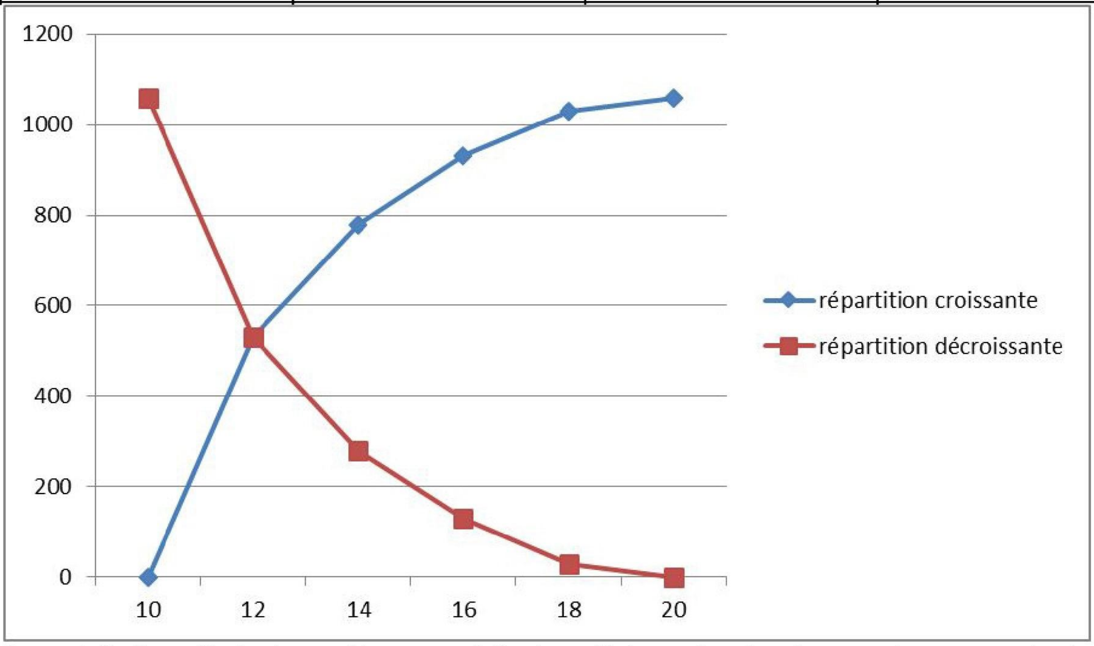
La courbe des fréquences cumulées croissantes se construit en portant les points correspondants à chaque classe à la limite supérieure de l'intervalle de classe.
La courbe des fréquences cumulées décroissantes se construit en portant les points correspondants à chaque classe à la limite inférieure de l'intervalle de classe.
La présence des classes d'amplitudes inégales n'entraine aucune modification en ce qui concerne la construction de cette courbe.

Exemple

La répartition des étudiants en fonction des notes du baccalauréat

Les notes du bac	Le nombre des étudiants	Effectifs cumulés croissants	Effectifs cumulés décroissant
[10,12[	530	530	1060
[12,14[	250	780	530
[14,16[	150	930	280
[16,18[	100	1030	130
[18,20[	30	1060	30

TOTAL	1060		



On contrôle l'exactitude du graphique, en vérifiant que l'intersection des deux courbes a pour abscisse la moitié de l'effectif total.