Statistique descriptive à une variable

✓ Introduction

La statistique est une méthode scientifique qui recueille, ordonne, analyse et interprète des données numériques ; afin d'avoir une meilleure lisibilité ces données sont représentées graphiquement. On utilise la statistiques dans plusieurs domaines parmi ces domaines : la biologie, l'économie, l'agronomie,...

✓ Vocabulaire

L'étude statistique ce porte sur un ensemble dite **population (**si notre population est très importante dans ce cas on prélève un **échantillon représentatif**) ces composantes (éléments) sont appelés **individu** ou **unité statistique** ($\boldsymbol{le\ nombre\ d'\text{é}tudiants\ n_i\ associ\text{é}s\ \text{à}\ la\ modalit\text{é}\ i}$)**,** la taille de la population est notée ***N***

Chaque individu se diffère d'un autre, cette différence s'appelle **un caractère** ou **une variable statistique** et chaque différence c'est une **modalité**
($\boldsymbol{la\ modalit\text{é}\ i\ est\ not\text{é}e\ x_i}$)

1) Ces modalités peuvent être mesurable dite **variable quantitative :**
   a) discrète ces valeurs sont isolées ou
   b) continue ces valeurs sont prises sur un intervalle.
   ou
2) ces valeurs ne sont pas mesurable dans ce cas c'est un **variable qualitative**.

Pour éclaircir, on prend quelques exemples.

Quelques exemples :

1) L'ensemble des étudiants au département MI est une population.
   a) Chaque étudiant est une unité statistique.
   b) On associé à chaque unité (étudiant) un caractère.
      i) Caractère de nature qualificatif, exemple : la couleur de la peau, la couleur des yeux, son adresse, la filière du baccalauréat…
      ii) Caractère de nature quantitatif (de type discrètes ou continue), exemple : les notes obtenus, les moyennes des étudiants, la tailles des étudiants en centimètres…

✓ Quelques définitions :

Effectif total : $N = n_1 + n_2 + \cdots + n_p = \sum_{i=1}^{p} n_i$, tel que p c'est le nombre de modalité.

La fréquence relative de la modalité i :

1) $f_i = \frac{n_i}{N}$, $0 \leq f_i \leq 1$ $1 \leq i \leq p$ et $\sum_{i=1}^{p} f_i = 1$.
2) Pour calculer le pourcentage des effectifs, on utilise les fréquences relatives en pourcentage c-à-d : $f_i\% = \frac{n_i}{N} \times 100$

Les données peuvent êtres données sous forme de liste ou de tableau, on réorganise nos données dans un tableau.

Exemples:

1) On a ci-dessous les notes de dix étudiants au module algèbre1 du groupe 3:

10, 11.5, 15, 10, 10, 11.5, 9, 9, 9, 9.

La population étudié c'est les étudiants du groupe 3.

Le caractère étudié c'est les notes du module algèbre 1, on remarque bien les notes sont des nombres donc notre variable est quantitative et ces valeurs sont isolés donc on est dans le cas discret.

Les notes x_i	9	10	11.5	15	Total
L'effectif n_i	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total
La fréquence relative f_i	$\frac{4}{10} = 0.4$	$\frac{3}{10} = 0.3$	$\frac{2}{10} = 0.2$	$\frac{1}{10} = 0.1$	$\sum_{i=1}^{4} f_i = 1$

- On a 4 modalités.

2) On a la taille en centimètres de 10 étudiants du groupe 3 :

La taille	[150,155[	[155 ,160[	[160,165[	[165,170[	Total
L'effectif n_i	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total
La fréquence relative en pourcentage $f_i\%$	$\frac{4}{10} \times 100 = 40$	$\frac{3}{10} \times 100 = 30$	$\frac{2}{10} \times 100 = 20$	$\frac{1}{10} \times 100 = 10$	$\sum_{i=1}^{4} f_i\,\% = 100\%$

On remarque bien, que nos modalités sont écrites sous la forme d'intervalle dite **classe**, donc notre caractère est quantitative continue.

Remarque : dans ce type de série la première des choses on calcule le centre de chaque classe. Une classe est écrite sous la forme suivante: $[a_i, b_i[$, le centre de cette classe est calculé comme suit :

$$c_i = \frac{a_i + b_i}{2}$$

On va réécrire, de nouveau notre tableau :

La taille	[150,155[	[155 ,160[	[160,165[	[165,170[	total
L'effectif n_i	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total
La fréquence relative en pourcentage $f_i\%$	$\frac{4}{10} \times 100 = 40$	$\frac{3}{10} \times 100 = 30$	$\frac{2}{10} \times 100 = 20$	$\frac{1}{10} \times 100 = 10$	$\sum_{i=1}^{4} f_i\,\% = 100\%$
Le centre des classes c_i	$\frac{150 + 155}{2} = 152.5$	$\frac{160 + 155}{2} = 157.5$	$\frac{160 + 165}{2} = 162.5$	$\frac{170 + 165}{2} = 167.5$	

✓ Les paramètres d'une série statistique

Une série statistique est l'ensemble des couples $\{(x_i\ ,\ n_i)\}$ où x_i est la modalité i avec l'effectif n_i.

1) Les paramètres de position :

1.1) **Le mode « Mo » :** pour trouver la valeur du mode, tout d'abord on cherche la valeur la plus grande des effectifs parmi toutes les valeurs données ;

1.1.a) dans le cas d'une série statistique discrète, on cherche la variable qui lui corresponde.

1.1.b) mais dans le cas d'une série statistique continue, on cherche la classe qui lui corresponde, le centre de cette classe c'est la valeur du mode.

Remarque : en général, le mode ce n'est pas une valeur unique, si notre série à un seule mode donc c'est une série dite **uni-modale**, si on a deux valeurs du mode dans ce cas c'est une série statistique **bimodales** si non **multimodales.**

Exemple :

1)

Les notes x_i	9	10	11.5	15	total
L'effectif n_i	4	3	2	1	$N=\sum_{i=1}^{4} n_i$ =10 Effectif total

On remarque bien, la plus grande des effectifs c'est 4 , la valeur de la modalité qui lui correspond c'est **9**, donc le mode : **Mo=9.**

2)

La taille	[150,155[	[155 ,160[	[160,165[	[165,170[	total
L'effectif n_i	4	3	2	1	$N=\sum_{i=1}^{4} n_i$ =10 Effectif total

On remarque bien, la plus grande des effectifs c'est 4 , la valeur de la classe modale qui lui correspond c'est [150,155[, donc le mode:

$\mathrm{Mo}=\frac{150+155}{2}=152.5.$

Remarque :

- Une série statistique continue dite uniforme si les amplitudes de toutes les classes de cette série sont égales.
- L'amplitude de la classe $[a_i, b_i[$ est calculée comme suit : $e_i = b_i - a_i$.
- Pour calculer le mode dans une série continue uniforme, on cherche directement la plus grande valeur des effectifs et la valeur du mode c'est le centre de la classe modale qui lui correspond.
- Mais pour une série statistique avec des amplitudes différentes, au lieu de chercher l'effectif on cherche l'effectif corrigé (on va voir un exemple explicatif pendant la révision).

1.2) La moyenne arithmétique pondérée $\overline{X}$: (on dit simplement la moyenne arithmétique)

1.2.a) dans le cas d'une série statistique discrète, la moyenne arithmétique se calcule comme suit :

$$\overline{X}=\frac{n_1x_1+n_2x_2+\cdots+n_px_p}{n_1+n_2+\cdots+n_p}=\frac{\sum_{i=1}^{p} n_ix_i}{\sum_{i=1}^{p} n_i}=\frac{\sum_{i=1}^{p} n_ix_i}{N}\neq\sum_{i=1}^{p}\frac{n_ix_i}{n_i}$$

Ou, on utilise les fréquences relatives donc :

$$\overline{X}=\frac{\sum_{i=1}^{p} n_ix_i}{N}=\sum_{i=1}^{p}\frac{n_ix_i}{N}=\sum_{i=1}^{p}\frac{n_i}{N}x_i=\sum_{i=1}^{p} f_i\, x_i$$

1.2.b) dans le cas d'une série statistique continue, la moyenne arithmétique se calcule comme suit :

$\overline{X}=\frac{n_1c_1+n_2c_2+\cdots+n_pc_p}{n_1+n_2+\cdots+n_p}=\frac{\sum_{i=1}^{p} n_ic_i}{\sum_{i=1}^{p} n_i}=\frac{\sum_{i=1}^{p} n_ic_i}{N}\neq\sum_{i=1}^{p}\frac{n_ic_i}{n_i}$, où c_i c'est le centre des classes.

Ou, on utilise les fréquences relatives donc :

$$\bar{X}=\frac{\sum_{i=1}^{p} n_i c_i}{N}=\sum_{i=1}^{p}\frac{n_i c_i}{N}=\sum_{i=1}^{p}\frac{n_i}{N}c_i=\sum_{i=1}^{p} f_i\, c_i$$

Remarque : y'a d'autre moyenne, mais pour le moment on va voir seulement la moyenne arithmétique.

Exemple : on calcule la moyenne avec les fréquences et les effectifs

1) Pour simplifier les calcules on rajoute une ligne

Les notes x_i	9	10	11.5	15	total
L'effectif n_i	4	3	2	1	$N=\sum_{i=1}^{4} n_i$ =10 Effectif total
$n_i x_i$	36	30	23	15	$\sum_{i=1}^{4} n_i\, x_i = 104$
f_i	0.4	0.3	0.2	0.1	1
$f_i x_i$	3.6	3	2.3	1.5	$\sum_{i=1}^{4} f_i\, x_i = 10{,}4$

$\bar{X}=\frac{\sum_{i=1}^{p} n_i x_i}{N}=\frac{104}{10}=10.4$ ou $\bar{X}=\sum_{i=1}^{p} f_i\, x_i=10.4.$

2)

La taille	[150,155[	[155 ,160[	[160,165[	[165,170[	total
L'effectif n_i	4	3	2	1	$N=\sum_{i=1}^{4} n_i$ =10 Effectif total
c_i	152.5	157.5	162.5	167.5	
$n_i c_i$	610	472.5	325	167.5	1575
f_i	0.4	0.3	0.2	0.1	1
$f_i x_i$	61	47.25	32.5	16.75	157.5

$\bar{X}=\frac{\sum_{i=1}^{p} n_i c_i}{N}=\frac{1575}{10}=157.5$ ou $\bar{X}=\sum_{i=1}^{p} f_i\, c_i=157.5.$

1.3) **Les quantiles :** on va voir seulement **la médiane « Me »**.

- Dans le cas discret on utilise les effectifs cumulés croissant (**E.C.C)**

x_i	x_1	x_2	....	x_p	total
L'effectif n_i	n_1	n_2	....	n_p	$N=\sum_{i=1}^{p} n_i$
ECC	n_1	n_1+n_2	....	$N=\sum_{i=1}^{p} n_i$	

Deuxième des choses on calcule le rang, on a deux cas :

a) Si *N* est impair, le rang de la médiane c'est $\frac{N+1}{2}$ cette valeur on va la trouver dans la dernière ligne, après l'avoir trouver on cherche la valeur de la médiane

$$Me = x_{\frac{N+1}{2}}$$

b) Si N est pair, la médiane se trouve dans l'intervalle [$x_{\frac{N}{2}}$, $x_{\frac{N}{2}+1}$]donc la valeur de la médiane est :

$$Me = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$$

-
- Dans le cas continu on utilise les effectifs cumulés croissant (**E.C.C)** ou les fréquences cumulées croissantes **(F.C.C)** ou les fréquences cumulées croissantes en pourcentage **(F.C.C %)**

Les classes	$[a_1, b_1[$	$[a_2, b_2[$	…… ….	$[a_p, b_p[$	total
L'effectif $\boldsymbol{n_i}$	$\boldsymbol{n_1}$	$\boldsymbol{n_2}$	…… …..	$\boldsymbol{n_p}$	$N = \sum_{i=1}^{p} n_i$
$\boldsymbol{ECC}$	$\boldsymbol{n_1}$	$\boldsymbol{n_1 + n_2}$	…… ……	$N = \sum_{i=1}^{p} n_i$	
$\boldsymbol{f_i}$	$\boldsymbol{f_1}$	$\boldsymbol{f_2}$	…… ……	$\boldsymbol{f_p}$	1
F.C.C	$\boldsymbol{f_1}$	$\boldsymbol{f_1 + f_2}$	…… ……	1	
$\boldsymbol{f_i}$%	$\boldsymbol{f_1}$***100**	$\boldsymbol{f_2}$***100**	…… ……	$\boldsymbol{f_p}$***100**	100
F.C.C %	$\boldsymbol{f_1}$***100**	$\boldsymbol{f_1 * 100 + f_2 * 100}$	…… ……	100	

Deuxième des choses on calcule La position, mais on ne va pas s'intéresser au cas pair et impair:

a) Si on utilise , donc le rang de la médiane c'est $\frac{N}{2}$ cette valeur on va la trouver dans la dernière ligne, après l'avoir trouver on utilise l'interpolation affine ou bien linéaire
b) Si on utilise , donc le position de la médiane c'est 0.5, on utilise l'interpolation affine.
c) Si on utilise $\%$, donc le position de la médiane c'est 50%, on utilise l'interpolation affine.

Exemple :

1)

Les notes $\boldsymbol{x_i}$	9	10	11.5	15	total
L'effectif $\boldsymbol{n_i}$	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total
$\boldsymbol{ECC}$	4	4+3=7	4+3+2=9	4+3+2+1=10	

On remarque **N=10** est pair donc $Me \in [x_{\frac{10}{2}}$, $x_{\frac{10}{2}+1}$]donc la valeur de la médiane est :

$$Me = \frac{x_{\frac{10}{2}} + x_{\frac{10}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{10+10}{2} = 10$$

2)

Les notes x_i	9	10	11.5	15	total
L'effectif n_i	4	3	2	2	$N = \sum_{i=1}^{4} n_i$ =11 Effectif total
ECC	4	4+3=7	4+3+2=9	4+3+2+2=11	

On a « **N**=11 » est impair, le rang de la médiane c'est $\frac{11+1}{2}$ =6 , donc la valeur de la médiane

$$Me = x_6 = 10$$

3)

La taille	[150,155[	[155 ,160[	[160,165[	[165,170[	total
L'effectif n_i	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total
ECC	4	4+3=7	7+2=9	9+1=10	
f_i	0.4	0.3	0.2	0.1	1
FCC	0.4	0.4+0.3=0.7	0.4+0.3+0.2=0.9	0.9+0.1=1	
$f_i\%$	40	30	20	10	100
FCC%	40	40+30=70	70+20=90	90+10=100	

Remarque : Ces valeurs ne change jamais .

- $\frac{10}{2} = 5$ donc la médiane, donc $Me \in [155{,}160[$

155
Me
160

4
5
7

$$\frac{Me - 155}{160 - 155} = \frac{5-4}{7-4}$$

Et on va résoudre une équation du premier degré :

$$Me = (160 - 155)\frac{5-4}{7-4} + 155 \approx 156.67$$

- 0.5 donc la médiane, donc $Me \in [155{,}160[$

155
Me
160

0.4
0.5
0.7

$$\frac{Me - 155}{160 - 155} = \frac{0.5-0.4}{0.7-0.4}$$

Et on va résoudre une équation du premier degré :

$$Me \approx 156.67$$

- 50% donc la médiane, donc $Me \in [155{,}160[$

155
Me

40
50

160	70

$$\frac{Me-155}{160-155}=\frac{50-40}{70-40}$$

Et on va résoudre une équation du premier degré :

$$Me \approx 156.67$$

2) Paramètres de dispersions :

2.1) Etendue

L'étendue d'une série statistique, notée e est la différence entre la plus grande valeur $x_{\max}$ et la plus petite valeur $x_{\min}$ du caractère : $e = x_{max} - x_{min}$.

Exemple :

1)

Les notes $\boldsymbol{x_i}$	9	10	11.5	15	total
L'effectif $\boldsymbol{n_i}$	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total

$$e = x_{max} - x_{min} = 15 - 9 = 6$$

2)

La taille	[150,155[	[155 ,160[	[160,165[	[165,170[	total
L'effectif $\boldsymbol{n_i}$	4	3	2	1	$N = \sum_{i=1}^{4} n_i$ =10 Effectif total

$$e = x_{max} - x_{min} = 170 - 150 = 20$$

2.2) la variance et l'écart type

2.2.a) la variance : c'est la moyenne des carrées des écarts à la moyenne arithmétique

- dans le cas discret :

$$v(x) = \sum_{i=1}^{p} \frac{n_i(x_i-\bar{X})^2}{N} = \sum_{i=1}^{p} \frac{n_i x_i{}^2 - 2n_i x_i \bar{X} + n_i \bar{X}^2}{N} = \sum_{i=1}^{p} \frac{n_i x_i{}^2}{N} - 2\sum_{i=1}^{p} \frac{n_i x_i \bar{X}}{N} + \sum_{i=1}^{p} \frac{n_i (\bar{X})^2}{N} = \overline{X^2} - 2(\bar{X})^2 + (\bar{X})^2 = \overline{X^2} - (\bar{X})^2.$$

- Dans le cas continue $v(x) = \sum_{i=1}^{p} \frac{n_i(c_i-\bar{X})^2}{N}$ $=\overline{X^2} - (\bar{X})^2$

Cette formule $v(x)=\overline{X^2} - (\bar{X})^2$, c'est la formule de **König**

2.2.b) l'écart type $\sigma(x) = \sqrt{v(x)}$

Remarque : L'écart type mesure la dispersion des x_i autour de $\bar{X}$, et il s'exprime dans la même unité que le caractère étudié.

Contrairement à la variance qui s'exprime dans l'unité au carré du caractère.

✓ La représentation graphique :

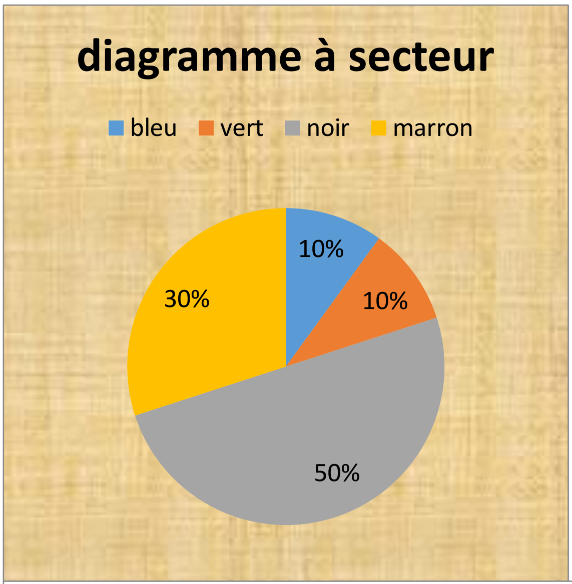
la représentation graphique sa dépendra de la nature de notre série statistique .

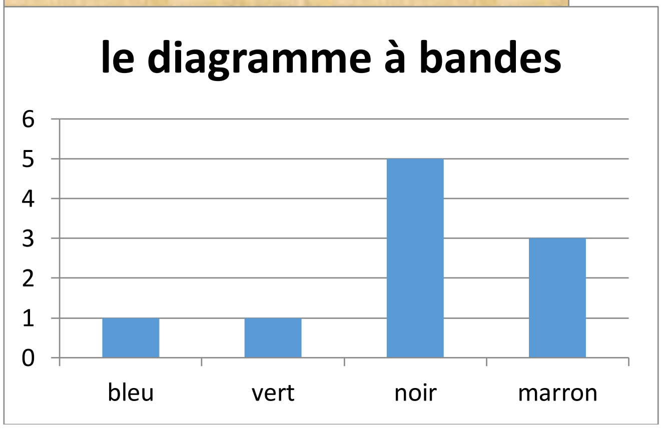
I. Dans le cas **qualificatif**, on peut utiliser deux types de diagramme : **le diagramme à secteur**

ou **le diagramme à bandes.**

<u>Exemple</u> on a pris dix personne et on a étudié la couleur des yeux, et on eu le tableau suivant :

La couleur des yeux	**effectif**	f_i	**Les mesures des secteurs**
bleu	1	0.1	0.1*360=36
vert	1	0.1	0.1*360=36
noir	5	0.5	0.5*360=180
marron	3	0.3	0.3*360=108
total	10	1	360°





II. Dans le cas **quantitatif discret**, on peut utiliser **le diagramme en bâtons.**

<u>Exemple :</u>

Les notes x_i	9	10	11	12	13	total
L'effectif n_i	1	1	4	3	1	N =10

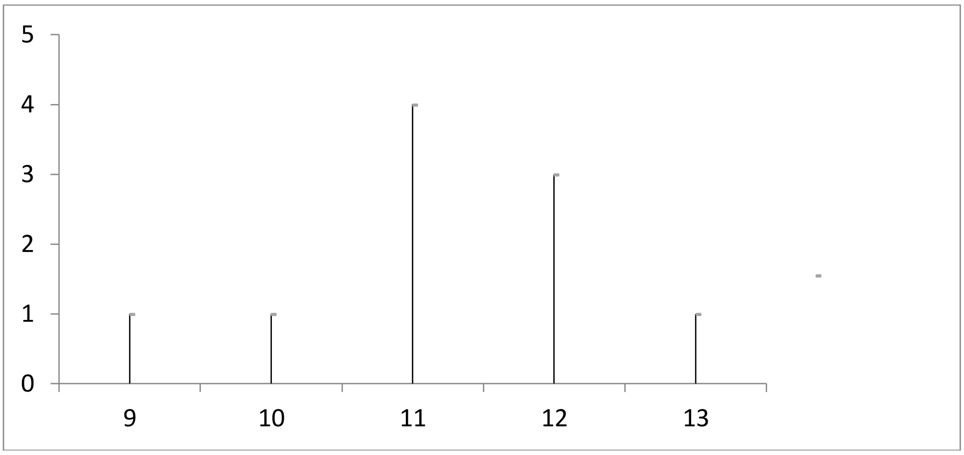


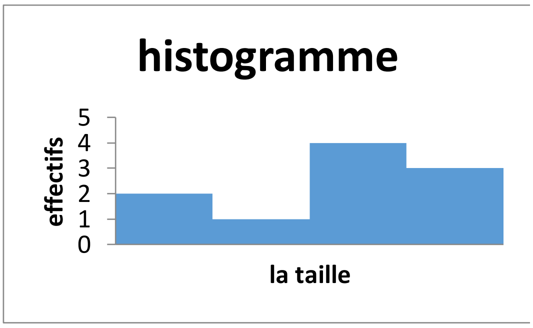
Diagramme en bâtons

III. Dans le cas **quantitatif continu**, L'histogramme est un diagramme composé de rectangles dont l'aire est proportionnelle à l'effectif de chaque classe. Il faut envisager le cas où les amplitudes des classes sont égales et le cas où ces amplitudes sont inégales.

Pour le moment on prendre seulement les séries à amplitudes avec classes égales.

Exemple :

La taille	effectif
[150,155[	2
[155,160[	1
[160,165[	4
[165,170[	3
total	10



.