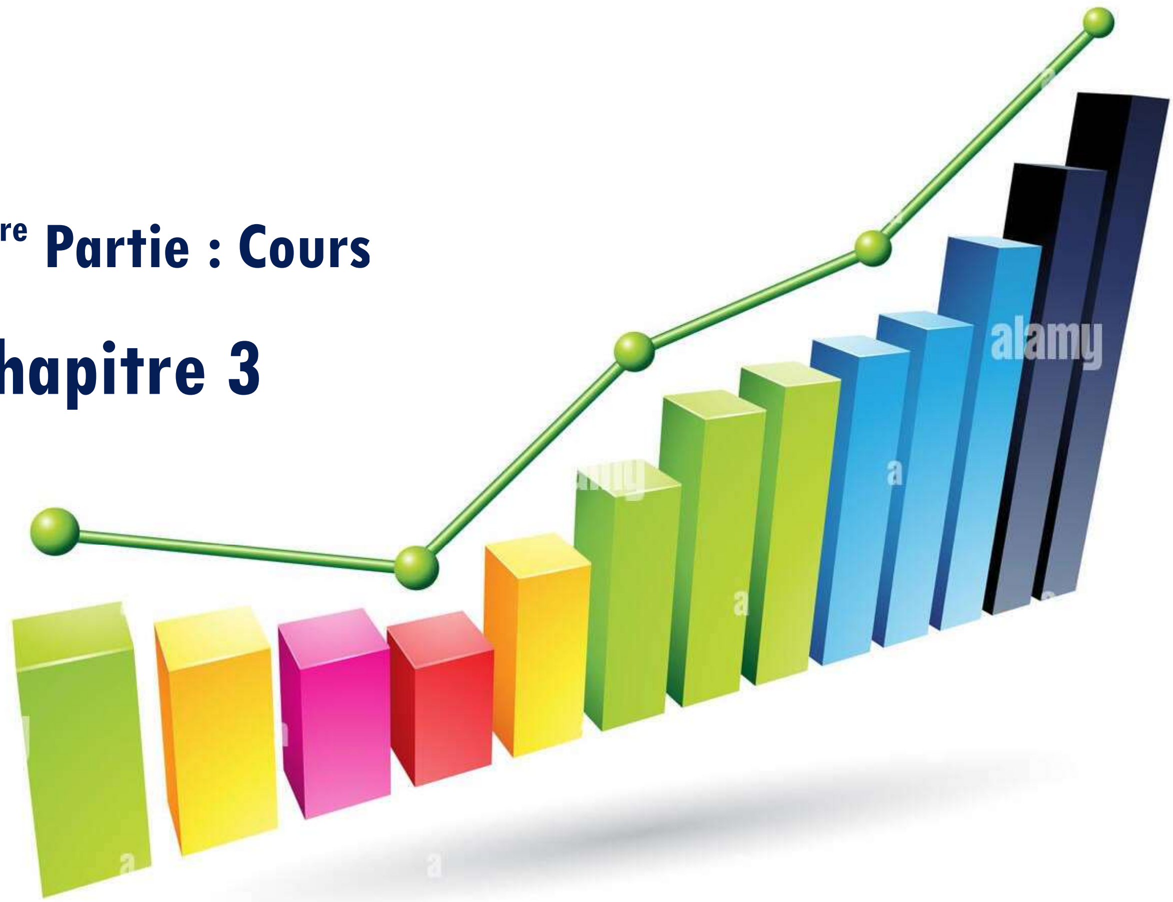
1ère Partie : Cours

Chapitre 3



Statistique descriptive bivariée

On s'intéresse à deux variables X et Y. Ces deux variables sont mesurées simultanément sur les mêmes individus. On obtient donc deux mesures. La série statistique est alors une suite de couples des valeurs prises par les deux variables sur chaque individu.
Chacune des deux variables peut être, soit quantitative, soit qualitative. On examine juste le cas des deux variables sont quantitatives.
Pour bien voir ce genre de série statistique on va commencer par un exemple et par la suite on généralise.

Exemple 3.0.1

On considère le tableau suivant, relatif à une population de 100 ménages, tel que :
$X =$ « le nombre d'enfants du ménage » et $Y =$ « le nombre de pièces du logement »

$X \backslash Y$	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	Total
$x_1 = 2$	10	8	05	23
$x_2 = 3$	39	5	10	54
$x_3 = 4$	8	5	0	13
$x_4 = 5$	10	0	0	10
Total	67	18	15	100

- La valeur 39 indique que, parmi les 100 ménages observés, il y'a 39 ménages qui ont 3 enfants et qui habitent dans des logements de 3 pièces.
- La valeur 67 indique que, parmi les 100 ménages observés, il y a 67 ménages qui habitent dans des logements de 3 pièces.
- La valeur 54 indique que, parmi les 100 ménages observés, il y a 54 ménages qui ont 3 enfants.

3.1 Distributions et caractéristiques

Soient X et Y deux variables mesurées sur N individus d'une population, Avec les modalités : $M(X) = \{x_1, x_2,\dots, x_r\}$, $M(Y) = \{y_1, y_2,\dots, y_k\}$.

3.1.1 Distribution conjointe de X et Y

C'est la liste des $r \times k$ modalités conjointes (x_i, y_j) associées chacune à son effectif n_{ij} ou à sa fréquence f_{ij} . Ce qui donne le tableau des contingents suivant :

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_k	Total
x_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2.}$
$\vdots$							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i.}$
$\vdots$							
x_r	n_{r1}	n_{r2}		n_{rj}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.k}$	N

- Les effectifs qui sont notés par n_{ij} est le nombre de fois où la modalité x_i de la variable X et la modalité y_j de la variable Y ont été observées simultanément.
- L'effectif $n_{i.}$ appelé **effectif marginal** de la variable X est le nombre total d'observations de la modalité x_i de la variable X.

- L'effectif $n_{.j}$ appelé **effectif marginal de** la variable Y est le nombre total d'observations de la modalité y_j de la variable Y.

3.1.2 Distributions marginales

La distribution marginale de X (respectivement de Y) est la distribution de X (respectivement de Y) sur l'échantillon, calculée à partir de la distribution conjointe.
Ces deux distributions peuvent se présenter sous forme de tableaux statistiques suivants :

Distribution marginale de X

X	*Effectif marginal*
x_1	$n_{1.}$
x_2	$n_{2.}$
$\vdots$	
x_i	$n_{i.}$
$\vdots$	
x_r	$n_{r.}$
Total	N

Distribution marginale de *Y*

Y	*Effectif marginal*
y_1	$n_{.1}$
y_2	$n_{.2}$
$\vdots$	
y_j	$n_{.j}$
$\vdots$	
y_k	$n_{.k}$
Total	N

Remarque 3.1.1 Pour deux variables X et Y mesurées sur N individus d'une population, la distribution conjointe se donne sous forme de tableaux des contingent des effectifs ou des fréquences comme suit :

Distribution conjointe en effectif de X et Y

$X \backslash Y$	y_1	y_2	...	y_j	...	y_k	Total
x_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2.}$
$\vdots$							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i.}$
$\vdots$							
x_r	n_{r1}	n_{r2}		n_{rj}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.k}$	N

Distribution conjointe en fréquence de X et Y

$X \backslash Y$	y_1	y_2	...	y_j	...	y_k	Total
x_1	f_{11}	f_{12}		f_{1j}		f_{1k}	$f_{1.}$
x_2	f_{21}	f_{22}		f_{2j}		f_{2k}	$f_{2.}$
$\vdots$							
x_i	f_{i1}	f_{i2}		f_{ij}		f_{ik}	$f_{i.}$
$\vdots$							
x_r	f_{r1}	f_{r2}		f_{rj}		f_{rk}	$f_{r.}$
Total	$f_{.1}$	$f_{.2}$		$f_{.j}$		$f_{.k}$	1

- L'effectif $n_{i.}$ appelé effectif marginal de X est le nombre total d'observations de la modalité x_i de la variable X.

$$n_{i.} = \sum_{j=1}^{j=k} n_{ij}$$

- L'effectif $n_{.j}$ appelé effectif marginal de Y est le nombre total d'observations de la modalité y_j de la variable Y.

$$n_{.j} = \sum_{i=1}^{i=r} n_{ij}$$

- L'effectif total de la distribution conjointe noté N, peut être obtenu à partir de l'effectif marginal de X ou bien à partir de l'effectif marginal de Y :

$$N = \sum_{i=1}^{i=r} n_{i.} = \sum_{j=1}^{i=k} n_{.j} = \sum_{i=1}^{i=r}\sum_{j=1}^{j=k} n_{ij}$$

- La fréquence conjointe, noté f_{ij} est $f_{ij} = \dfrac{n_{ij}}{N}$
- La fréquence $f_{i.}$ appelée **fréquence marginale** de X est le nombre

$$f_{i.} = \frac{n_{i.}}{N} = \sum_{j=1}^{j=k} f_{ij}$$

- La fréquence $f_{.j}$ appelée **fréquence marginale** de Y est le nombre

$$f_{.j} = \frac{n_{.j}}{N} = \sum_{i=1}^{i=r} f_{ij}$$

et $\sum_{i=1}^{i=r} f_{i.} = \sum_{j=1}^{i=k} f_{.j} = \sum_{i=1}^{i=r}\sum_{j=1}^{j=k} f_{ij} = 1$

- **Les moyennes marginales** et **les variances marginales**

Distribution marginale de X

X	Effectif marginal
x_1	$n_{1.}$
x_2	$n_{2.}$
$\vdots$	
x_i	$n_{i.}$
$\vdots$	
x_r	$n_{r.}$
Total	N

Distribution marginale de Y

Y	Effectif marginal
y_1	$n_{.1}$
y_2	$n_{.2}$
$\vdots$	
y_j	$n_{.j}$
$\vdots$	
y_k	$n_{.k}$
Total	N

Les moyennes marginales de X et de Y, ainsi que les variances marginales se calculent à partir des distributions marginales par les formules suivantes :

$$\bar{x} = \sum_{i=1}^{i=r} \frac{n_{i.}x_i}{N} = \sum_{i=1}^{i=r} f_{i.}x_i ; \qquad \bar{y} = \sum_{j=1}^{j=k} \frac{n_{.j}y_j}{N} = \sum_{j=1}^{j=k} f_{.j}y_j$$

$$V(x) = \frac{1}{N}\sum_{i=1}^{i=r} n_{i.}(x_i - \bar{x})^2 = \sum_{i=1}^{i=r} f_{i.}(x_i - \bar{x})^2 ;$$

$$V(y) = \frac{1}{N}\sum_{j=1}^{j=k} n_{.j}(y_j - \bar{y})^2 = \sum_{j=1}^{j=k} f_{.j}(y_j - \bar{y})^2$$

Exemple 3.1.1

On considère le tableau des effectifs suivant, relatif à une population de 20 adolescents, tel que : X = « la taille » et Y = « le poids ».

X \ Y	[40,50[	[50,70[	[70,90[	*Total*
[120,140[	2	1	0	3
[140,160[	2	6	0	8
[160,180[	1	3	5	9
Total	5	10	5	20

Donc la distribution conjointe en fréquences est : $f_{ij} = \frac{n_{ij}}{N}$

X \ Y	[40,50[	[50,70[	[70,90[	*Total*
[120,140[	0,10	0,05	0,00	0,15
[140,160[	0,10	0,30	0,00	0,40
[160,180[	0,05	0,15	0,25	0,45
Total	0,25	0,50	0,25	1,00

$\Rightarrow$

Distribution marginale en fréquence de X

X	$f_{i.}$
[120,140[	0,15
[140,160[	0,40
[160,180[	0,45
Total	1,00

et

Distribution marginale en fréquence Y

Y	$f_{.j}$
[40,50[	0,25
[50,70[	0,50
[70,90[	0,25
Total	1,00

De même on obtient à partir du même tableau les distributions marginales des effectifs.

En introduisant les centres des classes pour calculer les moyennes et les variances marginales :

Distribution marginale en effectif de X

X	$n_{i.}$	c_i	$n_{i.}\, c_i$	$n_{i.}\, c_i^2$
[120,140[	3	130	390	50700
[140,160[	8	150	1200	180000
[160,180[	9	170	1530	260100
Total	20		3120	490800

et

Distribution marginale en effectif Y

Y	$n_{.j}$	c_j	$n_{.j}\, c_j$	$n_{j.}\, c_j^2$
[40,50[	5	45	225	10125
[50,70[	10	60	600	36000
[70,90[	5	80	400	32000
Total	20		1225	78125

Donc les moyennes marginales de X et de Y sont :

$$\bar{x} = \frac{3120}{20} = 156 \quad cm \text{ et } \bar{y} = \frac{1225}{20} = 61{,}25 \quad kg$$

les variances marginales de X et de Y sont :

$$V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=3} n_{i.}c_i^{\,2}\right) - \bar{x}^2 = \frac{490800}{20} - (156)^2 = 24540 - 24336 \Rightarrow V(x) = 204.$$

$$\text{et } V(y) = \left(\frac{1}{N}\sum_{j=1}^{j=3} n_{.j}c_j^{\,2}\right) - \bar{y}^2 = \frac{78125}{20} - (61{,}25)^2 = 154{,}69$$