

Chapitre 4

Séries doubles

I. Tableau de contingence

On veut étudier la répartition de 100 étudiants selon le nombre d'année d'obtention de la licence et la mention. Soient :

- X : Nombre d'année d'obtention de la licence
- Y : Mention de la licence

X\Y	Passable (P)	Assez Bien (AB)	Bien (B)	Très Bien (TB)	Marge de X
3	20	10	10	5	45
4	15	10	5	5	35
5	10	5	5	0	20
Marge de Y	45	25	20	10	100

I. Tableau de contingence

X\Y	Passable (P)	Assez Bien (AB)	Bien (B)	Très Bien (TB)	Marge de X
3	20	10	10	5	45
4	15	10	5	5	35
5	10	5	5	0	20
Marge de Y	45	25	20	10	100



1. La distribution marginale de la variable **X** est :

X	3	4	5
Marge de X $n_{i\bullet}$	45	35	20

Le nombre $n_{i\bullet}$ ($i = 1, 2, 3$), est appelé la marge de X ou l'effectif marginal de X.

I. Tableau de contingence

X\Y	Passable (P)	Assez Bien (AB)	Bien (B)	Très Bien (TB)	Marge de X
3	20	10	10	5	45
4	15	10	5	5	35
5	10	5	5	0	20
Marge de Y	45	25	20	10	100



2. La distribution marginale de la variable Y est :

Y	P	AB	B	TB
Marge de Y $n_{\bullet j}$	45	25	20	10

Le nombre $n_{\bullet j}$ ($j = 1, 2, 3, 4$), est appelé la marge de Y ou l'effectif marginal de Y.

I. Tableau de contingence

Soient X et Y deux variables statistiques dont les valeurs sont:

$\square x_1, x_2, \dots, x_m$ et $\square y_1, y_2, \dots, y_k$

Les données de ces deux variables peuvent être regroupées dans un tableau dit tableau de contingence.

$X \backslash Y$	y_1	y_2	\dots	y_k	Marge de X $n_{i\bullet}$
x_1	n_{11}	n_{12}	\dots	n_{1k}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	\dots	n_{2k}	$n_{2\bullet}$
\dots	\dots	\dots	\dots	\dots	\dots
x_m	n_{m1}	n_{m2}	\dots	n_{mk}	$n_{m\bullet}$
Marge de Y $n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	\dots	$n_{\bullet k}$	N

I. Tableau de contingence

□ Le nombre n_{ik} est l'effectif concernant à la fois les modalités x_i et y_j , c'est-à-dire la modalité du couple (x_i, y_j) , où $i = 1, \dots, m$ et $j = 1, \dots, k$.

□ $n_{i\bullet} = \sum_{p=1}^k n_{ip}$ correspond au nombre d'individus ayant les modalités x_i où $i = 1, \dots, m$.

□ $n_{\bullet j} = \sum_{p=1}^m n_{pj}$ correspond au nombre d'individus ayant les modalités y_j où $j = 1, \dots, k$.

□ $N = \sum_{j=1}^k n_{\bullet j} = \sum_{i=1}^m n_{i\bullet} = \sum_{j=1}^k \sum_{k=1}^n n_{kj} = \sum_{i=1}^m \sum_{k=1}^n n_{ik}$ correspond à l'effectif total.

I. Tableau de contingence

Soient X et Y deux variables statistiques qui désignent respectivement le poids (Kg) et la taille (Cm) de 100 animaux.

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[2	4	2	5	1	14	
[5,10[4	5	5	2	2	18	
[10,15[2	10	5	15	4	36	
[15,20[4	4	2	1	1	12	
[20,25[5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	

I. Tableau de contingence

- Fréquence partielle du couple (x_i, y_j) où $i = 1, \dots, m$ et $j = 1, \dots, k$.
est donnée par :

$$f_{ij} = \frac{n_{ij}}{n}$$

- Fréquence marginale de x_i où $i = 1, \dots, m$ est donnée par:

$$f_{i\bullet} = \frac{n_{i\bullet}}{n}$$

- Fréquence marginale de y_j $j = 1, \dots, k$ est donnée par:

$$f_{\bullet j} = \frac{n_{\bullet j}}{n}$$

II. Tableau de fréquences

X\Y	y₁	y₂	...	y_k	Marge de X $f_{i\bullet}$
x₁	f ₁₁	f ₁₂	...	f _{1k}	$f_{1\bullet}$
x₂	f ₂₁	f ₂₂	...	f _{2k}	$f_{2\bullet}$
...	
x_m	f _{m1}	f _{m2}	...	f _{mk}	$f_{m\bullet}$
Marge de Y $f_{\bullet j}$	$f_{\bullet 1}$	$f_{\bullet 2}$...	$f_{\bullet k}$	1

$$f_{ij} = \frac{n_{ij}}{n}$$

$$f_{i\bullet} = \frac{n_{i\bullet}}{n}$$

$$f_{\bullet j} = \frac{n_{\bullet j}}{n}$$

II. Tableau de fréquences

Exemple :

Le tableau de fréquence du couple (x_i, y_j) et des fréquences marginales est :

X\Y	P	AB	B	TB	Fréquence Marginale de X
3	0,2	0,1	0,1	0,05	0,45
4	0,15	0,1	0,05	0,05	0,35
5	0,1	0,05	0,05	0	0,2
Fréquence Marginale de Y	0,45	0,25	0,2	0,1	1

II. Tableau de fréquences

Le tableau de fréquence du couple (x_i, y_j) et des fréquences marginales est :

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[0,02	0,04	0,02	0,05	0,01	0,14	
[5,10[0,04	0,05	0,05	0,02	0,02	0,18	
[10,15[0,02	0,10	0,05	0,15	0,04	0,36	
[15,20[0,04	0,04	0,02	0,01	0,01	0,12	
[20,25[0,05	0,02	0,01	0,10	0,2	0,20	
Marge de Y	$n_{\bullet j}$	0,17	0,25	0,15	0,33	0,10	N = 1

III. Moyennes et Variances Marginales

Soient $\{x_i : 1 \leq i \leq m\}$ et $\{y_j : 1 \leq j \leq k\}$ les valeurs discrètes ou les centres de classe pour les variables continues. Alors les moyennes et variances marginales sont:

Variable X	Variable Y
<p>Moyenne marginale de X</p> $\bar{x} = \frac{1}{n} \sum_{i=1}^m n_{i\cdot} x_i = \sum_{i=1}^m f_{i\cdot} x_i$	<p>Moyenne marginale de Y</p> $\bar{y} = \frac{1}{n} \sum_{j=1}^k n_{\cdot j} y_j = \sum_{j=1}^k f_{\cdot j} y_j$
<p>Variance marginale de X</p> $\sigma_X^2 = \frac{1}{n} \sum_{i=1}^m n_{i\cdot} x_i^2 - \bar{x}^2$	<p>Variance marginale de Y</p> $\sigma_Y^2 = \frac{1}{n} \sum_{j=1}^k n_{\cdot j} y_j^2 - \bar{y}^2$

III. Moyennes et Variances Marginales

Soit l'exemple suivant :

X \ Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[2	4	2	5	1	14	
[5,10[4	5	5	2	2	18	
[10,15[2	10	5	15	4	36	
[15,20[4	4	2	1	1	12	
[20,25[5	2	1	10	2	20	
Marge de Y	$n_{\bullet j}$	17	25	15	33	10	N = 100



Distribution
marginale de X

x_i	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
2,5	14	35	87,5
7,5	18	135	1012,5
12,5	36	450	5625
17,5	12	210	3675
22,5	20	450	10125
Total	100	1280	20525

III. Moyennes et Variances Marginales

Calcul de la moyenne marginale et la variance marginale de X

x_i	$n_{i\bullet}$	$n_{\bullet i}x_i$	$n_{i\bullet}x_i^2$
2,5	14	35	87,5
7,5	18	135	1012,5
12,5	36	450	5625
17,5	12	210	3675
22,5	20	450	10125
Total	100	1280	20525

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_{i\bullet} x_i = \frac{1}{100} \times 1280 = 12.8$$

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^m n_{i\bullet} x_i^2 - \bar{x}^2 = \frac{1}{100} \times 20525 - (12.8)^2 = 205.25 - 163.84 = 41.41$$

IV. Distributions Conditionnelles

La distribution conditionnelle est la distribution d'une variables sachant que l'autre a une valeur fixe. On écrit $X/Y = y_j$ (X sachant $Y = y_j$) ou $Y/X = x_i$ (Y sachant $X = x_i$)

Il existe deux types de distributions:

- ☐ Distribution conditionnelle de Y sachant $X = x_i$
- ☐ Distribution conditionnelle de X sachant $Y = y_j$.

IV. Distributions Conditionnelles

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[2	4	2	5	1	14	
[5,10[4	5	5	2	2	18	
[10,15[2	10	5	15	4	36	
[15,20[4	4	2	1	1	12	
[20,25[5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	

Soient:

- ☐ X une variable qui varie en parcourant toutes les classes
- ☐ Y est une variable qui fait parti à l'intervalle $[0, 5[$

La distribution conditionnelle de X sachant que $Y \in [0, 5[$ et les fréquences conditionnelles obtenues par les effectifs n_{i1} ($i = 1, \dots, 5$) divisés par $n_{\bullet 1}$

IV. Distributions Conditionnelles

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[2	4	2	5	1	14	
[5,10[4	5	5	2	2	18	
[10,15[2	10	5	15	4	36	
[15,20[4	4	2	1	1	12	
[20,25[5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	



X\Y	L'effectif conditionnel de $X / Y \in [0,5[$	Fréquence conditionnelle de $X / Y \in [0,5[$
[0,5[2	0,12
[5,10[4	0,23
[10,15[2	0,12
[15,20[4	0,23
[20,25[5	0,3
$n_{\bullet 1}$	17	1

IV. Distributions Conditionnelles

La distribution conditionnelle de Y sachant que $X \in [10,15[$ et les fréquences conditionnelles obtenues par les effectifs n_{3j} ($j = 1, \dots, 5$) divisés par $n_{3\bullet}$.

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[2	4	2	5	1	14	
[5,10[4	5	5	2	2	18	
[10,15[2	10	5	15	4	36	
[15,20[4	4	2	1	1	12	
[20,25[5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	



X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{3\bullet}$
L'effectif conditionnel de $Y/X \in [10,15[$	2	10	5	15	4	36	
Fréquence conditionnelle de $Y/X \in [10,15[$	0.055	0.277	0.138	0.416	0.111	1	

V. Moyennes et Variances Conditionnelles

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[Marge de X	$n_{i\bullet}$
[0,5[2	4	2	5	1	14	
[5,10[4	5	5	2	2	18	
[10,15[2	10	5	15	4	36	
[15,20[4	4	2	1	1	12	
[20,25[5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	

La distribution conditionnelle de X sachant [10,15[est une série univariée obtenue par l'élimination de toutes les classes sauf celle correspond à $n_{3\bullet}$.



X\Y	$n_{3\bullet}$
[0,5[2
[5,10[5
[10,15[5
[15,20[2
[20,25[1

V. Moyennes et Variances Conditionnelles

Calcul de la moyenne conditionnelle et la variance conditionnelle de X

X\Y	c_i	$n_{3\bullet}$	$c_i n_{3\bullet}$	$c_i^2 n_{3\bullet}$
[0,5[2,5	2	5	12,5
[5,10[7,5	5	37,5	281,25
[10,15[12,5	5	62,5	781,25
[15,20[17,5	2	35	612,5
[20,25[22,5	1	22,5	506,25
Total		15	162,5	2193,75

$$\bar{x}_{/[10,15]} = \bar{x}_{/3} = \frac{1}{n_{3\bullet}} \sum_{i=1}^m c_i n_{i3} = \frac{162.5}{15} = 10.83$$

$$\sigma_{/[10,15]}^2 = \sigma_{/3}^2 = \frac{1}{n_{3\bullet}} \sum_{i=1}^m n_{ij} x_i^2 - (\bar{x}_{/3})^2 = \frac{2193.75}{15} - (10.83)^2 = 34.7$$

V. Moyennes et Variances Conditionnelles

Variable X sachant $Y = y_j$	Variable Y sachant $X = x_i$
<p>Moyenne conditionnelle de X sachant $Y = y_j$</p> $\bar{x}_{/j} = \frac{1}{n_{\bullet j}} \sum_{i=1}^m n_{ij} x_i = \sum_{i=1}^m f_{i/j} x_i$	<p>Moyenne conditionnelle de Y sachant $X = x_i$</p> $\bar{y}_{/i} = \frac{1}{n_{i\bullet}} \sum_{j=1}^k n_{ij} y_j = \sum_{j=1}^k f_{j/i} y_j$
<p>Variance conditionnelle de X sachant $Y = y_j$</p> $\sigma_{X/j}^2 = \frac{1}{n_{\bullet j}} \sum_{i=1}^m n_{ij} (x_i - \bar{x}_{/j})^2$	<p>Variance conditionnelle de Y sachant $X = x_i$</p> $\sigma_{Y/i}^2 = \frac{1}{n_{i\bullet}} \sum_{j=1}^k n_{ij} (y_j - \bar{y}_{/i})^2$

VI. Covariance

La covariance de X et Y est la quantité

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^k n_{ij} (x_i - \bar{x})(y_j - \bar{y}) \\ &= \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^k n_{ij} x_i y_j - \bar{x} \times \bar{y} \\ &= \overline{xy} - \bar{x} \times \bar{y} \end{aligned}$$

Remarque:

Dans le cas des séries quantitatives non-groupées, on a $n_{ij} = 1$

VI. Covariance

Soient X et Y deux variables statistiques qui désignent respectivement le poids (Kg) et la taille (Cm) de 6 animaux.

Cas 1: Variables quantitatives non-groupées

	Poids (x_i)	Taille (y_i)	$x_i y_i$
1	12	10	120
2	14	11	154
3	15	13	195
4	6	7	42
5	9	11	99
6	7	5	35
Total	63	57	645

On a :

$$\bar{x} = \frac{63}{6} = 10.5 \quad \text{et} \quad \bar{y} = \frac{57}{6} = 9,5$$

$$\text{Cov}(X, Y) = \frac{645}{6} - (10.5 \times 9,5) = 9.25$$

VI. Covariance

Cas 2: Variables quantitatives groupées

X\Y	[0,5[[5,10[[10,15[[15,20[[20,25[x_i	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
[0,5[2	4	2	5	1	2,5	14	35	87,5
[5,10[4	5	5	2	2	7,5	18	135	1012,5
[10,15[2	10	5	15	4	12,5	36	450	5625
[15,20[4	4	2	1	1	17,5	12	210	3675
[20,25[5	2	1	10	2	22,5	20	450	10125
y_j	2,5	7,5	12,5	17,5	22,5			1280	20525
$n_{\bullet j}$	17	25	15	33	10		N = 100		
$n_{\bullet j}y_j$	42,5	187,5	187,5	577,5	225	1220			
$n_{\bullet j}y_j^2$	106,25	1406,25	486,75	10106,25	5062,5	17168			

$$\bar{x} = \frac{1280}{100} = 12.8$$

et

$$\bar{y} = \frac{1220}{100} = 12.2$$

$$V(Y) = \frac{17168}{100} - (12.2)^2 = 7.84$$

et

$$V(X) = \frac{20525}{100} - (12.8)^2 = 56.41$$

VI. Covariance

Calculons la somme suivante: $\sum_{i=1}^m \left(\sum_{j=1}^k n_{ij} y_j \right) \times x_i$. Pour chaque case on fixe i et on fait varier j.

Case N°1:

						Centre de X x_1
Effectif marginaux n_{1j} de [0,5[2	4	2	5	1	2,5
Centre de Y y_j	2,5	7,5	12,5	17,5	22,5	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (2 \times 2.5 + 4 \times 7.5 + 2 \times 12.5 + 5 \times 17.5 + 1 \times 22.5) \times 2.5 = 425$$

Case N°2:

						Centre de X x_2
Effectif marginaux n_{2j} de [5,10[4	5	5	2	2	7,5
Centre de Y y_j	2,5	7,5	12,5	17,5	22,5	

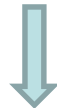


$$(n_{21}y_1 + n_{22}y_2 + n_{23}y_3 + n_{24}y_4 + n_{25}y_5) \times x_1 = (4 \times 2.5 + 5 \times 7.5 + 5 \times 12.5 + 2 \times 17.5 + 2 \times 22.5) \times 7.5 = 1425$$

VI. Covariance

Case N°3:

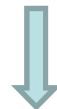
						Centre de X x_3
Effectif marginaux n_{3j} de [10,15[2	10	5	15	4	12,5
Centre de Y y_j	2,5	7,5	12,5	17,5	22,5	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (2 \times 2.5 + 10 \times 7.5 + 5 \times 12.5 + 15 \times 17.5 + 4 \times 22.5) \times 12.5 = 6187.5$$

Case N°4:

						Centre de X x_4
Effectif marginaux n_{4j} de [15,20[4	4	2	1	1	17,5
Centre de Y y_j	2,5	7,5	12,5	17,5	22,5	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (4 \times 2.5 + 4 \times 7.5 + 2 \times 12.5 + 1 \times 17.5 + 1 \times 22.5) \times 17.5 = 1837.5$$

VI. Covariance

Case N°5:

						Centre de X x_5
Effectif marginaux n_{5j} de [15,20[5	2	1	10	2	22,5
Centre de Y y_j	2,5	7,5	12,5	17,5	22,5	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (5 \times 2.5 + 2 \times 7.5 + 1 \times 12.5 + 10 \times 17.5 + 2 \times 22.5) \times 22.5 = 5850$$

$$\begin{aligned}
 \text{Cov}(X, Y) &= \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^k n_{ij} x_i y_j - \bar{x} \times \bar{y} \\
 &= \frac{1}{100} (425 + 1425 + 6187.5 + 1837.5 + 5850) - 12.2 \times 12.8 \\
 &= 157.25 - 156.16 = 1.09
 \end{aligned}$$

V. Corrélation

Le coefficient de corrélation linéaire de X et Y est donnée par:

$$R = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

R^2 désigne le coefficient de détermination de X et Y

Remarques importante :

- ☐ On a toujours $-1 \leq R \leq 1 \Rightarrow 0 \leq R^2 \leq 1$.
- ☐ La covariance dépend des unités ce qui est difficile à interpréter .
- ☐ La corrélation ne dépend pas des unités.

V. Corrélation

$$\bar{x} = \frac{1280}{100} = 12.8$$

et

$$\bar{y} = \frac{1220}{100} = 12.2$$

$$V(X) = \frac{17168}{100} - (12.8)^2 = 7.84$$

et

$$V(Y) = \frac{20525}{100} - (12.2)^2 = 56.41$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{7.84} = 2.8$$

et

$$\sigma_Y = \sqrt{V(Y)} = \sqrt{56.41} = 7.51$$

$$Cov(X, Y) = 1.09$$

$$R = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{1.09}{2.8 \times 7.51} = 0.051$$