

Statistique descriptive
Année académique 2020–2021
Carole.Baum@uliege.be

Chapitre 4 : Série statistique bivariable

Exercice 1. Le Tableau de contingence 1 décrit la répartition d’une population constituée de 535 ménages selon les deux variables suivantes : X représente le nombre de pièces de l’habitation et Y correspond au nombre d’enfants du ménage.

TABLE 1 – Tableau de contingence pour la série double “Nombre de pièces - Nombre d’enfants”.

Valeurs de X	Valeurs de Y				
	0	1	2	3	4
1	7	3	2	1	0
2	24	32	21	2	1
3	16	35	54	26	4
4	9	28	74	55	12
5	4	12	46	13	12
6	2	6	16	11	7

- (a) Que représentent les effectifs n_{23} et n_{54} ? Calculer les fréquences correspondantes.
- (b) Déterminer les distributions marginales des variables X et Y . Calculer les moyennes, médianes et variances marginales ainsi que les modes marginaux.
- (c) Déterminer la distribution conditionnelle de la variable Y sachant que le nombre de pièces du logement est égal à 4. Calculer la moyenne et la variance de cette distribution conditionnelle.
- (d) Calculer la covariance de la série double.

Solution :

- (a) $n_{23} = 21$ il s’agit du nombre de ménages ayant 2 pièces dans l’habitation et 2 enfants.
 $n_{54} = 13$ il s’agit du nombre de ménages ayant 5 pièces dans l’habitation et 3 enfants.
On a $f_{23} = 21/535 = 0.039$ et $f_{54} = 13/535 = 0.024$.
- (b) La distribution marginale de la variable X est donnée par

x_i	1	2	3	4	5	6
n_i	13	80	135	178	87	42

On peut alors déterminer les paramètres demandés.

- $\bar{x} = \frac{1}{535}(13 \cdot 1 + \dots + 42 \cdot 6) = 3.70$;
- $\frac{n}{2} = 267.5 \Rightarrow \tilde{x} = x_{(268)} = 4$ car $\begin{cases} N(3) = 228 \\ N(4) = 406 \end{cases}$;
- $s_x^2 = \frac{1}{535}(13 \cdot 1^2 + \dots + 42 \cdot 6^2) - 3.70^2 = 1.42$;

$$- x_M = 4.$$

De la même manière, la distribution marginale de Y et les paramètres demandés sont donnés par

y_j	0	1	2	3	4
n_j	62	116	213	108	36

$$\begin{aligned}
 - \bar{y} &= \frac{1}{535}(62 \cdot 0 + \dots + 36 \cdot 4) = 1.89; \\
 - \tilde{y} &= y_{(268)} = 2; \\
 - s_y^2 &= \frac{1}{535}(62 \cdot 0^2 + \dots + 36 \cdot 4^2) - 1.89^2 = 1.13; \\
 - y_M &= 2.
 \end{aligned}$$

(c) La distribution conditionnelle de Y sachant que $X = 4$ est donnée par

$y_i x = 4$	0	1	2	3	4	
n_i	9	28	74	55	12	178

On a alors

$$\begin{aligned}
 - \bar{y}_{|x=4} &= \frac{1}{178}(9 \cdot 0 + \dots + 12 \cdot 4) = 2.185; \\
 - s_{y_{|x=4}}^2 &= \frac{1}{178}(9 \cdot 0^2 + \dots + 12 \cdot 4^2) - 2.185^2 = 0.906.
 \end{aligned}$$

$$(d) s_{xy} = \frac{1}{535}(7 \cdot 1 \cdot 0 + 3 \cdot 1 \cdot 1 + \dots + 0 \cdot 1 \cdot 4 + 24 \cdot 0 \cdot 2 + \dots + 7 \cdot 6 \cdot 4) - 3.70 \cdot 1.89 = 0.47.$$

Exercice 2. Le directeur d'une entreprise vinicole a l'habitude d'offrir à ses employés une prime de fin d'année de 10, 15, 20 ou 25 unités monétaires. Le tableau 2 représente le tableau de contingence mettant en rapport la variable Y = "montant de la prime" avec la variable X = "taille de la cave à vin exprimée en nombre de bouteilles".

TABLE 2 – Tableau de contingence pour la série double "Taille de la cave - Montant de la prime".

X	Y			
	10	15	20	25
$[0, 100]$	11	13	5	3
$]100, 200]$	14	21	15	6
$]200, 300]$	5	8	9	7
$]300, 400]$	3	6	8	9
$]400, 500]$	2	3	5	4

- Quelle est la prime la plus souvent distribuée ? Que représente cette valeur dans la distribution marginale de Y ?
- Quelle est la somme totale dépensée par le directeur pour offrir les primes de fin d'année ?
- Si le directeur voulait être équitable par rapport à tous ses employés, quel montant donnerait-il à chacun ? Cette valeur est-elle proche des moyenne et médiane marginales de Y ?
- Comparer les nombres moyens de bouteilles détenues par les employés en tenant compte des différentes primes perçues.

Solution :

- (a) La prime la plus souvent distribuée est le mode de la distribution marginale de y :

y_i	10	15	20	25	
n_i	35	51	42	29	157

Ainsi, $y_M = 15$.

- (b) La somme totale dépensée est donnée par $10 \cdot 35 + 15 \cdot 51 + 20 \cdot 42 + 25 \cdot 29 + 25 \cdot 29 = 2680$.

- (c) Une somme “honnête” consisterait à diviser la somme totale par le nombre d’employés : $\frac{2680}{157} = 17.07$;
 $\bar{y} = 17.07$ (la moyenne est, par définition, la valeur donnée ci-dessus) ;
 $\tilde{y} = 15$.
- (d) $\bar{x}_{|y=10} = \frac{11 \cdot 50 + 14 \cdot 150 + 5 \cdot 250 + 3 \cdot 350 + 2 \cdot 450}{35} = 167.14$;
 $\bar{x}_{|y=15} = \frac{13 \cdot 50 + 21 \cdot 150 + 8 \cdot 250 + 6 \cdot 350 + 3 \cdot 450}{51} = 181.37$;
 $\bar{x}_{|y=20} = \frac{5 \cdot 50 + 15 \cdot 150 + 9 \cdot 250 + 8 \cdot 350 + 5 \cdot 450}{42} = 233.33$;
 $\bar{x}_{|y=25} = \frac{3 \cdot 50 + 6 \cdot 150 + 7 \cdot 250 + 9 \cdot 350 + 4 \cdot 450}{29} = 267.24$.

Exercice 3. Le tableau 3 reprend les âges de l’époux (variable X) et de l’épouse (variable Y) pour les 100 derniers mariages enregistrés dans une ville.

TABLE 3 – Ages des époux pour les 100 derniers mariages enregistrés dans une ville

Classes de X	Classes de Y			
	[18, 22]]22, 26]]26, 30]]30, 34]
[20, 24]	14	5	2	0
]24, 28]	15	19	7	3
]28, 32]	5	12	5	2
]32, 36]	0	1	2	2
]36, 40]	0	1	3	2

- (a) Calculer l’âge moyen des époux et des épouses au moment du mariage.
- (b) Comment peut-on obtenir à partir du tableau de contingence les informations suivantes : (1) 38% des femmes se marient entre 22 et 26 ans ; (2) 19% des mariages concernent des hommes âgés de 24 à 28 ans et des femmes âgées de 22 à 26 ans ; (3) 31,6% des femmes âgées de 22 à 26 ans se sont mariées avec des hommes de 28 à 32 ans.

Solution :

- (a) Les distributions marginales sont données par

x_i	[20; 24]]24; 28]]28; 32]]32; 36]]36; 40]
n_i	21	44	24	5	6
y_i	[18; 22]]22; 26]]26; 30]]30; 34]	
n_i	34	38	19	9	

dont on déduit

$$\bar{x} = \frac{21 \cdot 22 + 44 \cdot 26 + 24 \cdot 30 + 5 \cdot 34 + 6 \cdot 38}{100} = 27.24 \quad \text{et} \quad \bar{y} = \frac{34 \cdot 20 + 38 \cdot 24 + 19 \cdot 28 + 9 \cdot 32}{100} = 24.12$$

- (b) (1) Il s’agit de la fréquence marginale de la classe $]22; 26]$ chez les femmes, à savoir 38/100.
 (2) Il s’agit de la fréquence du couple $(]24; 28];]22; 26])$, à savoir 19/100.
 (3) On cherche la proportion de femmes entre 22 et 26 ans qui se sont mariées avec des hommes de 28 à 32 ans. Parmi les 38 femmes vérifiant la première condition, 12 vérifient la seconde. On a donc $12/38 = 31.6\%$.

Exercice 4. Le gérant d’un magasin d’appareils électroménagers a enregistré chaque semaine le nombre x_i de centaines d’appels téléphoniques reçus de l’extérieur et le chiffre d’affaires y_i réalisé en unités monétaires. Les résultats sont repris dans le tableau 4 où n_{ij} indique le nombre de semaines où le magasin a reçu x_i centaines d’appels téléphoniques et a fait y_j UM comme chiffre d’affaires.

TABLE 4 – Tableau de contingence pour la série double “Nombre d’appels téléphoniques - Chiffre d’affaires”.

Nombre d’appels téléphoniques	Chiffre d’affaires				
	1	2	3	4	5
2	9	5	3	1	0
3	4	5	7	3	1
4	0	6	9	6	3
5	0	5	14	14	5

- (a) Combien de semaines cette enquête a-t-elle duré ?
 (b) Combien d’appels téléphoniques le magasin reçoit-il en moyenne par semaine ?
 (c) Déterminer les distributions marginales ainsi que les moyennes et variances marginales.
 (d) Comparer les chiffres d’affaires moyens réalisés lorsque le magasin reçoit 2 ou 5 appels téléphoniques sur la semaine.
 (e) Calculer la covariance et le coefficient de corrélation linéaire entre les deux variables.

Solution :

- (a) Le nombre de semaines qu’a duré l’enquête est donné par la somme des n_{ij} .
 On a donc $n = 9 + 5 + \dots + 0 + 4 \dots + 5 = 100$.

- (b) La distribution marginale du nombre d’appels téléphoniques est donnée par

x_i	2	3	4	5
n_i	18	20	24	38

dont on déduit que $\bar{x} = \frac{18 \cdot 2 + 20 \cdot 3 + 24 \cdot 4 + 38 \cdot 5}{100} = 3.82$.

- (c) La distribution et la moyenne marginale de x sont données ci-dessus. On peut aussi calculer la variance marginale :

$$s_x^2 = \frac{18 \cdot 2^2 + 20 \cdot 3^2 + 24 \cdot 4^2 + 38 \cdot 5^2}{100} - 3.82^2 = 1.27.$$

De la même manière, la distribution marginale de y est donnée par

y_j	1	2	3	4	5
n_j	13	21	33	24	9

dont on déduit les moyenne et variance marginales :

$$\bar{y} = \frac{13 \cdot 1 + 21 \cdot 2 + 33 \cdot 3 + 24 \cdot 4 + 9 \cdot 5}{100} = 2.95$$

$$s_y^2 = \frac{13 \cdot 1^2 + 21 \cdot 2^2 + 33 \cdot 3^2 + 24 \cdot 4^2 + 9 \cdot 5^2}{100} - 2.95^2 = 1.33.$$

- (d) On veut comparer les moyennes conditionnelles de y lorsque $x = 2$ et lorsque $x = 5$. On a

$$\bar{y}_{|x=2} = \frac{9 \cdot 1 + 5 \cdot 2 + 3 \cdot 3 + 1 \cdot 4 + 0 \cdot 5}{18} = 1.78$$

$$\bar{y}_{|x=5} = \frac{0 \cdot 1 + 5 \cdot 2 + 14 \cdot 3 + 14 \cdot 4 + 5 \cdot 5}{38} = 3.5.$$

Comme on pouvait s’y attendre, le chiffre d’affaire moyen est plus élevé lorsque le nombre d’appels est élevé.

- (e) La covariance est donnée par

$$s_{xy} = \frac{1}{n} \sum_{i,j} n_{ij} x_i x_j - \bar{x} \bar{y}$$

$$= \frac{9 \cdot 2 \cdot 1 + \dots + 0 \cdot 2 \cdot 5 + 4 \cdot 3 \cdot 1 + \dots + 5 \cdot 5 \cdot 5}{100} - 3.82 \cdot 2.95$$

$$= 0.701.$$

Le coefficient de corrélation est quant à lui donné par

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{0.701}{\sqrt{1.27 \cdot 1.33}} = 0.539.$$

Exercice 5. Une étude statistique a été réalisée auprès des 500 entreprises d'un même secteur industriel. Deux variables étaient considérées : la variable X correspond à la taille de l'entreprise (en nombre de salariés) et la variable Y au niveau de leur salaire mensuel (en unités monétaires). L'étude a permis d'obtenir les renseignements du tableau 5.

TABLE 5 – Répartition de 500 entreprises en fonction du nombre de salariés et du salaire mensuel

Nombre de salariés (X)	Niveau de salaire mensuel (Y)		
	5000	10 000	30 000
$[0, 500]$	13	5	2
$]500, 1500]$	60	25	15
$]1500, 3000]$	90	30	30
$]3000, 5000]$	137	65	28

- (a) Déterminer les salaires mensuels moyen et médian des employés des entreprises considérées.
- (b) Déterminer, pour chaque classe de taille d'entreprise, le salaire mensuel moyen.
- (c) Quelle relation existe-t-il entre le salaire moyen global calculé en (a) et les salaires moyens calculés en (b).

Solution :

- (a) La distribution marginale du salaire mensuel est donnée par

y_j	5000	10 000	30 000
n_j	300	125	75

Dès lors, le salaire mensuel moyen est donné par

$$\bar{y} = \frac{300 \cdot 5000 + 125 \cdot 10\,000 + 75 \cdot 30\,000}{500} = 10\,000.$$

Le salaire mensuel médian est quant à lui donné par

$$\tilde{y} = \frac{y_{(250)} + y_{(251)}}{2} = \frac{5000 + 5000}{2} = 5000.$$

- (b) Les salaires moyens conditionnels sont donnés par

$$\begin{aligned}\bar{y}_{|x_1} &= \frac{13 \cdot 5000 + 5 \cdot 10\,000 + 2 \cdot 30\,000}{20} = 8750 \\ \bar{y}_{|x_2} &= \frac{60 \cdot 5000 + 25 \cdot 10\,000 + 15 \cdot 30\,000}{100} = 10\,000 \\ \bar{y}_{|x_3} &= \frac{90 \cdot 5000 + 30 \cdot 10\,000 + 30 \cdot 30\,000}{150} = 11\,000 \\ \bar{y}_{|x_4} &= \frac{137 \cdot 5000 + 65 \cdot 10\,000 + 28 \cdot 30\,000}{230} = 9456\end{aligned}$$

où x_1, \dots, x_4 représentent respectivement les classes $[0, 500], \dots,]3000, 5000]$ de nombres de salariés.

- (c) On a

$$\bar{y} = \frac{20 \cdot \bar{y}_{|x_1} + 100 \cdot \bar{y}_{|x_2} + 150 \cdot \bar{y}_{|x_3} + 230 \cdot \bar{y}_{|x_4}}{500}.$$

Autrement dit, le salaire moyen global n'est rien d'autre que la moyenne pondérée des salaires moyens conditionnels.