Statistique descriptive bivariée

Présentation et description des données

15 mai 2023

www.tifawt.com

Chapitre 4 : Présentation et description des données d'une distribution à deux caractères

Plan du Chapitre 4

1. **Définition des séries statistiques à deux variables**
    - **Tableau de contingence**
    - **Distribution conjointe**
    - **Effectifs partiels**
    - **Fréquences partielles sur effectif total**
2. **Distributions marginales et conditionnelles**
    - Effectifs marginaux et fréquences marginales
    - fréquences conditionnelles
3. **Caractéristiques numériques des séries statistiques à deux variables**
    - Moyennes et variances marginales
    - Moyennes et variances conditionnelles
    - Covariance
4. **Exercices**

1. Définition des séries statistiques à deux variables

La statistique bivariée est l'étude de deux variable statistiques simultanément, celles-ci pouvant être quantitatives ou qualitatives.
Elle cherche à déterminer si deux variables X et Y , étudiées sur la même population, sont liées, c'est à dire si l'une à une influence sur l'autre, ou bien si elles sont complètement indépendantes.

- Dans une population des salariées d'une entreprise, la variable X pourrait être l'âge et Y le salaire mensuel.
- Dans une population des élèves, la variable X pourrait être la note obtenue dans une matière et Y l'assiduité.

1. Définition des séries statistiques à deux variables

1.1. Tableau de contingence :

Le tableau de contingence est un tableau à double entrée et qui représente une série statistique à deux caractères.

Exemple 4.1

On considère le tableau de contingence suivant qui représente la distribution de 250 salariés d'une petite entreprise selon l'âge et le salaire mensuel en milles dirhams.

x_i (L'âge) \ y_j (Salaire)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

y_j (Salaire) / x_i (L'âge)	$[6, 8[$	$[8, 10[$	$[10, 12[$	Total
$[20, 30[$	15	25	20	60
$[30, 40[$	10	45	15	70
$[40, 50[$	30	50	40	120
Total	55	120	75	250

• Si l'on associe la ligne du bas "marge" et la ligne du haut (y_j), on obtient une distribution de 250 salariés selon leur salaire mensuel. Cette distribution est à une dimension, on l'appelle distribution marginale du caractère y.

• De même, la dernière colonne "marge" associée à la première constitue une distribution marginale du caractère x, distribution des salariés selon leurs âges.

Remarque

y_j (Salaire) / x_i (L'âge)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

• Pour cet exemple, ces deux distributions marginales sont insuffisantes pour déterminer si l'âge a une influence sur le niveau de salaire.

• Le tableau de contingence permet de voir comment se distribuent les effectifs de chaque modalité d'un caractère suivant les modalités de l'autre.

Présentation générale du tableau de contingence

On considère ici, deux variables quantitatives, notées X et Y qui sont définies sur une même population de taille n . On notera :

- x_i, $i = 1, 2, ..., p$: les p modalités distinctes de la variable X et rangées par ordre croissant.
- y_j, $j = 1, 2, ..., q$: les q modalités distinctes de la variable Y et rangées par ordre croissant.
- Les deux variables X et Y sont mesurées simultanément sur chacun des n individus de la population.
- n_{ij} l'effectif correspondant au couple (x_i, y_j) : représente le nombre d'individus qui prennent en même temps la valeur x_i et y_j.
- Si X et Y sont des variables quantitatives continues : x_i et y_j désignent les centres des classes de X et Y respectivement.

Présentation générale du tableau de contingence

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

- Les modalités du caractère y apparaissent au premier ligne.
- Les modalités du caractère x apparaissent au premier colonne.

1.2 Distribution conjointe

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

- Les effectifs apparaissent à l'intérieur du tableau de type n_{ij} appelés effectifs partiels.

1.2 Distribution conjointe

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Définition

On appelle *série conjointe* du couple (X, Y) l'ensemble des informations(x_i, y_j, n_{ij}), pour $i = 1, ..., p$ et $j = 1, ..., q$,et qu'on représente par un tableau de contingence.

Exemple

x_i (L'âge) \ y_j (Salaire)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

• Pour le tableau de l'exemple 4.1 on a : $n_{21} = 10$.
C'est l'effectif des salariés de l'entreprise qui touchent un salaire compris entre 6000 dirhams et 8000 dirhams et qui sont âgés entre 30 et 40 ans.
On a encore par exemple $n_{32} = 50$, $n_{13} = 20$...

• On note qu'en général $n_{ij} \neq n_{ji}$.
Pour l'exemple 4.1 on a $n_{12} = 25$ et $n_{21} = 10$
Donc $n_{12} \neq n_{21}$.

Fréquences partielles sur effectif total

On pourrait définir le tableau de contingence d'une série statistique bivariée à l'aide des fréquences partielles sur effectif total : $f_{ij} = \frac{n_{ij}}{n_{\bullet\bullet}}$, au lieu des effectifs partielles n_{ij}.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q
x_1	f_{11}	f_{12}	...	f_{1j}	...	f_{1q}
x_2	f_{21}	f_{22}	...	f_{2j}	...	f_{2q}
...					...	
x_i	f_{i1}	f_{i2}	...	f_{ij}	...	f_{iq}
...					...	
x_p	f_{p1}	f_{p2}	...	f_{pj}	...	f_{pq}

Fréquences partielles sur effectif total

x_i (L'âge) \ y_j (Salaire)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

- Pour le tableau de l'exemple 4.1 on a : $f_{21} = \frac{n_{21}}{n_{\bullet\bullet}} = \frac{10}{250} = 0.04$

Cela signifie que :

4% est le pourcentage des salariés de l'entreprise qui ont un salaire compris entre 6000 dirhams et 8000 dirhams et qui sont âgés entre 30 et 40 ans.

Effectifs marginaux de la variable X

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

- L'effectif $n_{i\bullet}$ désigne la somme des effectifs de la ligne i, l'indice j a varié de 1 à q, on le remplace par « • », ainsi

$$n_{i\bullet} = \sum_{j=1}^{q} n_{ij}$$

Effectifs marginaux de la variable X

Les effectifs $n_{i\bullet}$ sont appelés des effectifs marginaux de la variable X.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Effectifs marginaux de la variable X

x_i (L'âge) \ y_j (Salaire)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

- Pour le tableau de l'exemple 4.1 on a : $n_{1\bullet} = 60, \ n_{2\bullet} = 70 \cdots$

$n_{1\bullet} = 60$ représente :
L'effectif des salariés de l'entreprise qui sont âgés entre 20 et 30 ans.
$n_{2\bullet} = 70$ représente :
L'effectif des salariés de l'entreprise qui sont âgés entre 30 et 40 ans.

Effectifs marginaux de la variable Y

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

- L'effectif $n_{\bullet j}$ désigne la somme des effectifs de la colonne j, l'indice i a varié de 1 à p, on le remplace par « • », ainsi

$$n_{\bullet j} = \sum_{i=1}^{p} n_{ij}$$

Effectifs marginaux de la variable Y

Les effectifs $n_{\bullet j}$ sont appelés des effectifs marginaux de la variable Y.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Effectifs marginaux de la variable Y

x_i (L'âge) \ y_j (Salaire)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

- Pour le tableau de l'exemple 4.1 on a : $n_{\bullet 1} = 55,\ n_{\bullet 2} = 120 \cdots$

$n_{\bullet 1} = 55$ représente :
L'effectif des salariés de l'entreprise qui ont un salaire compris entre 6000 dirhams et 8000 dirhams.

$n_{\bullet 2} = 120$ représente :
L'effectif des salariés de l'entreprise qui ont un salaire compris entre 8000 dirhams et 10000 dirhams.

Remarques

$X \backslash Y$	y_1	y_2	$\dots$	y_j	$\dots$	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\dots$	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	$\dots$	n_{2j}	$\dots$	n_{2q}	$n_{2\bullet}$
$\dots$					$\dots$		
x_i	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	n_{iq}	$n_{i\bullet}$
$\dots$					$\dots$		
x_p	n_{p1}	n_{p2}	$\dots$	n_{pj}	$\dots$	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	$\dots$	$n_{\bullet j}$	$\dots$	$n_{\bullet q}$	$n_{\bullet\bullet}$

On a l'effectif total $n = n_{\bullet\bullet} = \sum_{j=1}^{q} n_{\bullet j} = \sum_{i=1}^{p} n_{i\bullet}$

Et $n_{\bullet\bullet} = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}$

Ou encore $n_{\bullet\bullet} = \sum_{j=1}^{q} \sum_{i=1}^{p} n_{ij}$

2.1.1 Distribution marginale de X

La distribution marginale de X est défini par les p modalités x_i de la variable X et les p effectifs marginaux $n_{i\bullet}$ correspondants, c'est à dire la distribution $\{x_i, n_{i\bullet}\}$.

X \ Y	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Voir moyenne marginale $\overline{\overline{x}}$

2.1.1 Distribution marginale de X

En reprenant les données de l'Exemple 4.1, la distribution marginale de X est donnée par :

x_i (L'âge)	Effectifs
$[20, 30[$	60
$[30, 40[$	70
$[40, 50[$	120
Total	250

2.1.2 Distribution marginale de Y

La distribution marginale de Y est défini par les q modalités y_i de la variable Y et les q effectifs marginaux $n_{\bullet j}$ correspondants, c'est à dire la distribution $\{y_i, n_{\bullet j}\}$.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

2.1.2 Distribution marginale de Y

En reprenant les données de l'Exemple 4.1, la distribution marginale de Y est donnée par :

y_j (Salaire)	Effectifs
$[6,8[$	55
$[8,10[$	120
$[10,12[$	75
Total	250

Fréquences marginales

Ce sont les rapports des effectifs marginaux à leurs totaux (l'effectif total $n_{\bullet\bullet}$), on a :

$$f_{i\bullet} = \frac{n_{i\bullet}}{n_{\bullet\bullet}} \quad \text{et} \quad f_{\bullet j} = \frac{n_{\bullet j}}{n_{\bullet\bullet}}$$

.

Fréquences marginales

y_j (Salaire) / x_i (L'âge)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

- En considérant encore le tableau de l'exemple 4.1 on a :

$f_{1\bullet} = \frac{n_{1\bullet}}{n_{\bullet\bullet}} = \frac{60}{250} = 0.24$

Cela signifie que : 24% est le pourcentage des salariés de l'entreprise qui sont âgés entre 20 et 30 ans.

On a : $f_{\bullet 2} = \frac{n_{\bullet 2}}{n_{\bullet\bullet}} = \frac{120}{250} = 0.48$

Cela signifie que : 48% est le pourcentage des salariés de l'entreprise ont un salaire compris entre 8000 dirhams et 10000 dirhams.

Propriétés et relation entre les fréquences partielles sur effectif total et fréquences marginales

Le tableau suivant donne les fréquences partielles sur effectif total du couple (x_i, y_i) et les fréquences marginales des variables X et Y. correspondantes aux données de l'Exemple 4.1

x_i \ y_j	[6,8[	[8,10[	[10,12[	Total
[20,30[	$f_{11} = 0,06$	$f_{12} = 0,1$	$f_{13} = 0,08$	$f_{1\bullet} = 0,24$
[30,40[	$f_{21} = 0,04$	$f_{22} = 0,18$	$f_{23} = 0,06$	$f_{2\bullet} = 0,28$
[40,50[	$f_{31} = 0,12$	$f_{32} = 0,2$	$f_{33} = 0,16$	$f_{3\bullet} = 0,48$
Total	$f_{\bullet 1} = 0,22$	$f_{\bullet 2} = 0,48$	$f_{\bullet 3} = 0,3$	1

Propriétés et relation entre les fréquences partielles sur effectif total et fréquences marginales

x_i \ y_j	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	$f_{11}=0,06$	$f_{12}=0,1$	$f_{13}=0,08$	$f_{1\bullet}=0,24$
$[30,40[$	$f_{21}=0,04$	$f_{22}=0,18$	$f_{23}=0,06$	$f_{2\bullet}=0,28$
$[40,50[$	$f_{31}=0,12$	$f_{32}=0,2$	$f_{33}=0,16$	$f_{3\bullet}=0,48$
Total	$f_{\bullet 1}=0,22$	$f_{\bullet 2}=0,48$	$f_{\bullet 3}=0,3$	1

- On a $f_{i\bullet}=\sum_{j=1}^{q} f_{ij}$ et $f_{\bullet j}=\sum_{i=1}^{p} f_{ij}$.
- La somme des fréquences marginales et la somme des fréquences partielles sur effectif total sont égales à 1 :

$$\boxed{\sum_{j=1}^{q} f_{\bullet j}=1,\quad \sum_{i=1}^{p} f_{i\bullet}=1,\quad \sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij}=1 \quad \text{et} \quad \sum_{j=1}^{q}\sum_{i=1}^{p} f_{ij}=1.}$$

2.2.1 Distribution conditionnelle

Les distributions conditionnelles s'obtiennent en fixant la valeur d'une des deux variables (ou la modalité d'un des deux caractères).

Dans le cas de l'Exemple 4.1, la distribution conditionnelle de X quand $Y = y_1$ est donnée par la première colonne du tableau.

x_i \ y_j	[6,8[	[8,10[	[10,12[	Total
[20,30[	15	25	20	60
[30,40[	10	45	15	70
[40,50[	30	50	40	120
Total	55	120	75	250

$X/_{Y=y1}$	Effectifs
[20,30[	15
[30,40[	10
[40,50[	30
Total	55

2.2.1 Distribution conditionnelle

De même dans le cas de l'Exemple 4.1, la distribution conditionnelle de X quand $Y = y_2$ est donnée par la deuxième colonne du tableau.

$x_i \backslash y_j$	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

$X_{/Y=y2}$	Effectifs
$[20,30[$	25
$[30,40[$	45
$[40,50[$	50
Total	120

2.2.1 Distribution conditionnelle

De même dans le cas de l'Exemple 4.1, la distribution conditionnelle de X quand $Y = y_2$ est donnée par la deuxième colonne du tableau.
Et la distribution conditionnelle de X quand $Y = y_3$ est donnée par la troisième colonne du tableau.

x_i \ y_j	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

$X_{/Y=y3}$	Effectifs
$[20,30[$	20
$[30,40[$	15
$[40,50[$	40
Total	75

2.2.1 Distribution conditionnelle

Voir caractéristiques conditionnelles

Dans le cas de l'Exemple 4.1, il y a trois distributions conditionnelles de X car Y ne prend ici que trois valeurs.

2.2.1 Distribution conditionnelle

Voir caractéristiques conditionnelles

En général, sachant que j varie de 1 à q, il y a q distributions conditionnelles de X.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Distribution conditionnelle de X selon $Y = y_1$

2.2.1 Distribution conditionnelle

Voir caractéristiques conditionnelles

En général, sachant que j varie de 1 à q, il y a q distributions conditionnelles de X.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Distribution conditionnelle de X selon $Y = y_2$

2.2.1 Distribution conditionnelle

Voir caractéristiques conditionnelles

En général, sachant que j varie de 1 à q, il y a q distributions conditionnelles de X.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Distribution conditionnelle de X selon $Y = yq$

2.2.1 Distribution conditionnelle

En reprenant les données de l'Exemple 4.1, la distribution conditionnelle de Y quand $X = x_1$ est donnée par la première ligne du tableau.

x_i \ y_j	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

$Y_{/X=x_1}$	Effectifs
$[6,8[$	15
$[8,10[$	25
$[10,12[$	20
Total	60

2.2.1 Distribution conditionnelle

En reprenant les données de l'Exemple 4.1, la distribution conditionnelle de Y quand $X = x_1$ est donnée par la première ligne du tableau.

x_i \ y_j	[6,8[	[8,10[	[10,12[	Total
[20,30[	15	25	20	60
[30,40[	10	45	15	70
[40,50[	30	50	40	120
Total	55	120	75	250

$Y_{/X=x_1}$	Effectifs
[6,8[	15
[8,10[	25
[10,12[	20
Total	60

2.2.1 Distribution conditionnelle

De même dans le cas de l'Exemple 4.1, la distribution conditionnelle de Y quand $X = x_2$ est donnée par la deuxième ligne du tableau.

x_i \ y_j	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

$Y_{/X=x_2}$	Effectifs
$[6,8[$	10
$[8,10[$	45
$[10,12[$	15
Total	70

2.2.1 Distribution conditionnelle

De même dans le cas de l'Exemple 4.1, la distribution conditionnelle de Y quand $X = x_2$ est donnée par la deuxième ligne du tableau.
Et la distribution conditionnelle de Y quand $X = x_3$ est donnée par la troisième ligne du tableau.

x_i \ y_j	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

$Y_{/X=x_3}$	Effectifs
$[6,8[$	30
$[8,10[$	50
$[10,12[$	40
Total	120

2.2.1 Distribution conditionnelle

Dans le cas de l'Exemple 4.1, il y a trois distributions conditionnelles de Y car X ne prend ici que trois valeurs.

2.2.1 Distribution conditionnelle

En général, sachant que i varie de 1 à p, il y a p distributions conditionnelles de Y.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Distribution conditionnelle de Y selon $X = x_1$

2.2.1 Distribution conditionnelle

En général, sachant que i varie de 1 à p, il y a p distributions conditionnelles de Y.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Distribution conditionnelle de Y selon $X = x_2$

2.2.1 Distribution conditionnelle

En général, sachant que i varie de 1 à p, il y a p distributions conditionnelles de Y.

$X \backslash Y$	y_1	y_2	...	y_j	...	y_q	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	$n_{2\bullet}$
...					...		
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	$n_{i\bullet}$
...					...		
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}	$n_{p\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet q}$	$n_{\bullet\bullet}$

Distribution conditionnelle de Y selon $X = x_p$

Fréquences conditionnelles

Les fréquences conditionnelles correspondent au rapport de l'effectif partiel sur l'effectif marginal.

- La fréquence conditionnelle de x_i sachant que $Y = y_j$ se note $f_{i/j}$ et se lit « f indice i sachant j » est :

$$f_{i/j} = \frac{n_{ij}}{n_{\bullet j}}$$

- La fréquence conditionnelle de y_j sachant que $X = x_i$ se note $f_{j/i}$ et se lit « f indice j sachant i » est :

$$f_{j/i} = \frac{n_{ij}}{n_{i\bullet}}$$

Fréquences conditionnelles

y_j (Salaire) / x_i (L'âge)	$[6,8[$	$[8,10[$	$[10,12[$	Total
$[20,30[$	15	25	20	60
$[30,40[$	10	45	15	70
$[40,50[$	30	50	40	120
Total	55	120	75	250

- En considérant encore le tableau de l'exemple 4.1 on a :

$f_{1/2} = \frac{n_{12}}{n_{\bullet 2}} = \frac{25}{120} = 0,2083$

Cela signifie que : 20,83% des individus qui touchent entre 8000 dirhams et 10000 dirhams sont âgés entre 20 et 30 ans.

On a : $f_{j=3/i=1} = \frac{n_{13}}{n_{1\bullet}} = \frac{20}{60} = 0,3333$

Cela signifie que : 33,33% des individus qui sont âgés entre 20 et 30 ans touchent un salaire 10000 dirhams et 12000 dirhams .

Propriété des fréquences conditionnelles et relation entre les différentes fréquences

- La somme des fréquences conditionnelles sont égales à 1 :

$$\boxed{\sum_{i=1}^{p} f_{i/j} = 1 \quad \text{et} \quad \sum_{j=1}^{q} f_{j/i} = 1.}$$

- La relation entre les différentes fréquences s'exprime par :

$$f_{ij} = f_{i/j} \times f_{\bullet j} = f_{j/i} \times f_{i\bullet}$$

ou encore :

$$\boxed{f_{i/j} = \frac{f_{ij}}{f_{\bullet j}} \quad et \quad f_{j/i} = \frac{f_{ij}}{f_{i\bullet}}}$$

3. Caractéristiques numériques des séries statistiques à deux variables

3.1 Moyennes et variances marginales

Les caractéristiques marginales de X (moyenne et variance) s'obtiennent à partir de la distribution marginale de X, c'est-à-dire à partir des p modalités (x_i) et des p effectifs marginaux $(n_{i\bullet})$.

Voir tableau de contingence

Définition : La moyenne marginale $\overline{\overline{x}}$

- Il s'agit d'une moyenne arithmétique, que l'on note $\overline{\overline{x}}$ (lue :« x double barre »).
- Sa détermination s'obtient de la même logique que celle de $\overline{x}$ pour les séries à une dimension :

$$\boxed{\overline{\overline{x}} = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{p} n_{i\bullet} x_i \quad \text{avec } n_{\bullet\bullet} = \sum_{i=1}^{p} n_{i\bullet}}$$

La moyenne marginale de X : $\overline{\overline{x}}$

Dans l'exemple 4.1, à partir du tableau de contingence, on a :

$$\overline{\overline{x}} = \frac{1}{250}[(60 \times 25) + (70 \times 35) + (120 \times 45)] = 37,4$$

L'âge moyen des individus est donc de 37,4 ans (soit 37 ans, 4 mois et 24 jours)

3. Caractéristiques numériques des séries statistiques à deux variables

Définition : La variance marginale de $X : (V(x))$

- La formule de définition est la même que pour les séries à une dimension.

$$V(x) = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{p} n_{i\bullet} \left(x_i - \overline{\overline{x}}\right)^2$$

- La formule développée est

$$V(x) = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{p} n_{i\bullet} x_i^2 - \overline{\overline{x}}^2$$

- L'écart type de X est : $\sigma_x = \sqrt{V(x)}$.

3. Caractéristiques numériques des séries statistiques à deux variables

La variance marginale de $X : (V(x))$

Dans l'exemple 4.1, la variance marginale de X est :

$$\begin{cases} \dfrac{1}{n_{\bullet\bullet}} \displaystyle\sum_{i=1}^{3} n_{i\bullet} x_i^2 = \dfrac{1}{250} \left[\left(60 \times 25^2\right) + \left(70 \times 35^2\right) + \left(120 \times 45^2\right) \right] = 1465 \\ \\ \overline{x}^2 = (37,4)^2 = 1398,76 \end{cases}$$

et $V(x) = 1465 - 1398,76 = 66,24$

$\sigma_x = 8,14 \, \text{ans}$

3. Caractéristiques numériques des séries statistiques à deux variables

Caractéristiques marginales de Y :

- Comme précédemment, la moyenne marginale de Y s'obtiennent à partir de la distribution marginale de Y, c'est-à-dire à partir des q modalités (y_j) et des q effectifs marginaux $(n_{\bullet j})$.

$$\boxed{\bar{\bar{y}} = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{q} n_{\bullet j} y_j \quad \text{avec } n_{\bullet\bullet} = \sum_{j=1}^{q} n_{\bullet j}}$$

- La Variance de Y s'écrit :

$$\boxed{V(y) = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{q} n_{\bullet j} \left(y_j - \bar{\bar{y}}\right)^2}$$

- La formule développée est

$$\boxed{V(y) = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{q} n_{\bullet j} y_j^2 - \bar{\bar{y}}^2}$$

- L'écart type de Y est : $\sigma_y = \sqrt{V(y)}$.

Exercice 1

Corrigé

Pour 25 mariages, les âges de l'époux et de l'épouse, relevés sur les registres d'un état civil sont les suivants :

Sachant que chaque couplet (x_i, y_j) représente respectivement l'âge de l'époux et l'âge de l'épouse au moment du mariage.

$(22,17)$;	$(23,18)$;	$(24,17)$;	$(24,18)$;	$(24,20)$;	$(24,21)$
$(25,18)$;	$(25,19)$;	$(25,20)$;	$(26,18)$;	$(26,19)$;	$(26,21)$
$(26,23)$;	$(27,19)$;	$(27,21)$;	$(28,21)$;	$(28,22)$;	$(30,22)$
$(30,23)$;	$(31,24)$;	$(31,25)$;	$(34,24)$;	$(35,24)$;	$(35,25)$
$(36,25)$.					

1. Rangez ces données en classes d'amplitude 5 dans un tableau de contingence
2. Dégager les distributions marginales.
3. Calculez l'âge moyen des époux et l'âge moyen des épouses au moment du mariage.

3.2 Moyennes et variances conditionnelles

- Les caractéristiques conditionnelles de X selon Y (moyenne, variance) s'obtiennent à partir des p modalités de X et des q colonnes d'effectifs, à l'intérieur du tableau de contingence.
- Rappelons qu'il y a en effet q distributions conditionnelles de X selon Y (du fait des q modalités du caractère Y).

Voir tableau de contingence

Définition : Les moyennes conditionnelles de X selon Y

La moyenne de X sachant que $Y = y_j$ (ou « de X si $Y = y_j$ ») s'écrit :

$$\boxed{\overline{x}_j = \frac{1}{n_{\bullet j}} \sum_{i=1}^{p} n_{ij} x_i \quad \text{avec } 1 \leqslant j \leqslant q}$$

3. Caractéristiques numériques des séries statistiques à deux variables

3.2 Moyennes et variances conditionnelles

Dans l'exemple 4.1, la moyenne conditionnelle $\overline{x}_3$ qui désigne la moyenne de X sachant que $Y = y_3$ est :

$$\overline{x}_3 = \frac{1}{75}[(20 \times 25) + (15 \times 35) + (40 \times 45)] = 37,66$$

L'âge moyen des individus qui gagnent entre 10000 DH et 12000 DH est donc de 37,66 ans ($\approx$ 37 ans et 8 mois)

3.2 Moyennes et variances conditionnelles

La variance conditionnelle de X sachant que $Y = y_j$ est :

$$\begin{aligned} V_j(x) &= \frac{1}{n_{\bullet j}} \sum_{i=1}^{p} n_{ij} \left(x_i - \overline{x}_j\right)^2 \\ &= \sum_{i=1}^{p} f_{i/j} \left(x_i - \overline{x}_j\right)^2 \end{aligned}$$

Ou bien, par la formule développée :

$$\begin{aligned} V_j(x) &= \frac{1}{n_{\bullet j}} \sum_{i=1}^{p} n_{ij} x_i^2 - \overline{x}_j^2 \\ &= \sum_{i=1}^{p} f_{i/j} x_i^2 - \overline{x}_j^2. \end{aligned}$$

3.2 Moyennes et variances conditionnelles

Dans l'exemple 4.1, la variance conditionnelle de X si $Y = y_3$ est $V_3(x)$

$$\begin{cases} \dfrac{1}{n_{\bullet 3}} \displaystyle\sum_{i=1}^{3} n_{i3} x_i^2 = \dfrac{1}{75}\left[\left(20 \times 25^2\right) + \left(15 \times 35^2\right) + \left(40 \times 45^2\right)\right] = 1491,66 \\ \\ (\overline{x}_3)^2 = (37,66)^2 = 1418,27 \end{cases}$$

Donc

$$V_3(x) = 1491,66 - 1418,27 = 73,39$$

Et

$$\sigma_3(x) = \sqrt{73,39} = 8,56$$

est l'écart type des individus qui gagnent entre 10000 DH et 12000 DH.

Caractéristiques conditionnelles de Y selon X :

Comme précédemment, les caractéristiques conditionnelles de Y selon X s'obtiennent à partir des q modalités de Y et des p colonnes d'effectifs, à l'intérieur du tableau de contingence.

- La moyenne de Y sachant que $X = x_i$ (ou « de Y si $X = x_i$ ») s'écrit :

$$\overline{y}_i = \frac{1}{n_{i\bullet}} \sum_{j=1}^{q} n_{ij} y_j \quad \text{avec } 1 \leqslant i \leqslant p$$

- La variance conditionnelle de Y sachant que $X = x_i$ est :

$$V_i(y) = \frac{1}{n_{i\bullet}} \sum_{j=1}^{q} n_{ij} (y_j - \overline{y}_i)^2 = \sum_{j=1}^{q} f_{j/i} (y_j - \overline{y}_i)^2$$

- Ou bien, par la formule développée :

$$V_i(y) = \frac{1}{n_{i\bullet}} \sum_{j=1}^{q} n_{ij} y_j^2 - \overline{y}_i^2 = \sum_{j=1}^{q} f_{j/i} y_j^2 - \overline{y}_i^2.$$

4. Relations entre les caractéristiques marginales et conditionnelles

4.1 Relation entre les moyennes

- La moyenne marginale est égale à la moyenne des moyennes conditionnelles, pondérée par les effectifs marginaux :

$$\boxed{\overline{\overline{x}} = \frac{1}{n_{\bullet\bullet}} \sum_j n_{\bullet j} \overline{x}_j}$$

et

$$\boxed{\overline{\overline{y}} = \frac{1}{n_{\bullet\bullet}} \sum_i n_{i\bullet} \overline{y}_i}$$

4.2 Relation entre les variances

- La variance marginale est égale à la moyenne des variances conditionnelles, augmentée de la variance des moyennes conditionnelles.

$$V(x) = \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_j n_{\bullet j} \left(\overline{x}_j - \overline{\overline{x}}\right)^2}_{} + \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_j n_{\bullet j} V_j(x)}_{}$$

$$\left(\begin{array}{c} \text{Variance des moyennes} \\ \text{conditionnelles } \overline{x}_j \end{array} \right) + \left(\begin{array}{c} \text{Moyenne des variances} \\ \text{conditionnelles } V_j(x) \end{array} \right)$$

- Autre notation :

$$V(x) = V\left(\overline{x}_j\right) + \overline{V_j(x)}$$

4.2 Relation entre les variances

- De même pour la variable Y :

$$V(y) = \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_i n_{i\bullet} \left(\overline{y}_i - \overline{\overline{y}}\right)^2}_{} + \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_i n_{i\bullet} V_i(y)}_{}$$

$$\left(\begin{array}{c} \text{Variance des moyennes} \\ \text{conditionnelles } \overline{y}_i \end{array} \right) + \left(\begin{array}{c} \text{Moyenne des variances} \\ \text{conditionnelles } V_i(y) \end{array} \right)$$

- Autre notation :

$$V(y) = V(\overline{y}_i) + \overline{V_i(y)}$$

5. La covariance

Définition : Covariance de deux variables

- Pour mesurer la dépendance entre les deux variables X et Y, on calcule la covariance, qui généralise à deux variables la notion de variance. Sa formule s'exprime par :

$$\begin{aligned} Cov(x,y) &= \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}(x_i - \overline{\overline{x}})(y_j - \overline{\overline{y}}) \\ &= \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij}(x_i - \overline{\overline{x}})(y_j - \overline{\overline{y}}) \end{aligned}$$

ou encore, par la formule développée suivante : (plus pratique)

$$\begin{aligned} Cov(x,y) &= \left(\frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij} x_i y_j \right) - \overline{\overline{x}}.\overline{\overline{y}} \\ &= \left(\sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} x_i y_j \right) - \overline{\overline{x}}.\overline{\overline{y}} \end{aligned}$$

Exercice 2

Corrigé

Le tableau de contingence suivant donne la répartition des âges de l'époux et de l'épouse, relevés sur les registres d'un état civil :

x_i \ y_j	$[15;20[$	$[20;25]$	$[25;30]$	$n_{i\bullet}$
$[20;25[$	4	2	0	6
$[25;30[$	5	6	0	11
$[30;35[$	0	4.	1	5
$[35;40[$	0	1	2	3
$n_{\bullet j}$	9	13	3	25

1. Donner les valeurs des effectifs partiels suivants : n_{11} ; n_{23} ; n_{32} ; n_{21}.
2. Calculer $f_{2\bullet}$, $f_{\bullet 2}$, et donner leur signification.
3. Calculer f_{22}, donner sa signification
4. Calculer $f_{i=3/j=2}$ et donner sa signification

Exercice 3

Corrigé

En reprenant les données du tableau de contingence de l'exercice 1, calculer la $Cov(x, y)$

x_i \ y_j	$[15; 20[$	$[20; 25]$	$[25; 30]$	$n_{i\bullet}$
$[20; 25[$	4	2	0	6
$[25; 30[$	5	6	0	11
$[30; 35[$	0	4.	1	5
$[35; 40[$	0	1	2	3
$n_{\bullet j}$	9	13	3	25

Retour

1.
On choisit des bornes de classes divisibles par 5, le tableau de contingence est le suivant :

x_i \ y_j	$[15;20[$	$[20;25]$	$[25:30]$	$n_{i\bullet}$
$[20;25[$	4	2	0	6
$[25:30[$	5	6	0	11
$[30;35[$	0	4.	1	5
$[35;40[$	0	1	2	3
$n_{\bullet j}$	9	13	3	25

Solution de l'Exercice 1

2.

- La colonne $n_{i\bullet}$ associée aux x_i est la distribution marginale de X.
  Elle exprime l'âge des époux quel que soit l'âge des épouses.
  La distribution marginale de X est $\{x_i, n_{i\bullet}\}$
- La distribution marginale de Y est $\{y_j, n_{\bullet j}\}$

3.

- L'âge moyen des époux au moment du mariage est :
  $$\overline{\overline{x}} = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{4} n_{i\bullet} x_i = \frac{(6 \times 22,5) + (11 \times 27,5) + (5 \times 32,5) + (3 \times 37,5)}{25} = 28,5$$
  Il est donc de 28 ans et 6 mois.
- L'âge moyen des épouses au moment du mariage est :
  $$\overline{\overline{y}} = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{3} n_{\bullet j} y_j = \frac{(9 \times 17,5) + (13 \times 22,5) + (3 \times 27,5)}{25} = 21,3$$
  Il est donc de 21 ans, 3 mois et 18 jours.

Solution de l'Exercice 2

Retour

1.

$n_{11} = 4; \quad n_{23} = 0 \quad ; \quad n_{32} = 4 \quad ; \quad n_{21} = 5$

2.

$f_{2\bullet} = \frac{n_{2\bullet}}{n_{\bullet\bullet}} = \frac{11}{25} = 0,44.$

Signification : 44% des hommes de l'échantillon se marient entre 25 et 30 ans.

$f_{\bullet 2} = \frac{n_{\bullet 2}}{n_{\bullet\bullet}} = \frac{13}{25} = 0,52.$

Signification : 52% des femmes de l'échantillon se marient entre 20 et 25 ans.

3.

$f_{22} = \frac{n_{22}}{n_{\bullet\bullet}} = \frac{6}{25} = 0,24.$

Signification : 24% des couples de l'échantillon se sont mariés entre 20 et 25 ans pour la femme et 25 et 30 ans pour l'homme.

4.
$f_{i=3/j=2} = \frac{n_{32}}{n_{\bullet\bullet}} = \frac{4}{13} = 0,308.$
Signification : $30,8\%$ des femmes de l'échantillon de 20 à 25 ans se sont mariées avec des hommes de 30 et 35 ans.

Solution de l'Exercice 3

Retour

En reprenant les données du tableau de contingence de l'exercice 1, calculons la $Cov(x,y)$

On a :

$$\begin{cases} \dfrac{1}{n_{\bullet\bullet}}\displaystyle\sum_{i=1}^{4}\sum_{j=1}^{3} n_{ij}x_i y_j \\ = \frac{1}{25}\left[(4\times 22,5\times 17,5)+(2\times 22,5\times 22,5)+(0\times 22,5\times 27,5)+(5\times 27,5\times 17,5)+\right. \\ (6\times 27,5\times 22,5)+(0\times 27,5\times 27,5)+(0\times 32,5\times 17,5)+(4\times 32,5\times 22,5)+ \\ \left.(1\times 32,5\times 27,5)+(0\times 37,5\times 17,5)+(1\times 37,5\times 22,5)+(2\times 37,5\times 27,5)\right] \\ = \frac{1}{25}\left[1575+1012,5+0+2406,25+3712,5+0+0+2925+893,75+0+843,75\right. \\ \left.+2062,5\right] \\ = 617,25 \\ \overline{\overline{x}}\times\overline{\overline{y}} = 28,5\times 21,3 = 607,05 \end{cases}$$

Donc

$$Cov(x,y) = 617,25 - 607,05 = 10,2$$