Chapter 3: Two-variable statistical series

A dataset with two variables contains what is called bivariate data. It's a statistical series where two characters are studied simultaneously, Xi and Yj. Each one can be either quantitative or qualitative.

Example 1

- A sample of 200 families, the number of children X and the number of rooms Y are studied simultaneously.
- A sample of 20 families, we study the monthly income X in Da and the monthly expenses Y.
- Students randomly selected from a section of L2 civil engineering, we observe the hair color X and the color of the eyes Y.

A.3.1 Data tables (contingency table):

In this type of table, the numbers **nij are partial counts**. These are the numbers of individuals that present both the xi and the yj modality.

Y X	y1	y2	y3	…………..	Yq	Marginal Column ni.
x1	n11	n12	n13	…………..	n1q	n1.
x2	n21	n22	n23	…………..	n2q	n2.
x3	n31	n32	n33	…………..	n3q	n3.
. .	. .	. .	. .	. .	. .	. .
xp	np1	np2	np3	…………..	npq	np.
Marginal line n.j	n.1	n.2	n.3	…………..	n.q	N=n..

Example 1 :

- A sample of 80 families, the number of children X and the number of rooms Y are studied simultaneously.

Number of rooms **Number of children**	**0**	**1**	**2**	**3**	**Total**
1	4	8	1	4	17
2	5	8	7	5	25
3	12	8	9	9	38
Total	21	24	17	18	80

Solution: Contingency Table

Y **X**	$y_{(j=1)}=0$	$y_{(j=2)}=1$	$y_{(j=3)}=2$	$y_{(j=4)}=3$	**Total**
$x_{(i=1)}=1$	**n11=4**	**n12=8**	**n13=1**	**n14=4**	**n1.=17**
$x_{(i=2)}=2$	**n21=5**	**n22=8**	**n23=7**	**n24=5**	**n2.=25**
$x_{(i=3)}=3$	**n31=12**	**n32=8**	**n33=9**	**n34=9**	**n3.=38**
Total	**n.1=21**	**n.2=24**	**n.3=17**	**n.4=18**	**N=n..=80**

- **n24** indicates that 5 families with 3 bedrooms and 2 children.

A.3.1.1. Simple Scatter plot:

The graphical representation of a two-variable statistical series is made by a scatter plot that is plotted, from the raw data of X and Y.

- If X and Y are continious quantitative variables, we plot the graph by their class centers Ci.

The coordinates of the mean point G are the arithmetic means of X and Y $(\bar{x} \ ; \ \bar{y})$.

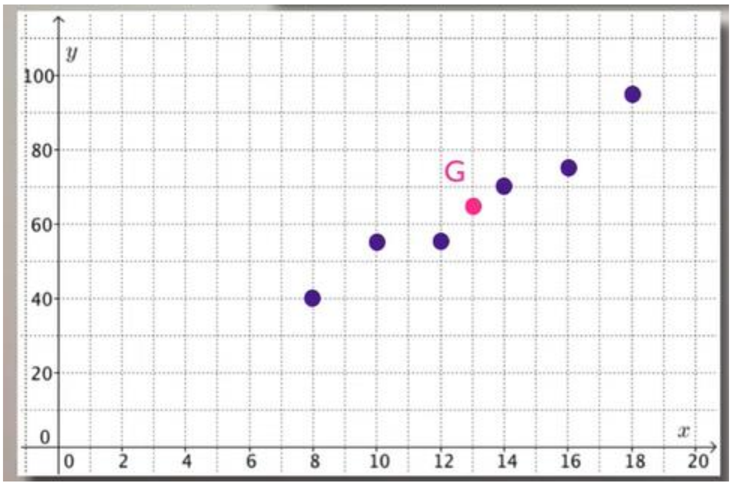
Example 2 :

The following table shows the evolution of a company's advertising budget and revenues over the last 6 years.

X Budget PUB $x10^3$	8	10	12	14	16	18
Y Revenue $x10^3$	40	55	55	70	75	95

1. In a coordinate system, represent the plot scatter (x,y).
2. Determine the coordinates of the mean point G of the plot scatter.

Solution:

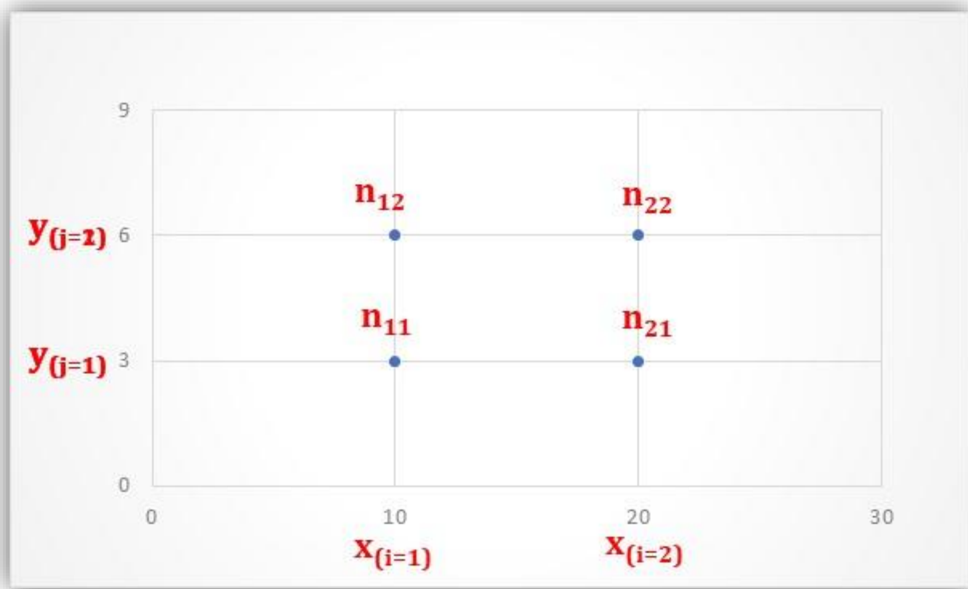


$$\bar{x} = \frac{8 + 10 + 12 + 14 + 16 + 18}{6} = 13$$

$$\bar{y} = \frac{40 + 55 + 55 + 70 + 75 + 95}{6} = 65$$

A.3.1.2. Weighted Scatter plot:

X \ Y	$y_{(j=1)}=3$	$y_{(j=2)}=6$	Total
$x_{(i=1)}=10$	n11	n12	n1.
$x_{(i=2)}=20$	n21	n22	n2.
Total	n.1	n.2	N=n..



A.3.2 Marginal and conditional distributions:

By examining the marginal distributions of the contingency table, we can obtain data solely for variable X or variable Y

A.3.2.1 Marginal distributions of X :

We take the first column and the last column. The counts (ni.) represent individuals with modality xi independently of the modalities of the second character studied Y. They are reported to define the marginal distribution of X. (This statistical series is a single-character statistical series).

Counts and marginal frequencies with respect to X : we have, for i = 1... p,

$$fi \bullet = \frac{ni \bullet}{N = n \bullet\bullet}$$

X	ni.	fi.
x1	n1.	f1.
x2	n2.	f2.
x3	n3.	f3.
. .	. .	. .
xp	np.	fp.
Total	N=n..	F=f..=1

A.3.2.2 Marginal distributions of Y:

We take the first line and the last line. Counts and marginal frequencies with respect to Y : we have, for i = 1... q,

$$f \bullet j = \frac{n \bullet j}{N = n \bullet\bullet}$$

Y	n.j	f.j
y1	n.1	f.1
y2	n.2	f.2
y3	n.3	f.3
. .	. .	. .
yq	n.q	f.q
Total	N=n..	F=f..=1

A.3.2.3 Conditional distributions of X:

We take the first column and the 2nd column will be taken according to the value of j. Notation X/(Y=yj). It is said to define the "conditional distribution of X given that Y=yj".

X/(Y=yj)	ni/j	fi/j
x1	n1j	f1j
x2	n2j	f2j
x3	n3j	f3j
. .	. .	. .
xp	npj	fpj
Total	N=n.j	F=f.j

A.3.2.4 Conditional distributions of Y:

We take the first line and the 2nd line will be taken according to the value of i. Notation Y/(X=xi). It is said to define the "Conditional distribution of Y given that X=xi".

Example 3:

- Déterminer la distribution conditionnelle de X pour $Y_{(j=2)}$.

X \ Y	3	6	Total
10	**n11**	**n12**	**n1.**
20	**n21**	**n22**	**n2.**
Total	**n.1**	**n.2**	**N=n..**

Solution :

X	$n_{i/j=2}$ $y_{(j=2)}=6$	fi/j=2
$x_{(i=1)}=10$	**n12**	**f12**
$x_{(i=2)}=20$	**n22**	**f22**
Total	**n.2**	**f..**

A.3.2.5 Statistical independence:

X and Y are independent if the counts observed nij are identical to the counts of Independence.

$$nij = \frac{(ni\bullet * n\bullet j)}{n\bullet\bullet}$$ or $$fij = fi \bullet * f \bullet j$$

Example 3 :

A survey was carried out on 100 families by observing "monthly expenditure" X and "monthly income" Y (in thousands of DA), the results are given in the table below.

Y / X	**[25-40[**	**[40-55[**	**[55-70]**	**Total**
[20-25[	n11=30	n12=20	n13=10	n1•=60
[25-30[	n21=10	n22=30	n23=10	n2•=50
[30-35]	n31=10	n32=10	n33=20	n3•=40
Total	n•1=50	n•2=60	n•3=40	n••=150

- Check the dependence of the variables X and Y for i=2 and j=1.

Solution:

For the independence of the variables X and Y, for i=2 and j=1, we obtain:

$$n21 = \frac{(n2 \bullet * \ n \bullet 1)}{n \bullet\bullet}$$

$n21 * n \bullet\bullet = n2 \bullet * n \bullet 1$

$n21 * n \bullet\bullet = 10 \ * 150 = 1500$

$n2 \bullet * \ n \bullet 1 = 50 * 50 = 2500$

Since $n21 * n \bullet\bullet \neq n2 \bullet * n \bullet 1$, Therefore, X and Y are not independent