



SÉRIES STATISTIQUES À DEUX VARIABLES

D) Série statistique à deux variables

Une série statistique à deux variables est une série pour laquelle deux caractères mesurables sont relevés pour chaque individu.

L'étude statistique suivante porte sur une population de nouveau-nés. Deux caractères sont étudiés : la masse et la taille.

Enfant	1	2	3	4	5	6	7	8	9	10
Masse en kg	2,4	2,6	2,7	3,0	3,2	3,3	3,5	3,6	3,8	4
Taille en cm	45	47	48	50	51	52	53	54	54	56

Une série statistique à deux variables peut être rapportée par un tableau double entrée. L'étude statistique suivante porte sur une population de 100 ménages. Deux caractères sont étudiés : - le caractère X est le nombre d'enfants

- le caractère Y est le nombre de pièces de l'appartement occupé.

$y_i \backslash x_i$	0	1	2	3	4	5	
1	6	4	1	0	0	0	11
2	3	11	10	5	1	0	30
3	1	3	16	13	4	1	38
4	0	1	3	5	8	4	21
	10	19	30	23	13	5	100

LOI
M
A
R
G
I
N
A
L
E
de Y

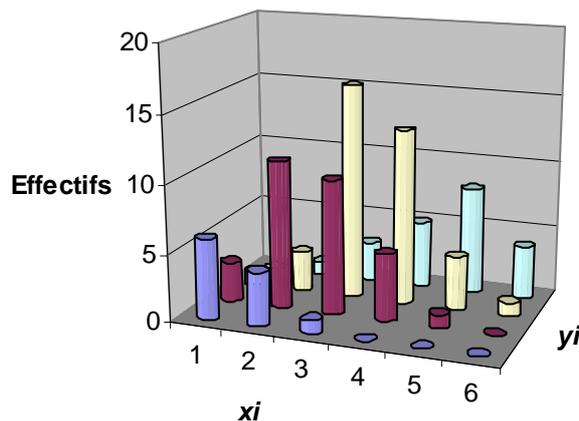
LOI MARGINALE de X

Si on additionne les nombres d'une même ligne du tableau, nous obtenons l'effectif marginal de y_i et l'ensemble de ces nombres constitue la loi marginale du caractère Y.

Si on additionne les nombres d'une même colonne du tableau, nous obtenons l'effectif marginal de x_i et l'ensemble de ces nombres constitue la loi marginale du caractère X.

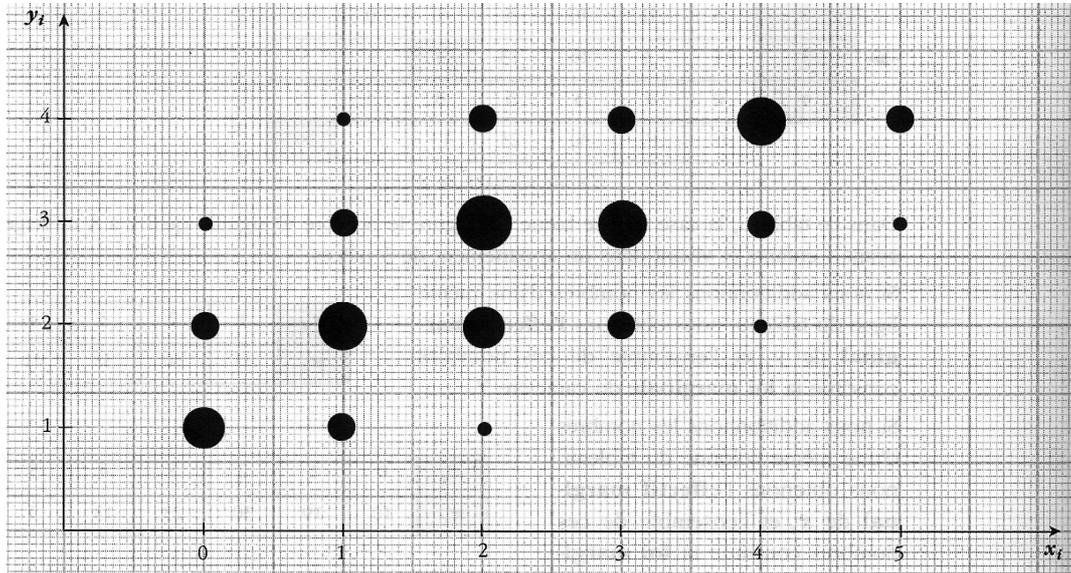
On appelle stéréogramme une représentation spatiale en perspective. Sur la case $(x_i ; y_i)$ on dessine une barre de hauteur n_i .

ETUDE STATISTIQUE SUR 100 MENAGES





La représentation graphique, dans un repère orthogonal, d'une série statistique double, dont les caractères sont quantitatifs, forme un nuage de points.



II) Point moyen

Le point moyen d'un nuage de points a pour :

- abscisse, la moyenne des abscisses des points constituant le nuage ;
- ordonné, la moyenne des ordonnées.

Ce point est le point $G(\bar{x} ; \bar{y})$ avec $\bar{x} = \frac{1}{n} \sum n_i x_i$ et $\bar{y} = \frac{1}{n} \sum n_i y_i$

III) Droite d'ajustement

L'intérêt d'une telle étude réside dans la recherche du lien éventuel existant entre les deux caractères étudiés. On peut mathématiser cette liaison en établissant une loi permettant de prédire avec un faible degré d'incertitude la valeur d'un caractère en fonction de la valeur de l'autre caractère. On réalise alors un ajustement affine. Plusieurs méthodes sont possibles :

- les méthodes d'ajustement graphiques ;
- la méthode de Mayer ;
- la méthode des moindres carrés.

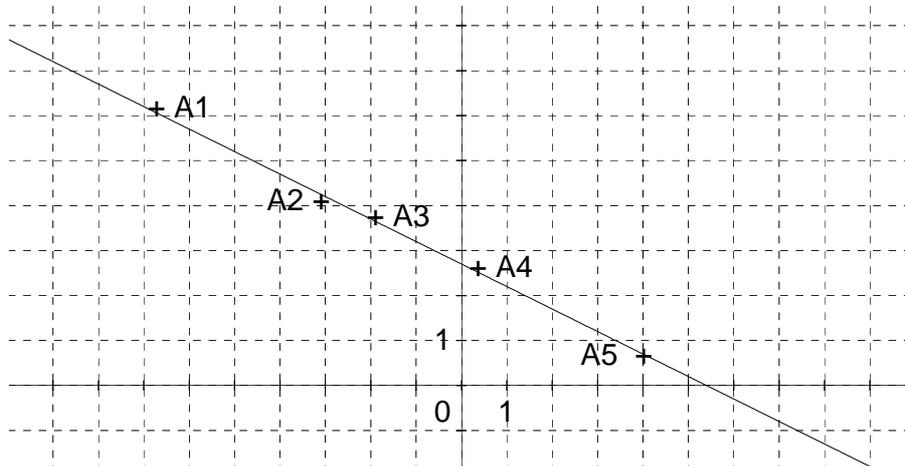


• **La méthode d'ajustement affine**

La forme du nuage nous permet de dire qu'il existe une fonction affine $f: x \mapsto ax + b$ qui lie les valeurs x_i aux valeurs y_i .

On trace une droite d'ajustement par une méthode empirique, en dessinant une droite qui passe par G et qui passe au plus près des six points du nuage.

Cette méthode n'est pas complètement fautive mais dépend de l'opérateur



• **La méthode de Mayer**

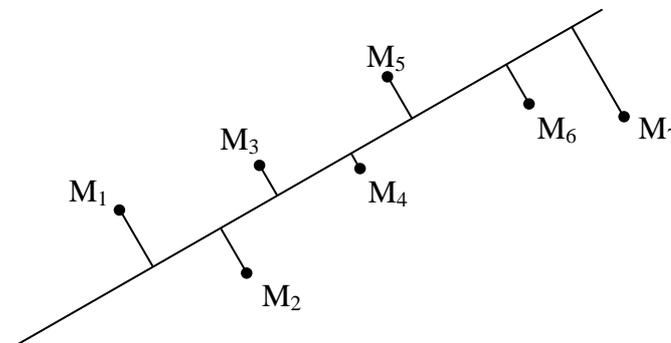
Elle consiste à :

- fractionner le nuage en deux nuages dont les effectifs sont égaux ou diffèrent de 1 ;
- calculer les coordonnées des points moyens G_1 et G_2 des deux parties du nuage.
- Déterminer l'équation de la droite (G_1G_2) .

• **La méthode des moindres carrés**

Elle consiste à rechercher la droite telle que la somme des carrés des distances des points M_i ($1 \leq i \leq n$) à la droite soit minimale.

C'est la raison pour laquelle cette méthode est appelée méthode des moindres carrés.



La droite d'ajustement permet d'estimer la valeur d'un caractère quand on connaît la valeur du deuxième caractère ou d'établir des prévisions.