Série Statistique à deux variables

1. Nuage de points

Le plan est muni d'un repère orthogonal .

Pour une population donnée, on peut s'intéresser simultanément à deux caractères x et y.

Les valeurs prises par x sont $x_1, x_2, \ldots, x_n$.

Les valeurs prises par y sont $y_1, y_2 \ldots, y_n$.

La représentation graphique des points Mi (xi, yi) est le nuage de points associé à cette série statistique.

2. Point moyen

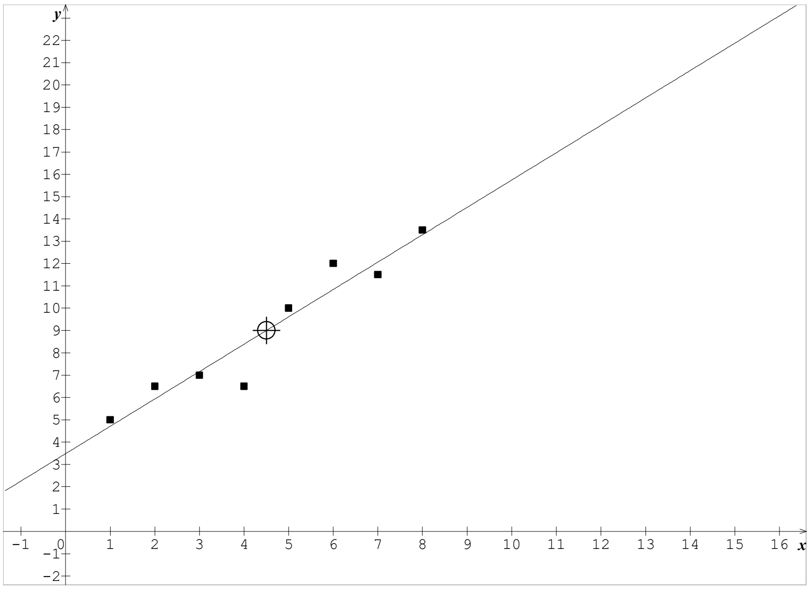
Le point moyen du nuage de points associé à la série statistique double (x_i; y_i) est le point G ($\bar{x} ; \bar{y}$).

où $\bar{x}$ et $\bar{y}$ sont les moyennes respectives des variables x_i et y_i .

Exemple

Le tableau suivant représente le montant en million d'Ariary des importations d'une société pendant 8 années consécutives :

Rang de l'année x_i	1	2	3	4	5	6	7	8
Montant y_i	5	6,5	7	6,5	10	12	11,5	13,5



3. Covariance

La covariance de la série statistique double est le nombre Cov (X, Y) définie par :

$$\text{cov (X, Y)} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}$$

On démontre que la covariance joue un rôle analogue à la variance dans le cas d'une série statistique simple . Si on fait X= Y

on obtient la formule de la variance .

Exemple

Dans l'exemple de la série statistique précédente, on a $\sum_{i=1}^{n} x_i \cdot y_i$ = 1.5 +2.6,5 +...+8.13,5 = 375,5 ;

Cov(X, Y) = $\frac{375,5}{8}$ - 4,5 .9 ; on a Cov(X,Y) = 6,44

4. Coefficient de corrélation linéaire

On appelle coefficient de corrélation linéaire entre X et Y le réel défini par :

$$r = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$$

- Le coefficient de corrélation linéaire r vérifie : $-1 \leq r \leq 1$
- La corrélation entre X et Y est forte si $|r| \geq 0,866$

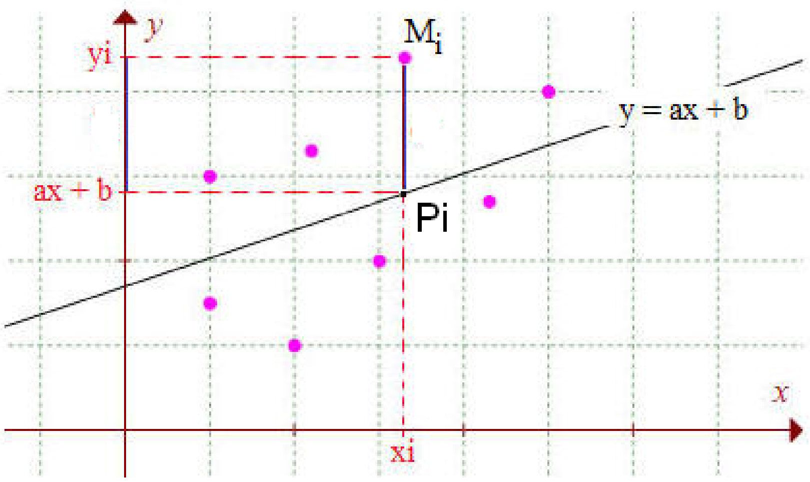
Un ajustement affine entre X et Y consiste à déterminer une droite qui passe le plus près possible des différents nuage points associé à la série (X, Y). Cette droite est alors utilisée pour estimer la valeur de l'une des variables X ou Y connaissant l'autre.

Si la corrélation entre X et Y est forte, on peut effectuer un ajustement affine entre X et Y.

5. Ajustement affine par la méthode des moindre carrées

5.1 Le principe de la méthode

On considère dans le plan muni d'un repère orthogonal le nuage de points associé à la série statistique double (X, Y). On trace un droite (D) : y = ax + b et on note P_i les projetés de M_i sur (D) parallèlement à l'axe des ordonnées. M_iP_i représente la distance entre les point M_i et P_i .



5.2 Droite de régression de y en x

On appelle droite de régression de y en x (ou droite d'ajustement de y en x) par la méthode des moindres carrés la droite (D) telle que la somme $M_1P_1{}^2 + M_2P_2{}^2 + ...+M_nP_n{}^2$ soit minimale.

Cette droite passe par le point moyen G

La droite de régression de y en x a pour équation $y - \bar{y} = a(x - \bar{x})$ avec :

$$\frac{Cov(X,Y)}{V(X)}$$

Exemple

reprenons l'exemple de la série statistique étudiée dans ce chapitre, on sait que :

Cov(X, Y) = 6,44 V(X) = 5,25 et en tenant compte des résultats $\bar{x}$ =4,5 et $\bar{y}$ =9

(D) : y = 1,23x + 3,47.

5.3 Droite de régression de x en y

La droite de régression de x en y a pour équation $x - \bar{x} = a'(y - \bar{y})$ avec :

$$a' = \frac{Cov(X,Y)}{V(Y)}$$

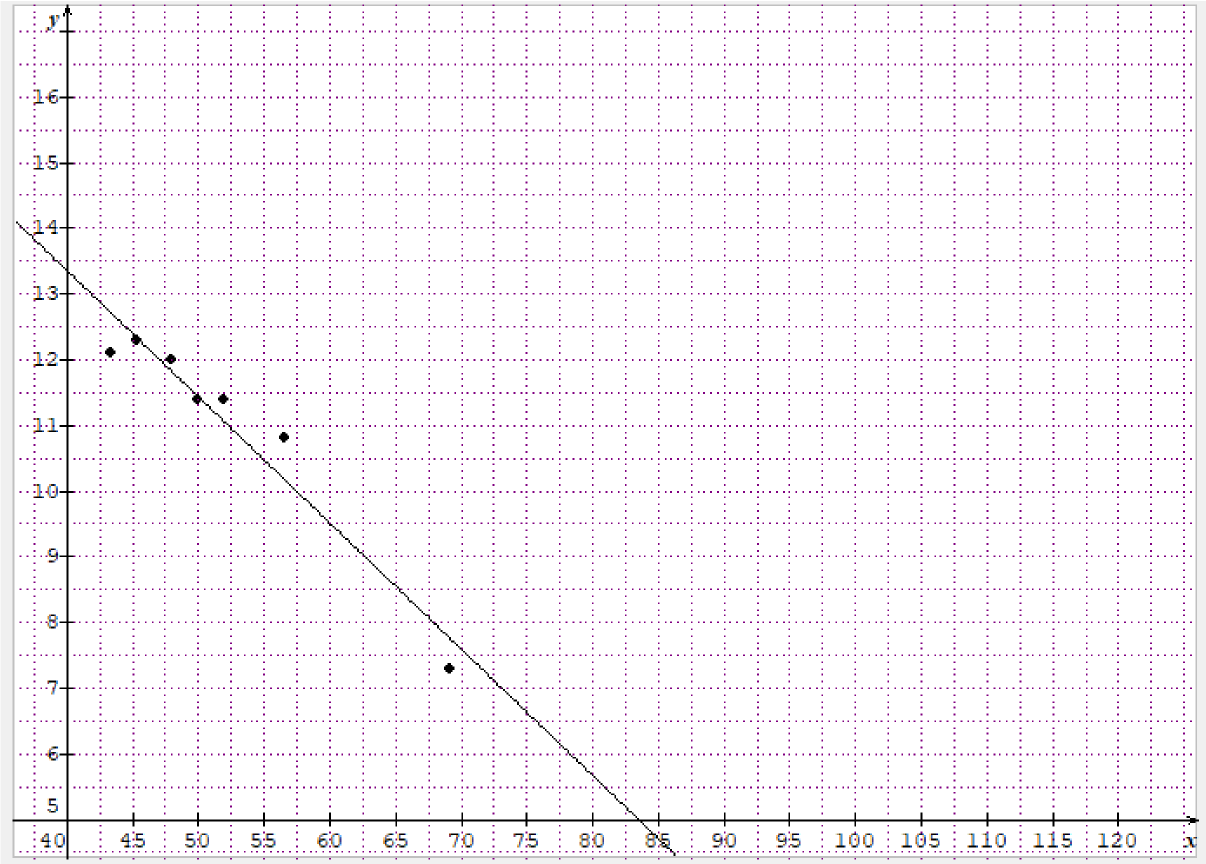
Exemple

Considérons La série statistiques double représenté par le tableau suivant.

xi	69,1	56,6	52,	50	48	45,3	43,3
yi	7,3	10,8	11,4	11,4	12	12,3	12,1

1) Représenter le nuage de points associé à cette série.
2) Calculer les coordonnées du point moyen G
3) Calculer le coefficient de corrélation r.
4) Déterminer l'équation de la droite de régression de y en x.
5) estimer la valeur de y pour x = 40.

1) Nuage de points



2) Coordonnées de G

avec sinéquanon, G(52,0429 ; 11,0429)

3) Coefficient de corrélation

$$r = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)} = 0{,}967072$$

4) Droite de régression de y en x

(D) : y = - 0,19x + 21,05

5) Estimation pour x = 40

si x = 40 ; y = - 0,19(40) + 21,05 ; y = 13,45