

Chapitre 5

Statistiques à deux variables

5.1 Tableau de données et nuages de points

Il existe parfois une relation entre deux caractères d'une population.

On définit alors une série statistique à deux variables (ou série statistique double) x et y , prenant les valeurs x_1, \dots, x_n et y_1, \dots, y_n .

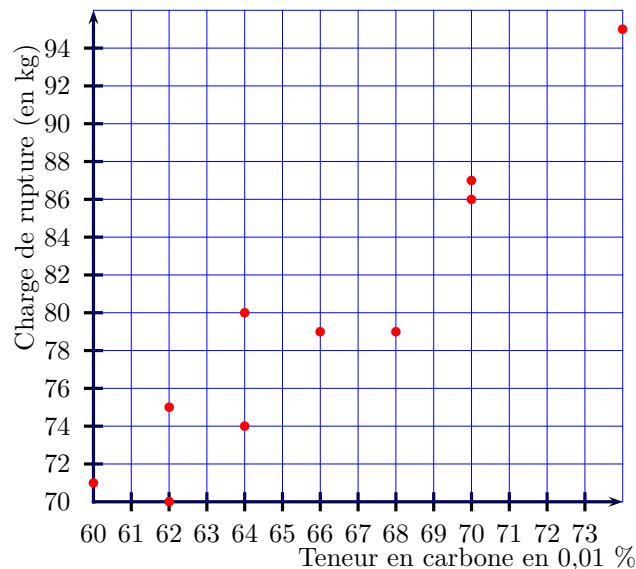
Tableau de données :

Le tableau suivant donne les résultats obtenus à partir de 10 essais de laboratoire concernant la charge de rupture d'un acier en fonction de sa teneur en carbone.

Teneur en carbone x_i (en 0,01 %)	70	60	68	64	66	64	62	70	74	62
Charge de rupture y_i (en kg)	87	71	79	74	79	80	75	86	95	70

Nuages de points :

Dans le plan muni d'un repère, le **nuage de points** de la série statistique est l'ensemble des points du plan de coordonnées $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$.



Point moyen

On appelle **point moyen** d'un nuage de n points $M_i(x_i; y_i)$ le point G de coordonnées $(\bar{x}; \bar{y})$:

$$\begin{cases} x_G = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ y_G = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases}$$

Dans l'exemple donné les coordonnées du point moyen sont $G(66; 79,6)$.

5.2 Ajustements affines

Lorsque les points du nuage semblent alignés, on recherche une droite passant au plus près de tous les points. On cherche à déterminer une équation de la forme $y = ax + b$ de cette droite.

La fonction affine f qui à x associe $f(x) = ax + b$ est un **ajustement affine** de la série statistique.

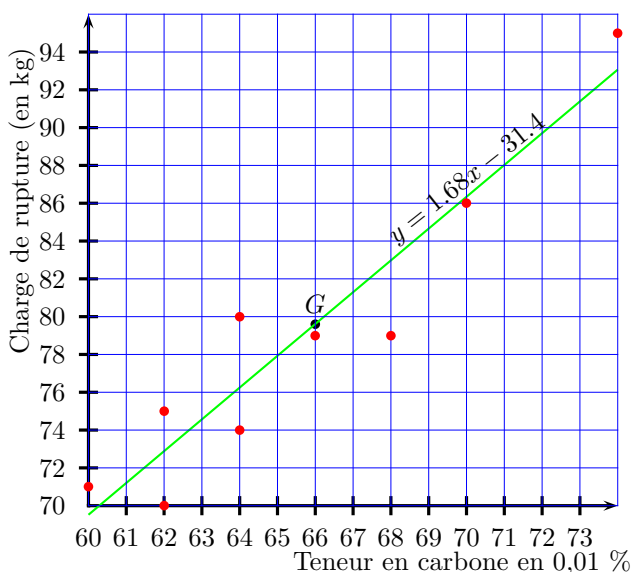
Méthode graphique

On trace au jugé une droite D (on peut s'aider du point moyen) en s'efforçant d'équilibrer le nombre de points situés de part et d'autre.

Ensuite on détermine par lecture graphique les réels a et b . Cette méthode est peu précise.

Méthode des moindres carrés

À l'aide d'une calculatrice ou d'un tableur, on détermine l'équation de la droite qui minimise les écarts du nuage de points à cette droite que l'on appelle **droite de régression**.



Voici un bref descriptif de l'emploi de certaines calculatrices, pour plus d'informations consulter le manuel de l'utilisateur.

Avec une TI

Appuyer sur la touche **STAT** puis choisir **1 : EDIT**
 Saisir les valeurs x_i et y_i dans les colonnes **L1** et **L2**.
 Appuyer sur la touche **STAT** puis dans **CALC**
 Pour obtenir l'équation de la droite de régression de y en x
 Choisir **4 : LinReg(ax+b) L1,L2**

Avec une Casio

Choisir le menu **STAT**
 Saisir les valeurs x_i et y_i dans les colonnes **List 1** et **List 2**.
 Paramétrer (uniquement la première fois) ainsi
CALC **SET** **2 VAR X List : List 1**
2 VAR Y List : List 2
 puis **EXE**
 Pour obtenir l'équation de la droite de régression de y en x
2 VAR puis **REG** **X**.

5.3 Estimations et prévisions

Lorsqu'il semble exister une relation affine entre les données x_i et y_i , on peut donner une estimation de y_0 connaissant x_0 .

Pour cela on peut :

- faire une lecture graphique sur la droite tracée ;
- faire le calcul en utilisant l'équation de la droite de régression $y = ax + b$.

On a pour une teneur en carbone de 77, $y = 1,682 \times 77 - 31,4 = 98,114$

Une estimation de sa charge de rupture est donc 98,114 kg.