

# Statistiques à deux variables

## 1 Définition - Nuage de points

Le principe est le suivant : on étudie à la fois deux caractères sur une même population. L'objectif est le plus souvent de déterminer s'il y a un lien statistique entre les deux, ce qu'on appelle une **corrélacion**.

On définit alors une série statistique à deux variables  $x$  et  $y$ , prenant les valeurs  $x_1, \dots, x_p$  et  $y_1, \dots, y_q$ . À chaque couple de valeurs  $(x_i; y_j)$ , on peut faire correspondre son effectif  $n_{ij}$ .

### Définition 1

Dans le plan muni d'un repère, on peut associer à chaque couple  $(x_i; y_i)$  de la série statistique, le point  $M_i$  de coordonnées  $(x_i; y_i)$ . L'ensemble des points  $M_i$  s'appelle le **nuage de points** représentant la série statistique.

### Définition 2

On appelle point moyen d'un nuage le point  $G$  de coordonnées  $(\bar{x}; \bar{y})$ .

**Exemple** : on mesure la taille et la masse d'une série de trucs ; on obtient le tableau suivant :

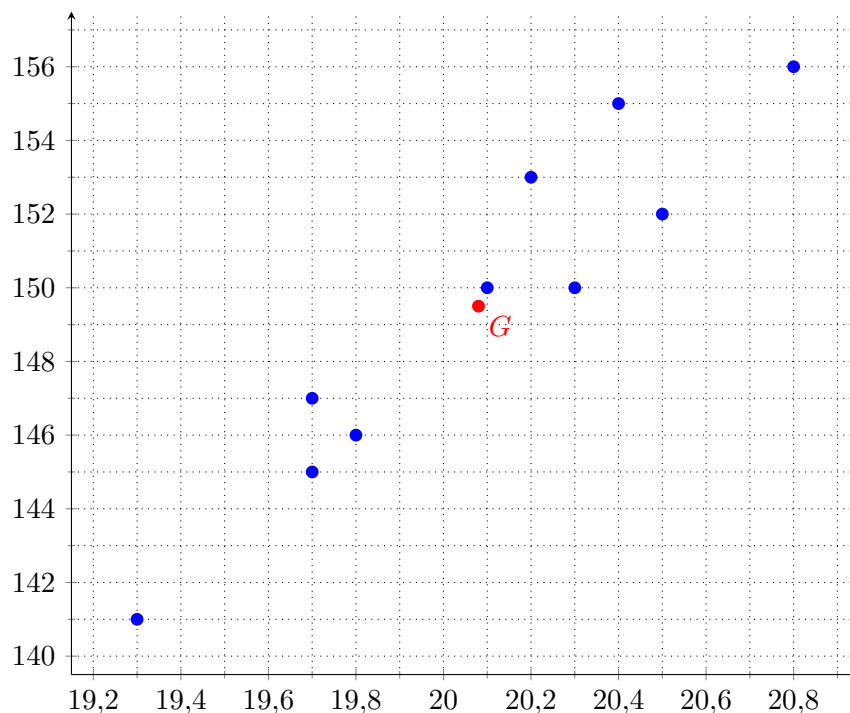
Truc n°	1	2	3	4	5	5	7	8	9	10
Taille (cm)	19,3	20,5	19,8	19,7	20,2	20,4	20,3	19,7	20,8	20,1
Masse (g)	141	152	146	147	153	155	150	145	156	150

On va donc placer les points de coordonnées  $(19,3; 141)$ ,  $(20,5; 152)$ , etc.

Calculons les coordonnées du point moyen :

$$\bar{x} = \frac{19,3 + 20,5 + 19,8 + 19,7 + 20,2 + 20,4 + 20,3 + 19,7 + 20,8 + 20,1}{10} = 20,08$$

$$\bar{y} = \frac{141 + 152 + 146 + 147 + 153 + 155 + 150 + 145 + 156 + 150}{10} = 149,5$$



## 2 Ajustement affine

### 2.1 Méthode graphique

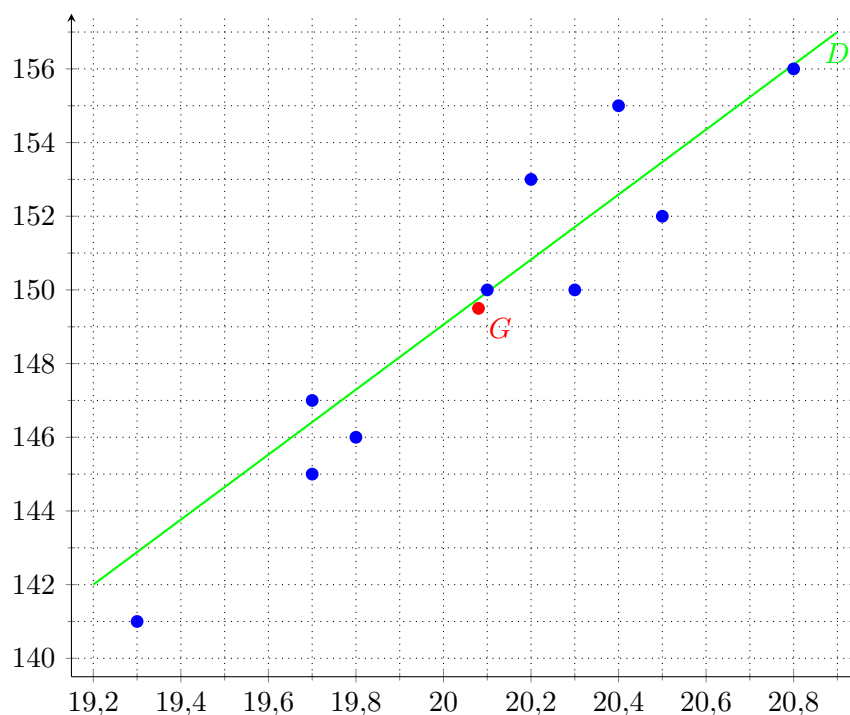
Lorsque les points  $M_i$  semblent à peu près alignés, on cherche à construire une droite qui représente au mieux la forme du nuage, c'est à dire qui passe « au plus près » de chacun des points.

L'équation de cette droite constitue une formule qui relie, de façon approchée, les deux variables  $x$  et  $y$ .

#### a. Ajustement à la règle :

On trace au jugé une droite  $D$  en s'efforçant d'équilibrer le nombre de points situés de part et d'autre. Ensuite on détermine son équation par lecture graphique.

Illustration, à partir de l'exemple précédent :



#### b. Ajustement par la méthode de Mayer :

On partage le nuage de points en deux nuages d'affectif à peu près égal, de part et d'autre du point moyen. On détermine les coordonnées des points moyens  $G_1$  et  $G_2$  de chacun de ces nuages.

La droite  $(G_1G_2)$ , appelée droite de **Mayer**, constitue en général une bonne droite d'ajustement.

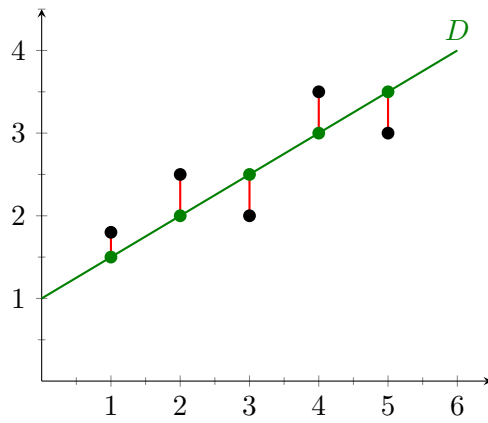
### 2.2 Méthode des moindres carrés - Droite de régression

#### Définition 3

Soit  $D$  une droite d'équation  $y = ax + b$ . Appelons  $P_i$  le point d'abscisse  $x_i$  situé sur la droite  $D$  (son ordonnée est donc  $ax_i + b$ ).

On appelle **droite de régression de  $y$  en  $x$**  la droite  $D$  choisie de telle sorte que la somme des  $M_i P_i^2$  soit minimale.

Sur le dessin, c'est la droite qui minimise la somme des carrés des distances en rouge.



Les points  $M_i$  ont pour coordonnées  $(x_i; y_i)$ ; les points  $P_i$  ont pour coordonnées  $(x_i; ax_i + b)$ .

La quantité minimisée est donc la somme des  $[y_i - (ax_i + b)]^2$ .

On démontre que la droite de régression de  $y$  en  $x$  passe toujours par le point moyen du nuage.

### 2.3 Utilisation de la calculatrice

On veut obtenir les coefficients  $a$  et  $b$  de la droite de régression après avoir entré les valeurs.

T.I.	Casio
Touche <b>STAT</b>	Menu <b>STAT</b>
Menu <b>EDIT</b>	Entrer les valeurs $x_i$ dans <i>List1</i>
Entrer les valeurs $x_i$ dans $L_1$	Entrer les valeurs $y_i$ dans <i>List2</i>
Entrer les valeurs $y_i$ dans $L_2$	Menu <b>STAT</b>
Touche <b>STAT</b>	Item <b>CALC</b>
Menu <b>CALC</b>	Régler les paramètres avec <b>set</b>
Item <b>LinReg</b>	Item <b>REG</b>
LinReg $L_1, L_2$	Choisir <b>X</b>

Si on reprend notre exemple, on trouve :  $a \simeq 9,967$  et  $b \simeq -50,64$ . On peut donc tracer  $D$ , la droite de régression de  $y$  en  $x$ , d'équation  $y = 9,967x - 50,64$ . On vérifie qu'elle passe par le point moyen.

