5.2 RÉGRESSION LINÉAIRE

cours 28

Il arrive souvent qu'on recueille plusieurs données sur une même population.

Âge, taille, poids, années d'études, salaire, etc.

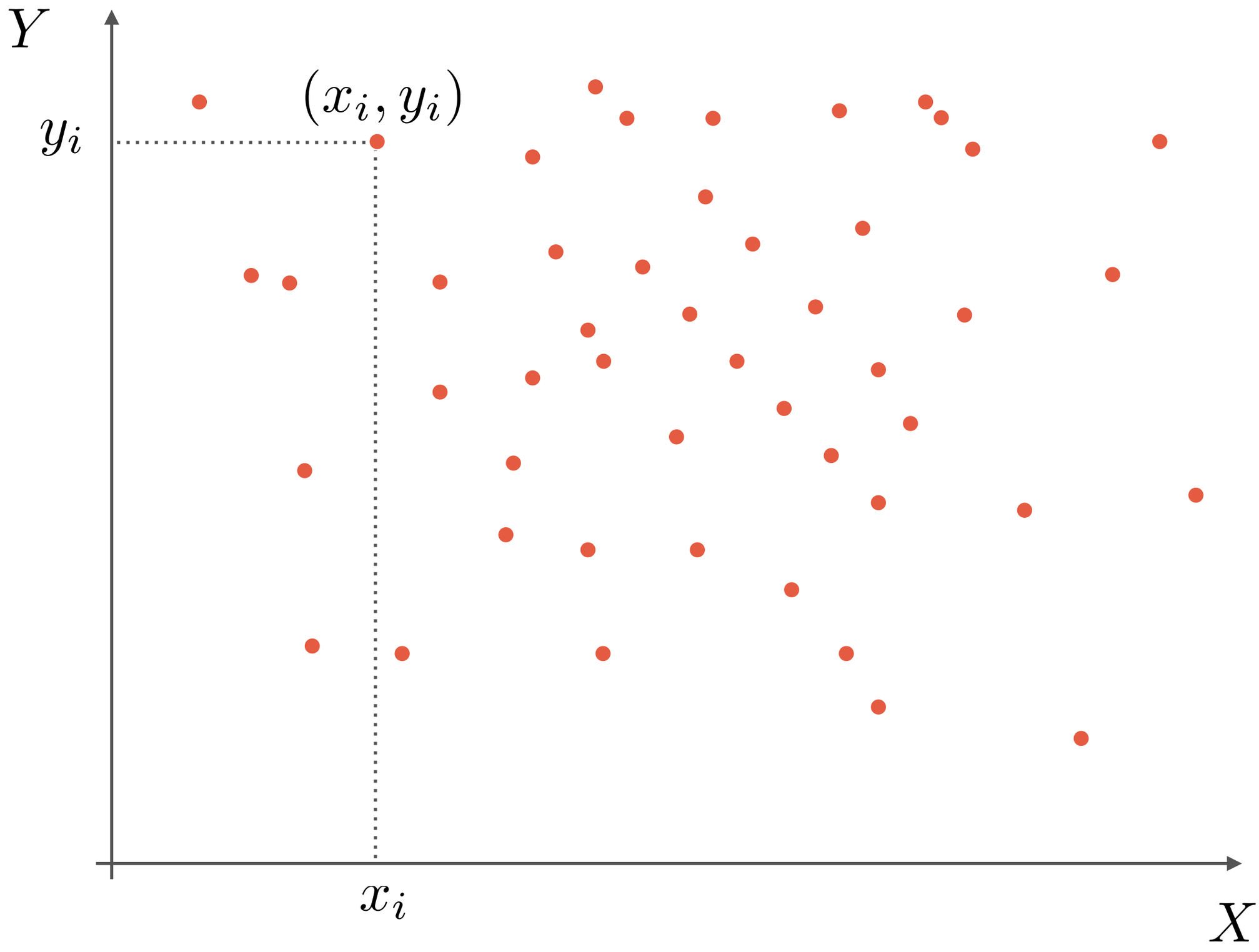
On se demande souvent s'il y a des liens entre ces variables statistiques.

Concentrons-nous sur deux variables statistiques X et Y d'une même population.

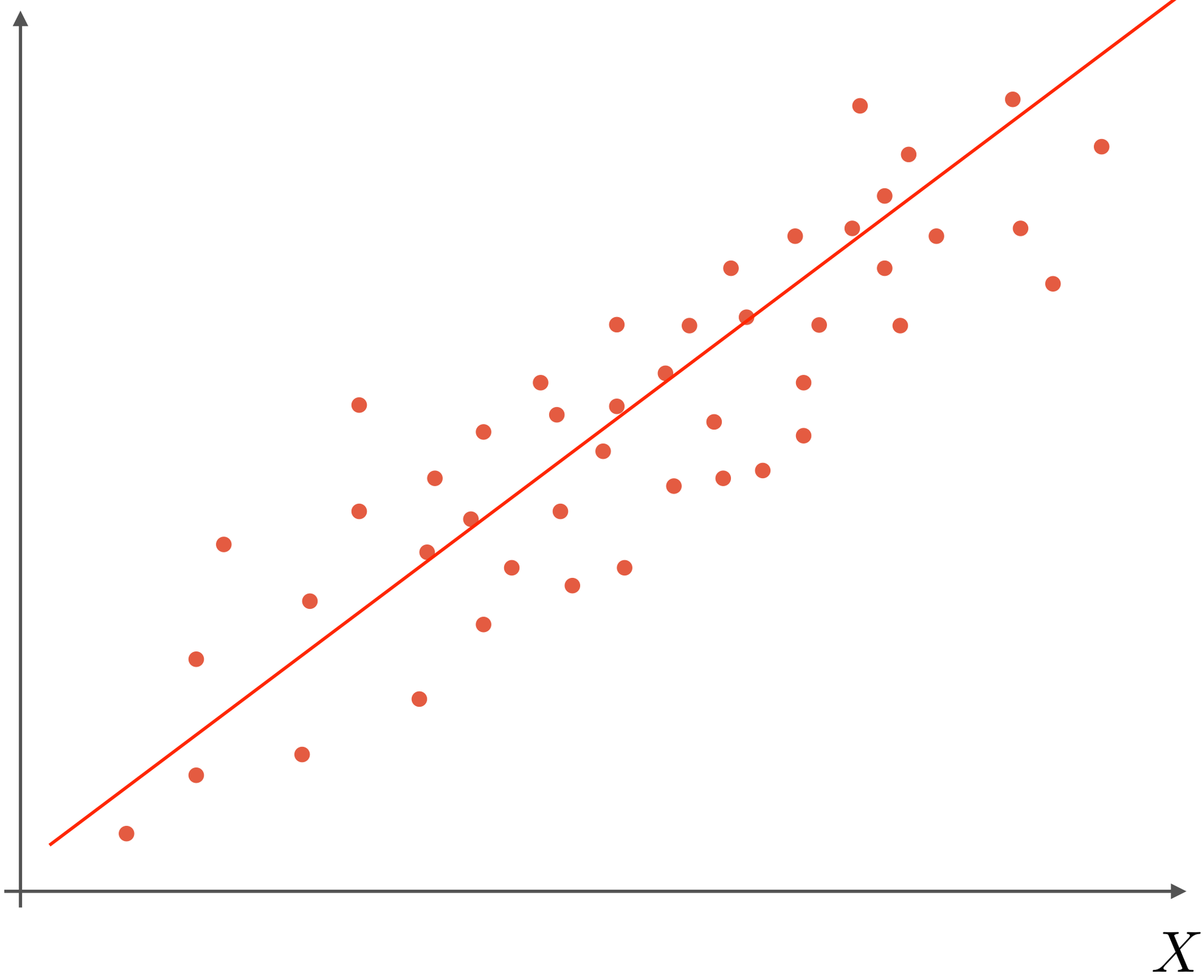
Pour chaque individu de la population ou d'un échantillon, associons-lui le couple de ses valeurs pour ces deux variables.

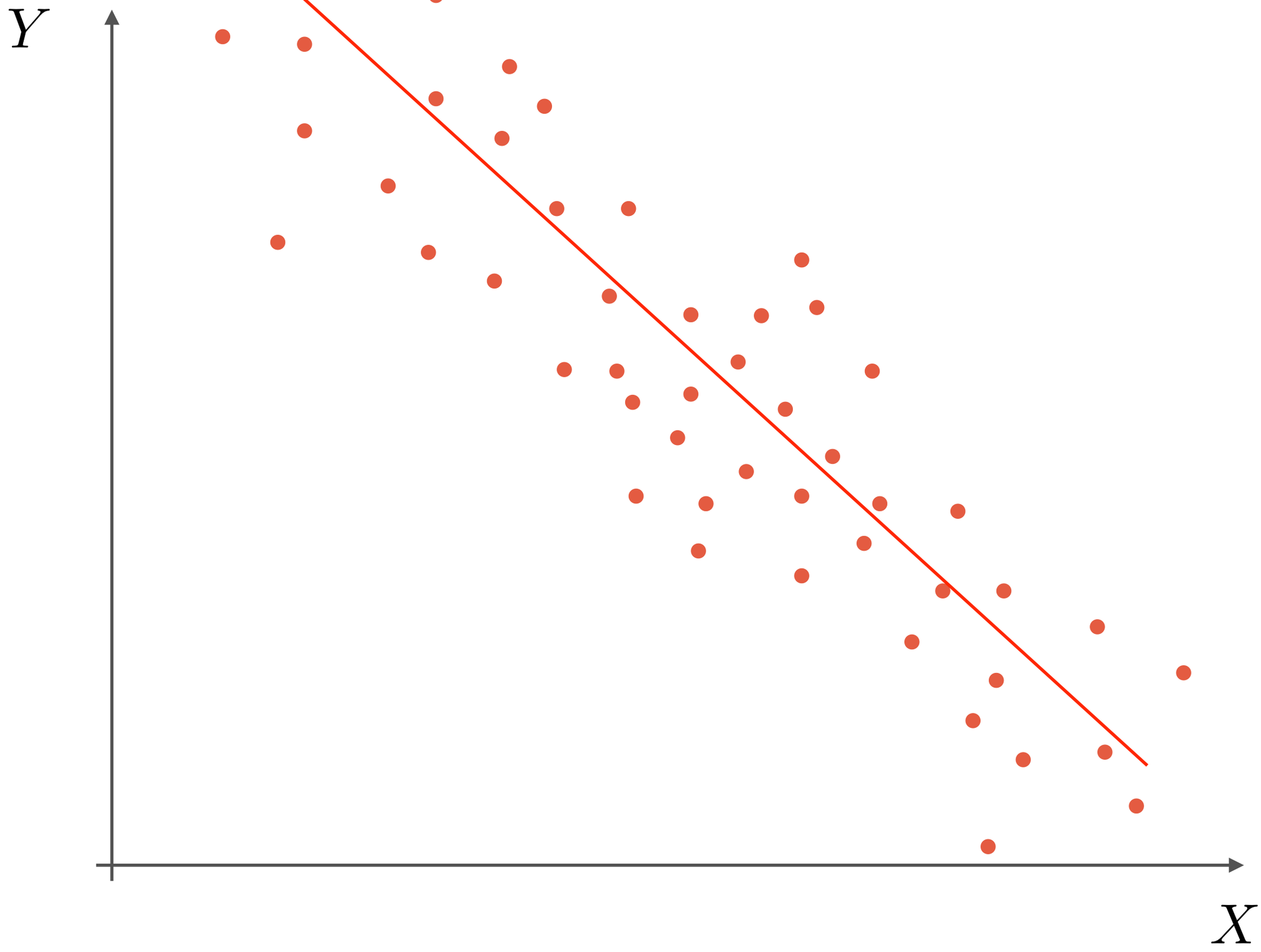
$$(x_i, y_i)$$

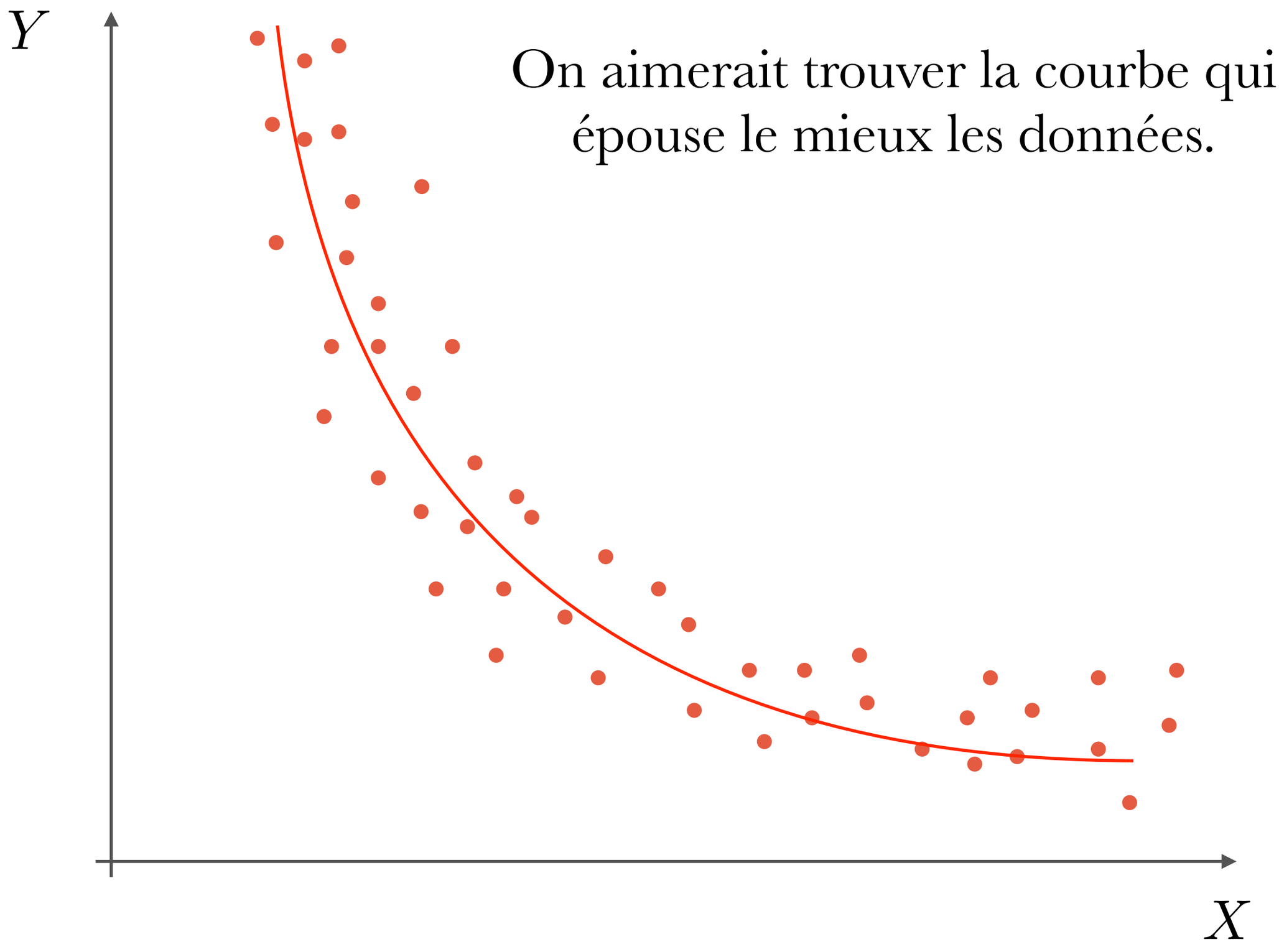
Associons à chacun de ces couples un point du plan.



S’il y a un lien entre les variables, on s’attend à voir un « pattern » entre les points


Y
X

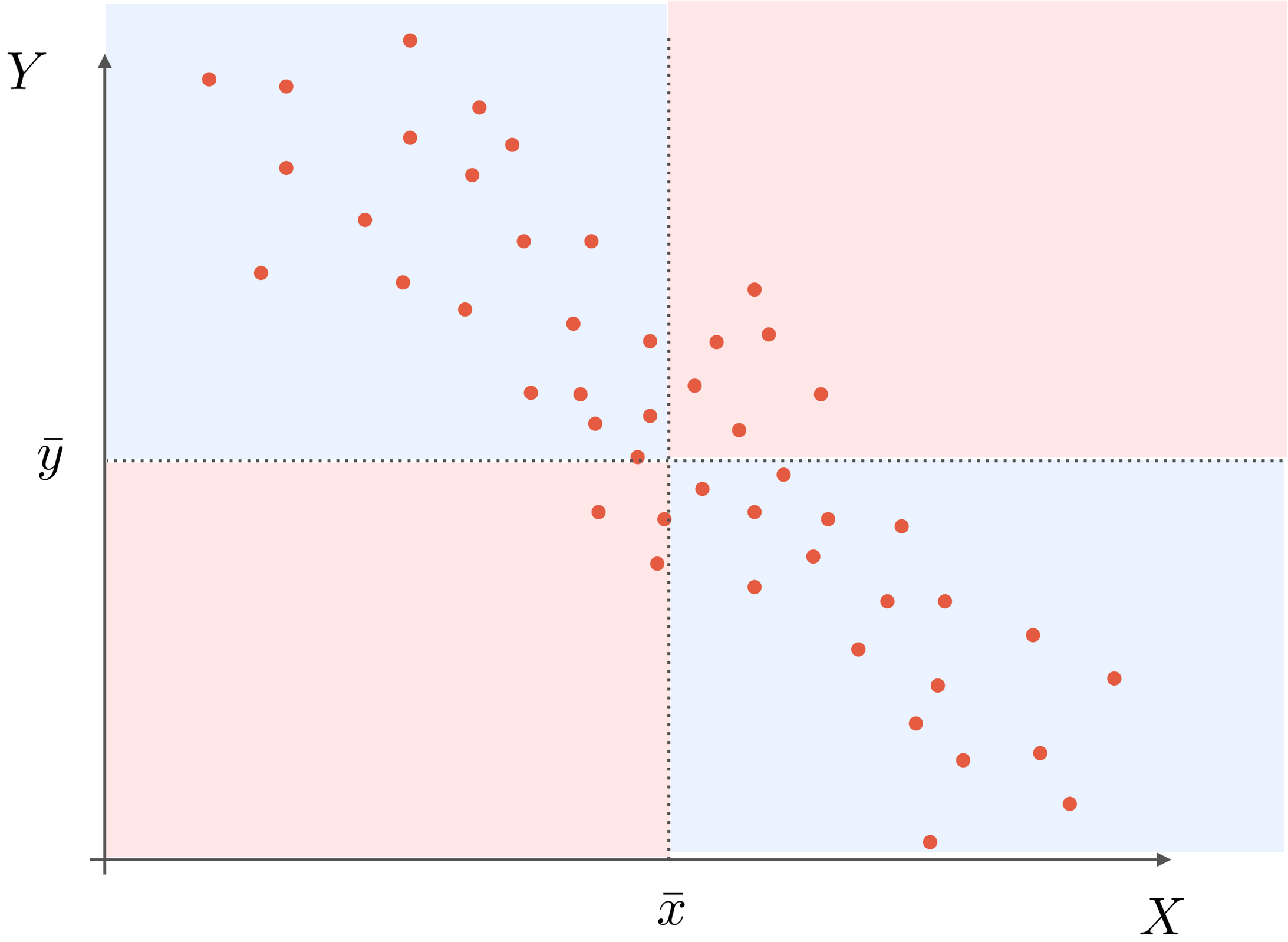

Y
X



On aimerait trouver la courbe qui épouse le mieux les données.

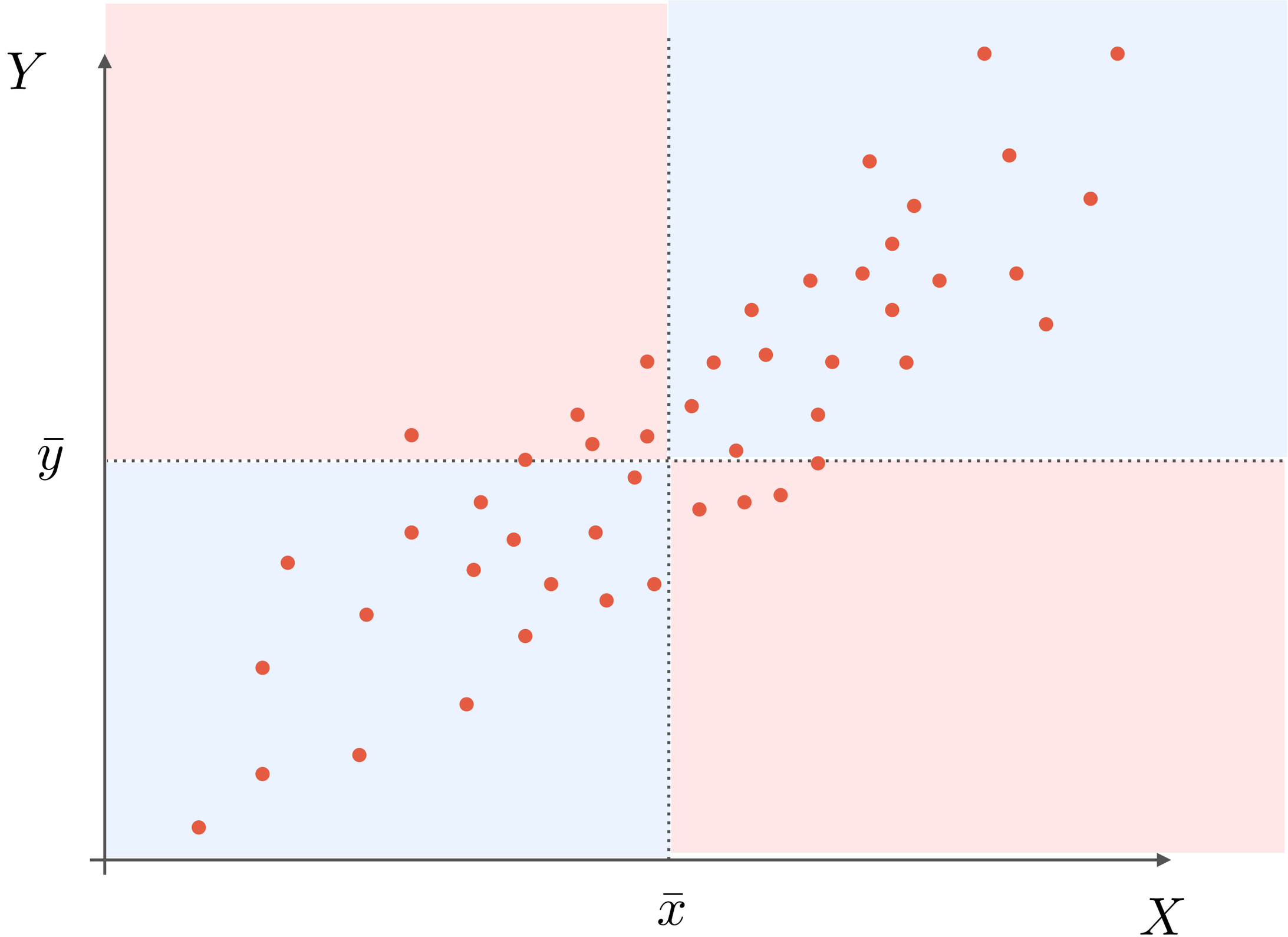
Mais on va se contenter de trouver la droite qui épouse le mieux les données.

S'il y a un lien linéaire, on s'attend à avoir beaucoup de points ici

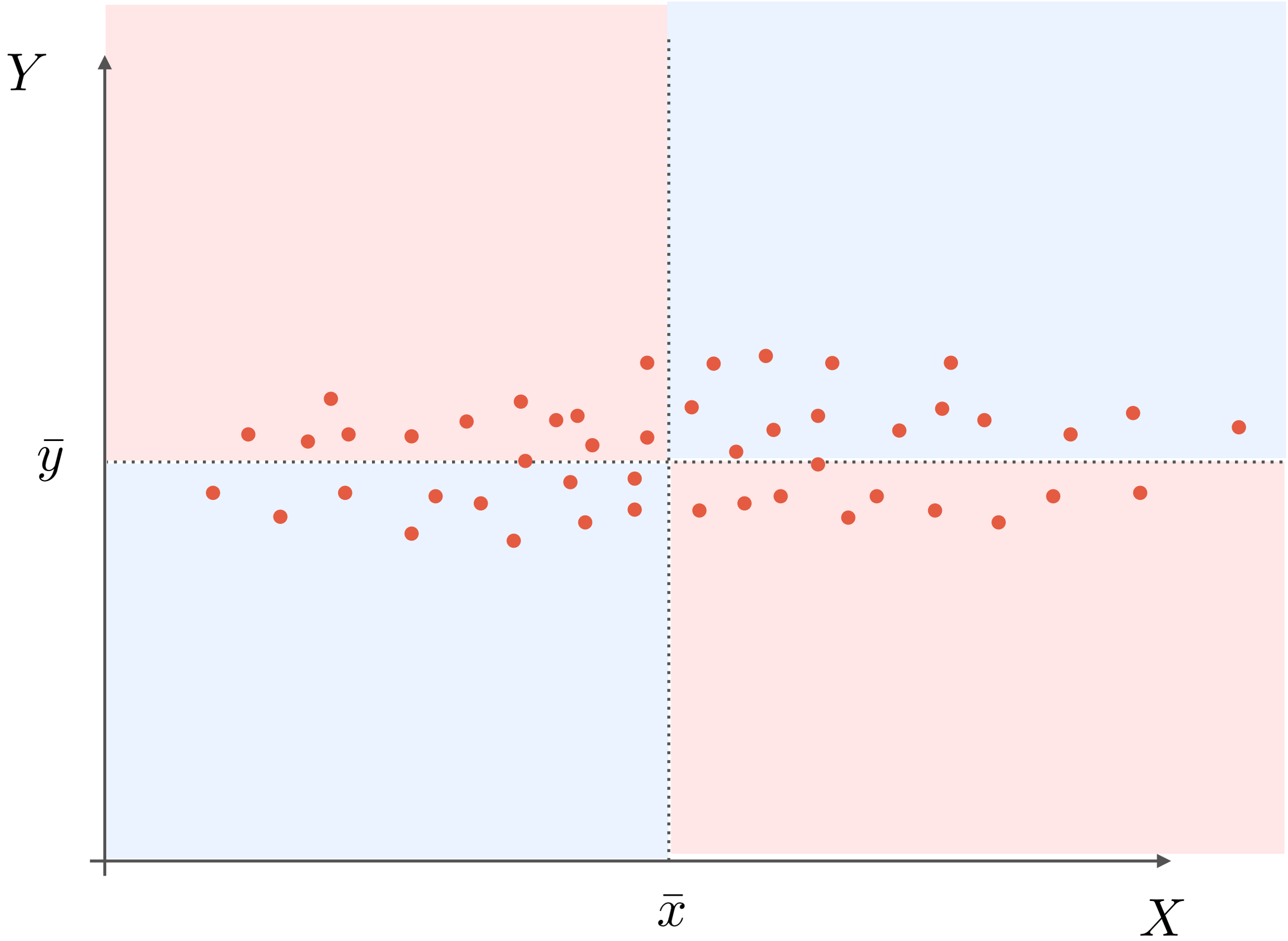


et très peu ici

Ou vice versa



Or on a un problème si X a une grande variance et Y une petite



C'est pour cette raison qu'on va travailler avec les variables centrées réduites.

$$Z_X = \frac{X - \bar{x}}{s_x} \qquad Z_Y = \frac{Y - \bar{y}}{s_y}$$

en divisant par l'écart type échantillonnal, car on travaille habituellement avec des échantillons.

$z_{x_i} < 0$

$z_{y_i} > 0$

$z_{x_i} > 0$

$z_{y_i} > 0$

$z_{x_i} z_{y_i} < 0$

$z_{x_i} z_{y_i} > 0$

$z_{x_i} z_{y_i} > 0$

$z_{x_i} z_{y_i} < 0$

$z_{x_i} < 0$

$z_{y_i} < 0$

$z_{x_i} > 0$

$z_{y_i} < 0$

On défini le **coefficient de corrélation**

$$r = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$$

on utilise n-1 car on utilise

s

et non

σ

Plus r est grand positivement plus les points se retrouvent dans la région bleue

Plus r est grand négativement plus les points se retrouvent dans la région rose

Plus r est près de 0, plus les points sont autant dans le bleu que dans le rouge.

Essayons de trouver une manière plus conviviale de trouver r

$$
\begin{aligned}
r &= \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1} \\
&= \frac{\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n-1} \\
&= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1) s_x s_y} \\
&= \frac{\sum_{i=1}^{n} (x_i y_i - y_i \bar{x} - x_i \bar{y} + \bar{x}\bar{y})}{(n-1) s_x s_y} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \bar{x} - \sum_{i=1}^{n} x_i \bar{y} + \sum_{i=1}^{n} \bar{x}\bar{y}}{(n-1) s_x s_y}
\end{aligned}
$$

$$
\begin{aligned}
r &= \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \bar{x} - \sum_{i=1}^{n} x_i \bar{y} + \sum_{i=1}^{n} \bar{x}\bar{y}}{(n-1)s_x s_y} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + \bar{x}\bar{y}\sum_{i=1}^{n} 1}{(n-1)s_x s_y} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - \bar{x}n\bar{y} - \bar{y}n\bar{x} + \bar{x}\bar{y}n}{(n-1)s_x s_y} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}
\end{aligned}
$$

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

Coefficient de corrélation

Exemple On prend 10 poissons et on mesure leurs longueurs et leurs diamètres.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
L	172	156	170	200	171	171	201	170	186	188
D	1,16	1,1	0,69	1,45	1,04	1,18	1,14	1,1	1,07	0,76

$\bar{x} = 178,5$ $\quad s_{\bar{x}} = 14,608$ $\quad \sum_{i=1}^{10} x_i y_i = 1916,08$

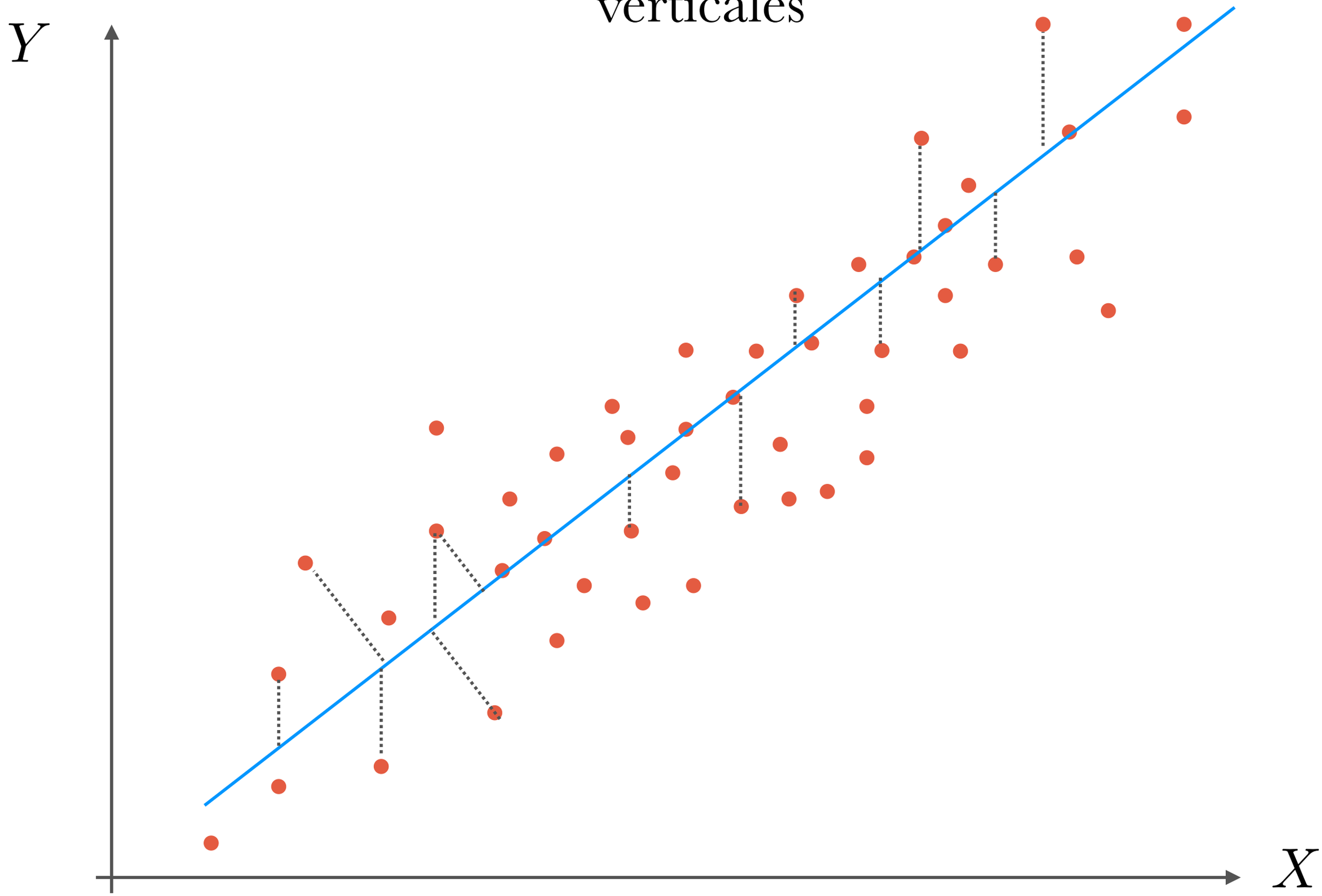
$\bar{y} = 1,069$ $\quad s_{\bar{y}} = 0,214$

$$r = \frac{1916,08 - (10)(178,5)(1,069)}{(9)(14,608)(0,214)} = 0,2813$$

Essayons de trouver une bonne droite qui épouse bien les données

Idéalement on aimerait minimiser les distances à la droite

Mais c'est algébriquement plus simple de minimiser les distances verticales



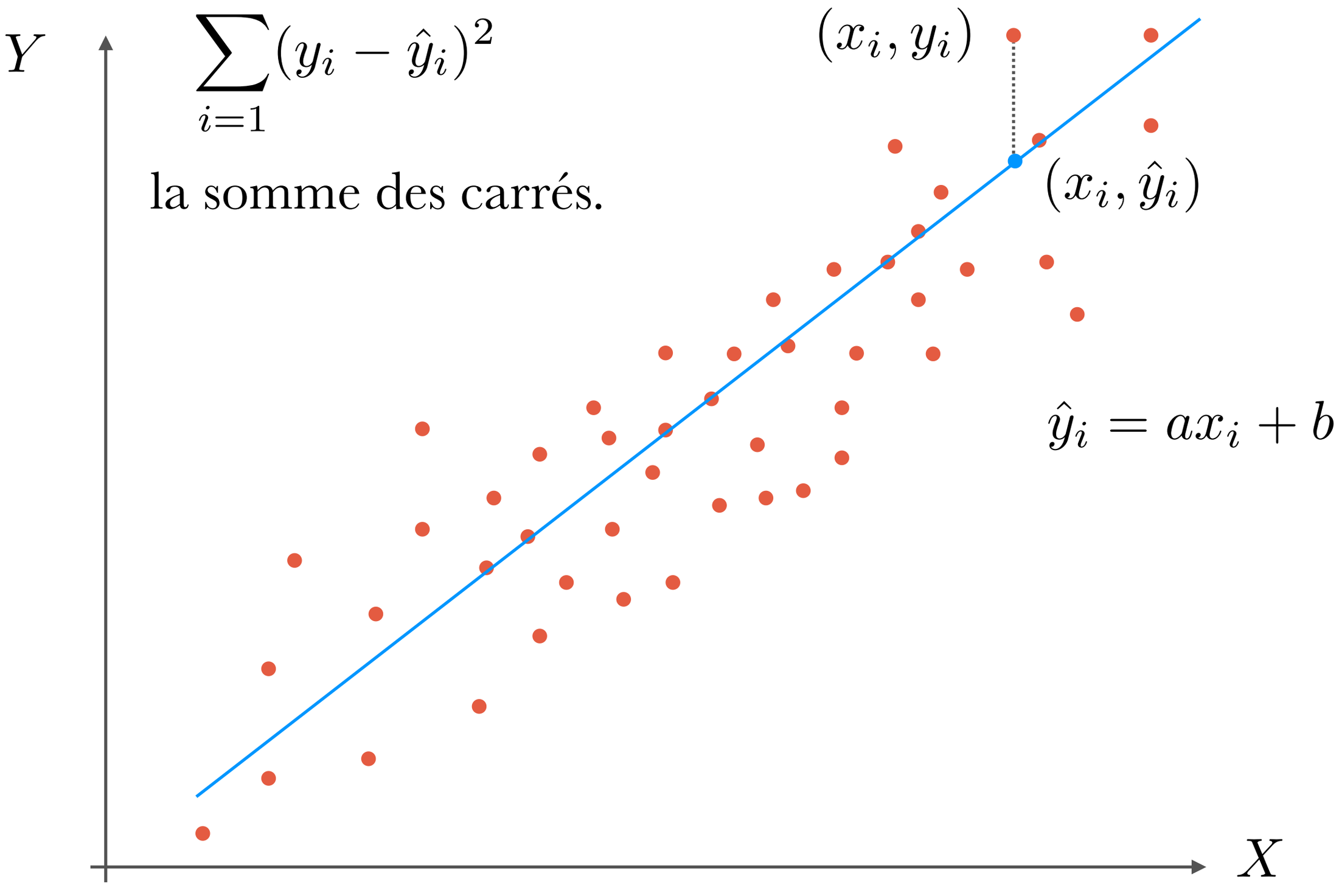
On veut donc que les $|y_i - \hat{y}_i|$ soient le plus petit possible

À la place, on va minimiser

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

la somme des carrés.

$$y = ax + b$$



$$y = ax + b \qquad \hat{y}_i = ax_i + b$$

À la place, on va minimiser $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ la somme des carrés.

$$f(a,b) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (ax_i + b))^2$$

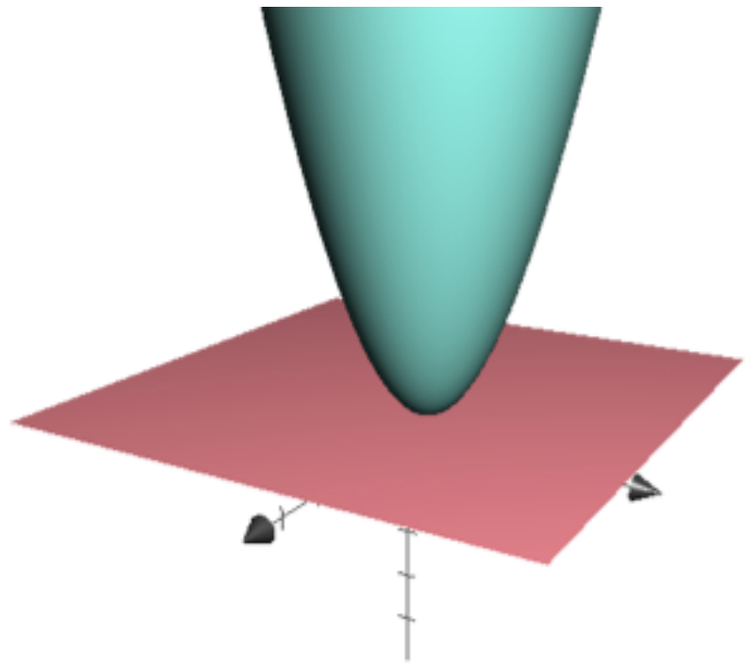
On peut voir cette expression comme une fonction qui dépend de *a* et de *b*

Et on cherche son minimum.

Avez vous déjà vu ça trouver des minimums?

Dérivée!!!

$f(a,b)$ est une fonction à deux variables donc ce n'est pas une courbe mais une surface.



On va donc chercher les points critiques

$$\frac{\partial f(a,b)}{\partial a} = 0 \qquad \text{et} \qquad \frac{\partial f(a,b)}{\partial b} = 0$$

On pourra conclure que ce point critique est automatiquement le minimum.

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 = \sum_{i=1}^{n}(y_i - ax_i - b)^2$$

$$= \sum_{i=1}^{n}(y_i^2 + a^2x_i^2 + b^2 - 2ay_ix_i - 2by_i + 2abx_i)$$

$$= \sum_{i=1}^{n} y_i^2 + a^2\sum_{i=1}^{n} x_i^2 + b^2\sum_{i=1}^{n} 1 - 2a\sum_{i=1}^{n} y_ix_i - 2b\sum_{i=1}^{n} y_i + 2ab\sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} y_i^2 + a^2\sum_{i=1}^{n} x_i^2 + b^2n - 2a\sum_{i=1}^{n} y_ix_i - 2bn\bar{y} + 2abn\bar{x}$$

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 \;\; = f(a,b)$$

$$= \sum_{i=1}^{n} y_i^2 + a^2 \sum_{i=1}^{n} x_i^2 + b^2 n - 2a \sum_{i=1}^{n} y_i x_i - 2bn\bar{y} + 2abn\bar{x}$$

$$\frac{\partial f(a,b)}{\partial b} = 2bn - 2n\bar{y} + 2an\bar{x} = 0$$

$$2bn = 2n\bar{y} - 2an\bar{x}$$

$$b = \bar{y} - a\bar{x}$$

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 \quad = f(a,b)$$

$$b = \bar{y} - a\bar{x}$$

$$= \sum_{i=1}^{n} y_i^2 + a^2 \sum_{i=1}^{n} x_i^2 + b^2 n - 2a \sum_{i=1}^{n} y_i x_i - 2bn\bar{y} + 2abn\bar{x}$$

$$\frac{\partial f(a,b)}{\partial a} = 2a \sum_{i=1}^{n} x_i^2 - 2 \sum_{i=1}^{n} y_i x_i + 2bn\bar{x}$$

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 \quad = f(a,b)$$

$$b = \bar{y} - a\bar{x}$$

$$\frac{\partial f(a,b)}{\partial a} = 2a\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} y_i x_i + 2bn\bar{x}$$

$$2a\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} y_i x_i + 2(\bar{y} - a\bar{x})n\bar{x} = 0$$

$$a\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i + (\bar{y} - a\bar{x})n\bar{x} = 0$$

1

$$a\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}y_ix_i+(\bar{y}-a\bar{x})n\bar{x}=0 \qquad b=\bar{y}-a\bar{x}$$

$$a\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)=a\sum_{i=1}^{n}x_i^2-an\bar{x}^2=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}$$

$$a=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}=\frac{n\sum_{i=1}^{n}x_iy_i-n^2\bar{x}\bar{y}}{n\sum_{i=1}^{n}x_i^2-n^2\bar{x}^2}$$

$$=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}$$

1

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \bar{y} - a\bar{x}$$

$$b = \bar{y} - \left(\frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\right)\bar{x}$$

$$= \bar{y} - \left(\frac{n\bar{x}\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\right)$$

1

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \bar{y} - \left(\frac{n\bar{x}\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \right)$$

$$= \frac{n\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{y}\left(\sum_{i=1}^{n} x_i\right)^2 - \left(n\bar{x}\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

1

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{n\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{y}\left(\sum_{i=1}^{n} x_i\right)^2 - \left(n\bar{x}\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$= \frac{n\bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}\sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

1

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{n\bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}\sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$= \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

1

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Donc la droite de régression est

$$y = ax + b$$

2

$$\sum_{i=1}^{n}(y_i-(ax_i+b))^2 \;=f(a,b)$$

$$b=\bar{y}-a\bar{x}$$

$$a\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}y_ix_i+(\bar{y}-a\bar{x})n\bar{x}=0$$

$$s_x^2=\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}$$

$$(n-1)s_x^2=\sum_{i=1}^{n}(x_i-\bar{x})^2 \;=\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)$$

$$=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+\bar{x}^2\sum_{i=1}^{n}1$$

$$=\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2 \;=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

2

$$\sum_{i=1}^{n}(y_i-(ax_i+b))^2 \;=f(a,b)$$

$$b=\bar{y}-a\bar{x}$$

$$a\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}y_ix_i+(\bar{y}-a\bar{x})n\bar{x}=0$$

$$a\left((n-1)s_x^2+n\bar{x}^2\right)-\sum_{i=1}^{n}y_ix_i+(\bar{y}-a\bar{x})n\bar{x}=0$$

$$(n-1)s_x^2=\sum_{i=1}^{n}(x_i-\bar{x})^2 \;=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

$$\sum_{i=1}^{n}x_i^2=(n-1)s_x^2+n\bar{x}^2$$

2

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 \;\; = f(a,b)$$

$$b = \bar{y} - a\bar{x}$$

$$a\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i + (\bar{y} - a\bar{x})n\bar{x} = 0$$

$$a\Big((n-1)s_x^2 + n\bar{x}^2\Big) - \sum_{i=1}^{n} y_i x_i + (\bar{y} - a\bar{x})n\bar{x} = 0$$

$$n$$

$$a(n-1)s_x^2 - \sum_{i=1} y_i x_i + n\bar{y}\bar{x} = 0$$

2

$$\sum_{i=1}^{n}(y_i-(ax_i+b))^2 = f(a,b)$$

$$b=\bar{y}-a\bar{x}$$

$$a(n-1)s_x^2-\sum_{i=1}^{n}y_ix_i+n\bar{y}\bar{x}=0$$

$$a(n-1)s_x^2-\left(r(n-1)s_xs_y+n\bar{x}\bar{y}\right)+n\bar{y}\bar{x}=0$$

$$r=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{(n-1)s_xs_y}$$

$$r(n-1)s_xs_y=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}$$

$$\sum_{i=1}^{n}x_iy_i=r(n-1)s_xs_y+n\bar{x}\bar{y}$$

2

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 = f(a,b)$$

$$b = \bar{y} - a\bar{x}$$

$$a(n-1)s_x^2 - \sum_{i=1}^{n} y_i x_i + n\bar{y}\bar{x} = 0$$

$$a(n-1)s_x^2 - \left(r(n-1)s_x s_y + n\bar{x}\bar{y}\right) + n\bar{y}\bar{x} = 0$$

$$a(n-1)s_x^2 - r(n-1)s_x s_y = 0$$

$$a(n-1)s_x^2 = r(n-1)s_x s_y$$

$$a = \frac{r(n-1)s_x s_y}{(n-1)s_x^2}$$

$$a = \frac{rs_y}{s_x}$$

2

$$\sum_{i=1}^{n}(y_i - (ax_i + b))^2 \quad = f(a,b)$$

$$a = \frac{r s_y}{s_x} \qquad b = \bar{y} - a\bar{x} = \bar{y} - \frac{r s_y}{s_x}\bar{x}$$

Donc la droite de régression est

$$y = ax + b$$

$$y = \left(\frac{r s_y}{s_x}\right) x + \left(\bar{y} - \frac{r s_y}{s_x}\bar{x}\right)$$

2

Donc la droite de régression est

$$f(x) = y = \left(\frac{rs_y}{s_x}\right)x + \left(\bar{y} - \frac{rs_y}{s_x}\bar{x}\right)$$

En posant $x = \bar{x}$

$$f(\bar{x}) = \left(\frac{rs_y}{s_x}\right)\bar{x} + \left(\bar{y} - \frac{rs_y}{s_x}\bar{x}\right) = \bar{y}$$

Donc la droite de régression passe par le point

$$(\bar{x}, \bar{y})$$

3

Pour le fun!

$$y_i = ax_i + b$$

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\mathbf{XD} = \mathbf{Y}$$

$$\mathbf{X}^T\mathbf{XD} = \mathbf{X}^T\mathbf{Y}$$

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

3

Pour le fun! $y_i = ax_i + b$

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \qquad \mathbf{X}^T\mathbf{Y} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

$$\mathbf{XD} = \mathbf{Y} \qquad \mathbf{X}^T\mathbf{XD} = \mathbf{X}^T\mathbf{Y}$$

$$\mathbf{D} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

3

Pour le fun! $y_i = ax_i + b$

$$\mathbf{D} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

$$= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} n\sum x_i y_i - \sum x_i \sum y_i \\ \\ \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \end{pmatrix}$$

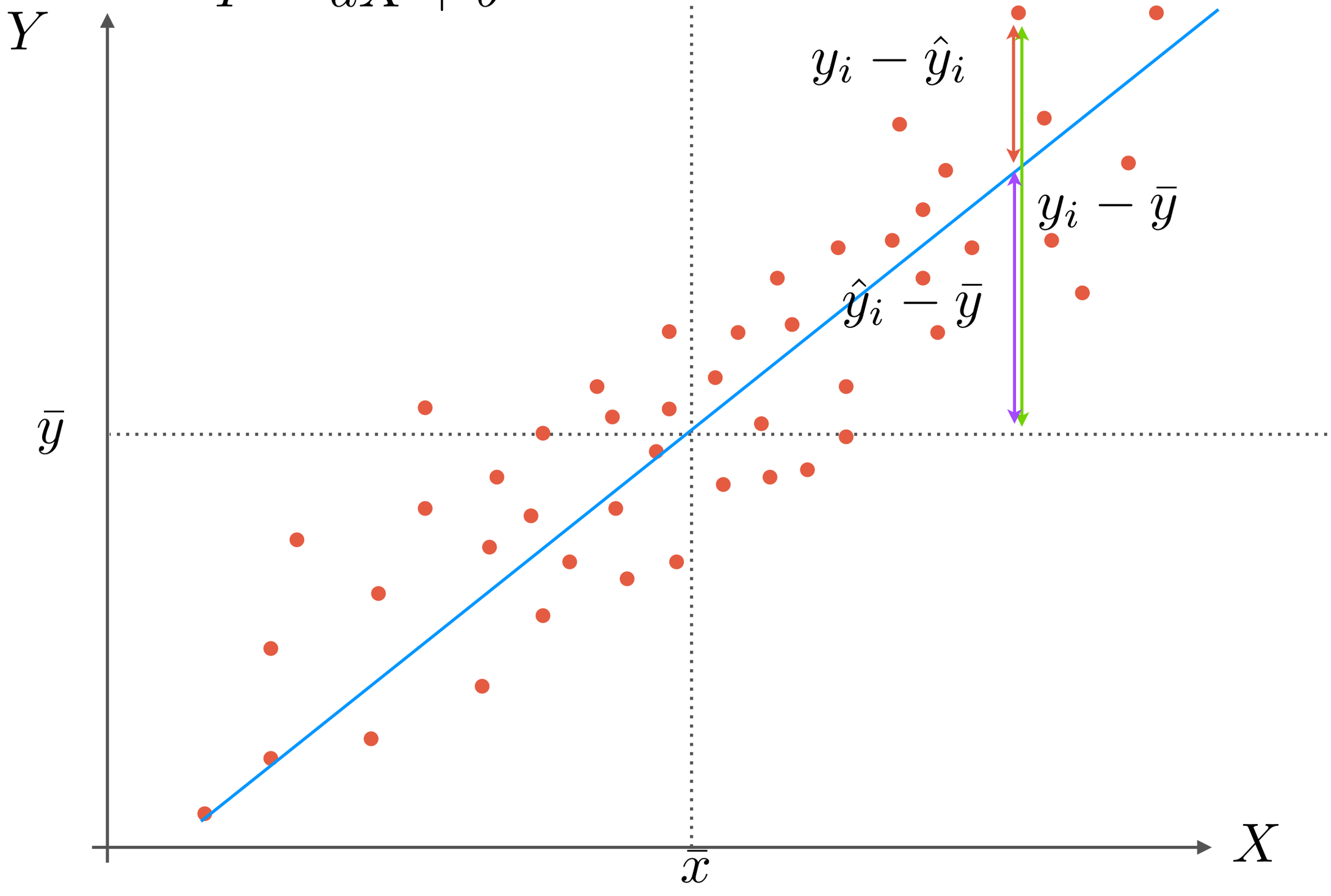
$$= \begin{pmatrix} \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \\ \\ \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

Exactement ce qu'on a eu plus tôt!!!

Lorsqu'on a une variable statistique Y on s'attend à s'écarter de la moyenne

Or une partie de cet écart est dû au lien entre les variables

$$Y = aX + b$$



On introduit donc le **coefficient de détermination** qui permet de mesurer quelle partie de la variation est expliquée par le lien entre les variables.

Variations expliquées: $\hat{y}_i - \bar{y}$

Variations totales: $y_i - \bar{y}$

$$\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}((ax_i + b) - (a\bar{x} + b))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(ax_i - a\bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(ax_i - a\bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{a^2\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \frac{a^2(n-1)s_x^2}{(n-1)s_y^2} = a^2\frac{s_x^2}{s_y^2}$$

$$= \left(\frac{rs_y}{s_x}\right)^2 \frac{s_x^2}{s_y^2} = r^2$$

Donc le coefficient de détermination est tous simplement

$$r^2 \qquad \text{où} \qquad r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

Devoir: 5.12 et 5.13