Faculté de médecine
1ère année
Cours de Biostatistique

Mme Medouer Nawel

Cour 4

Statistique bivariée

Statistique bivariée

Lorsqu'on effectue sur un échantillon simultanément, deux mesures, on

obtient une série double de mesures ; par exemple; on parlera de la série double de mesures sur la taille et le poids des étudiants .. Ces mesures

sont en fait des réalisations de deux variables quantitatives et la question qui se pose est celle de l'existence d'une liaison entre ces deux

variables.

Statistique bivariée

L'ajustement linéaire consiste à tracer une droite qui passe

au plus près des observations d'un nuage de points.

Cette droite est ensuite utilisée pour faire des prévisions.

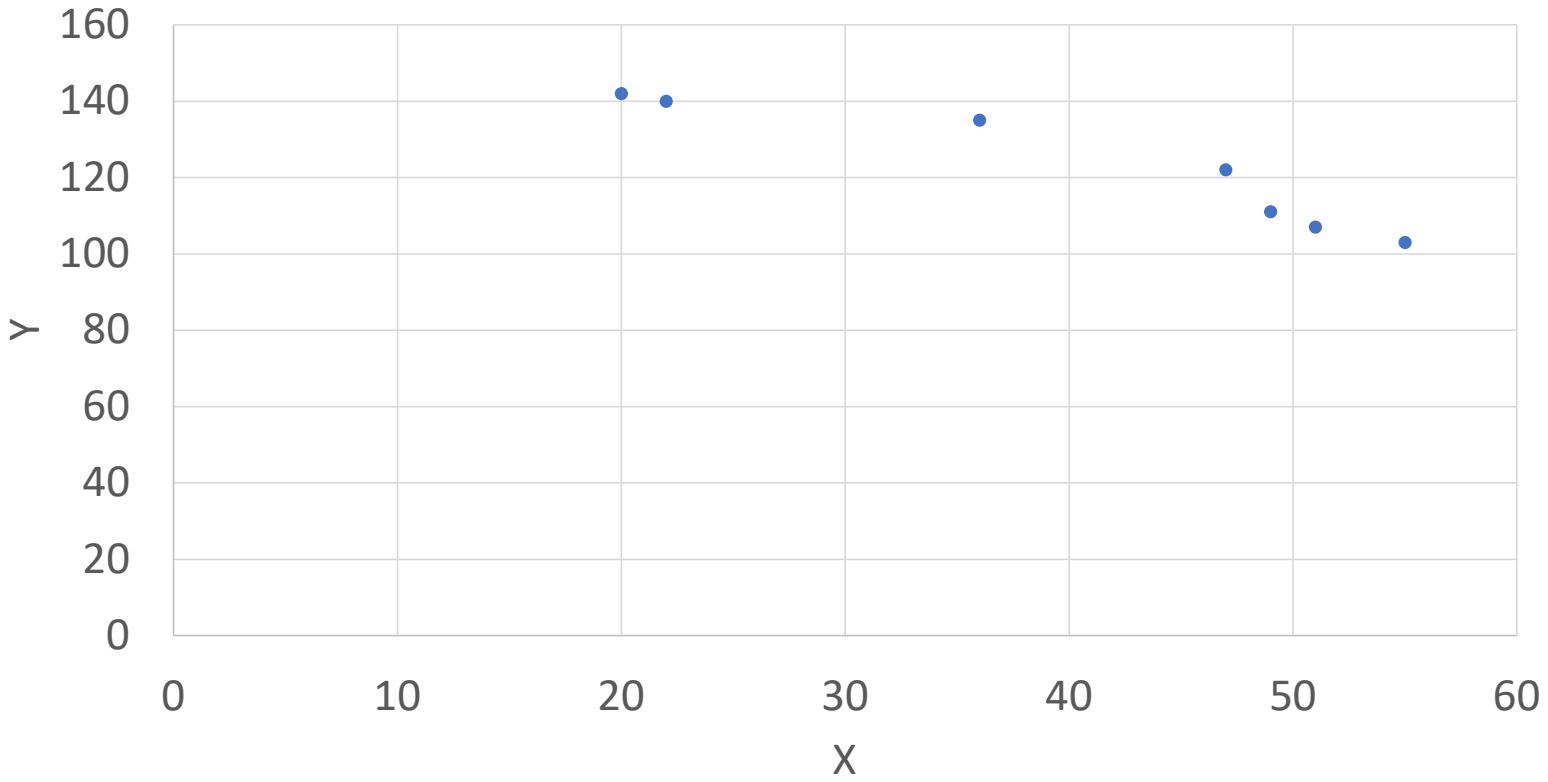
Exemple 1 introductif

X	55	51	49	47	36	22	20
Y	103	107	111	122	135	140	142

Nuage de points *Figure 2.1*

Des points dispersés entre les deux axes.

l'axe des abscisse et l'axe des ordonnés



Interprétation de nuage des points

Les deux variable se varient dans deux sens opposés,

quand X augmente Y déminue

Qu'est-ce la régression linéaire ?

Lorsque le nuage de points associé à une série statistique double a une forme "allongée " c'est-à-dire lorsque les points sont sensiblement alignés, on peut tracer des droites passant « au plus près de ces points ». On dit alors que chacune de ces droites réalise un **ajustement affine** du nuage de points.

X	Y
Variable explicative	Variable expliquée
Variable contrôlée	Variable réponse
Variable indépendante	Variable dépendante

On distingue trois cas

Cas 1: Tableau de deux lignes

X=x_i	x_1	x_2	.....	x_{k-1}	x_N
Y=y_i	y_1	y_2		y_{k-1}	y_N

On distingue trois cas

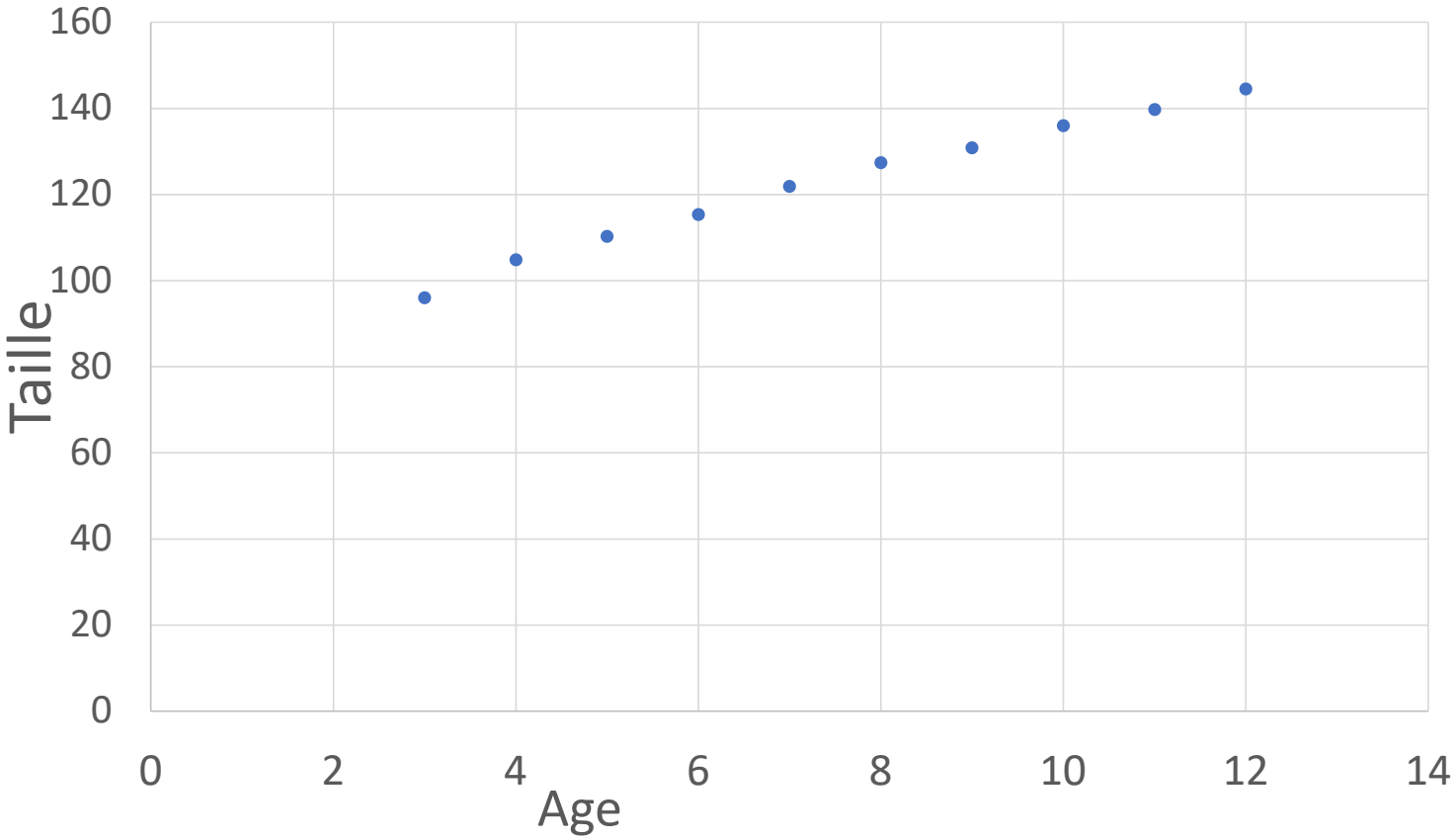
Cas 1: Tableau de deux lignes

Exemple 2:

Age	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
taille	*96*	*104,8*	*110,3*	*115,3*	*121,9*	*127,4*	*130*	*136,9*	*139,7*	*144,5*

Interprétation de nuage des points

Les deux variable évoluent dans le même sens
quand X augmente Y augmente



Ajustement graphique

Vous choisissez deux couples (5;110,3) et (10;136,9)

$$\begin{cases} 110{,}3 = a5 + b \ldots\ldots (1) \\ 136{,}9 = a10 + b \ldots\ldots (2) \end{cases}$$

(2)-(1) $5a=26{,}6$ $a = \frac{26{,}6}{5} = 5{,}32$

De (2) $b = 136{,}9-10(5{,}32) = 83{,}7$

Alors la droite d'ajustement graphique est de

$$y = 5{,}32x + 83{,}7$$

Ajustement graphique

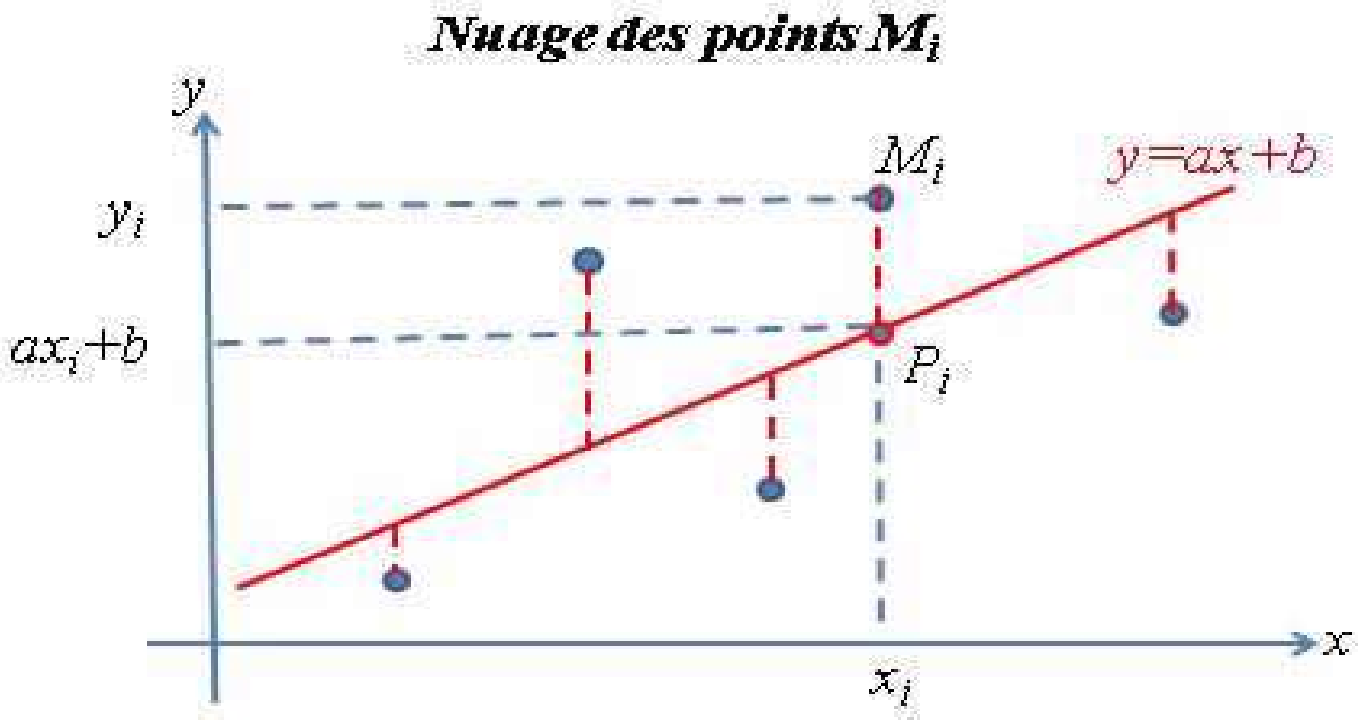
Les inconvénients *de l'ajustement graphique, c'est que ce dernier est une méthode subjective , chaque étudiant trouve une droite et finalement vous n'avez pas des critères objectifs pour choisir entre ces droites .*

Problématique ?

Comment faire pour avoir une droite unique?

Droite de La méthode des moindres carrés (MOC)

Cette méthode consiste à minimiser la somme des carrés des écarts



Démonstration

Minimiser les écarts, soit la fonction

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

Après une chaine d'opérations mathématiques, on obtient:

Méthode des moindres carrées

$$\begin{cases} \mathrm{a} = \dfrac{\sum_{i=1}^{N} x_i y_i \; - \bar{Y} \sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2 - \bar{X} \sum_{i=1}^{N} x_i} \\ \mathrm{b} = \bar{\mathrm{Y}} - \mathrm{a}\bar{X} \end{cases}$$

- *La 2ème équation signifie que la droite de régression passe la*

point moyen $G(\bar{X}; \bar{Y})$

Droite de la méthode des moindres carrés

Droite de Y en fonction de X: ***Notée par***

$$\boldsymbol{D_Y(x): y = ax + b}$$

Tels que:

$$\begin{cases} \mathrm{a} = \dfrac{\sum_{i=1}^{N} x_i y_i - \bar{Y} \sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2 - \bar{X} \sum_{i=1}^{N} x_i} \\ \mathrm{b} = \overline{\mathrm{Y}} - \mathrm{a}\overline{\mathrm{X}} \end{cases}$$

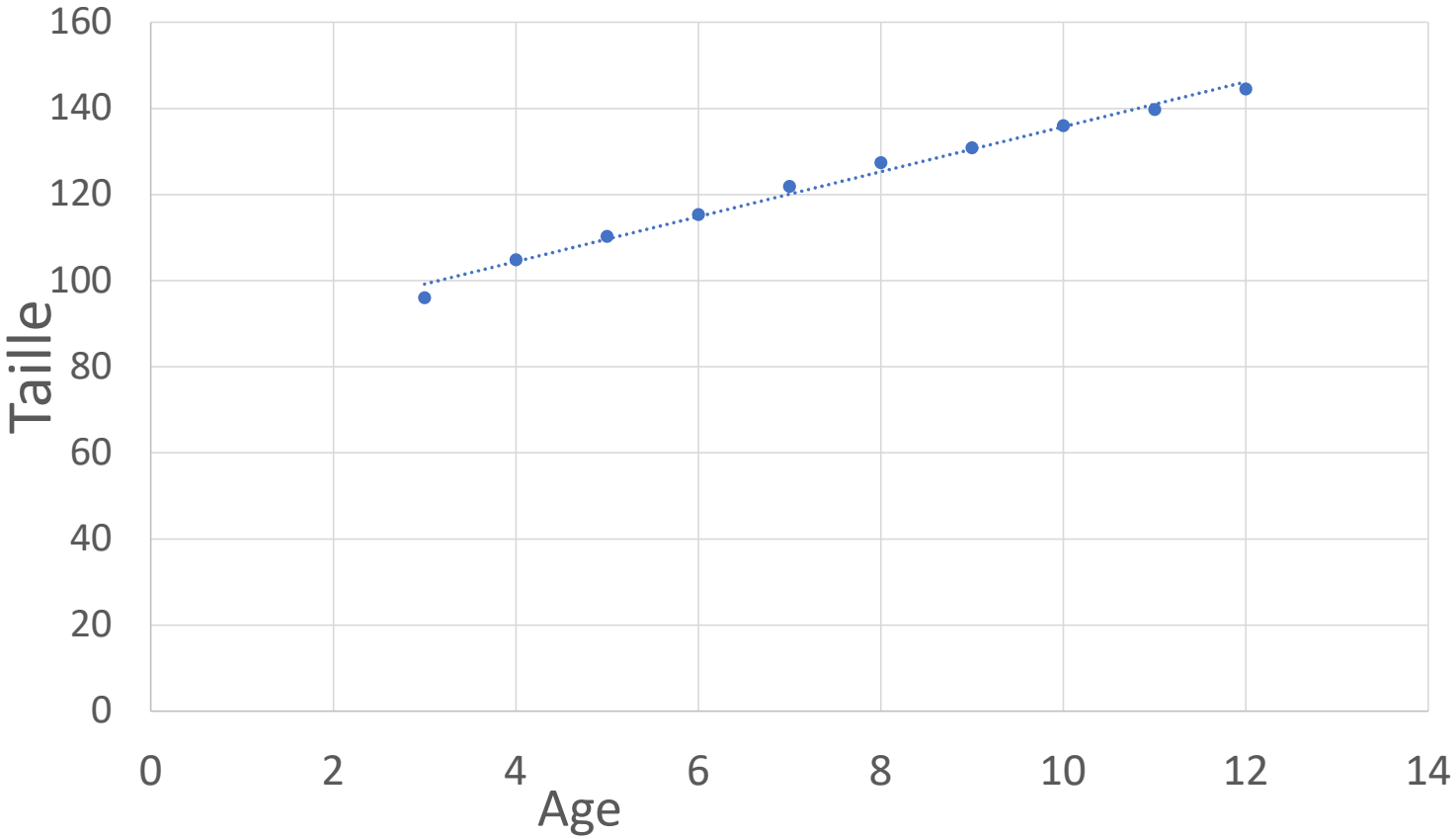
Propriétés de la droite de régression $\boldsymbol{D_Y(x)}$

1. *C'est une droite unique*

2. *Elle passe toujours par le points moyen* $G(\bar{X};\bar{Y})$

Interprétation de nuage des points

Les deux variable se varient dans le même sens quand X augmente Y augmente



Notations:

1) La covariance $Cov(X;Y) = \sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{X}\bar{Y}$

2) La variance de X $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{X}^2$

3) La variance de Y $\sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - Y^2$

Alors $a = \frac{\sigma_{XY}}{\sigma_X^2}$

Coefficient de corrélation linéaire

Symbolisé par $r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Le coefficient de corrélation est une valeur sans unité comprise toujours entre (-1) et (+1)

$$-1 \leq \quad r \leq +1$$

Coefficient de corrélation linéaire

- *Les valeurs positives de r indiquent une corrélation positive*

Lorsque les valeurs des deux variables tendent augmenter ensemble

- *Les valeurs négatives de r indiquent une corrélation négative lorsque les valeurs d'une variable tend à augmenter et que les valeurs de l'autre variable diminuent*

Signification de r

- $r = 0$ liaison nulle

- $r = +1$ forte liaison positive
- $r = -1$ forte liaison négative

On distingue trois cas

Cas 1: Tableau de deux lignes

Exemple 2:

Age	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
taille	*96*	*104,8*	*110,3*	*115,3*	*121,9*	*127,4*	*130*	*136,9*	*139,7*	*144,5*

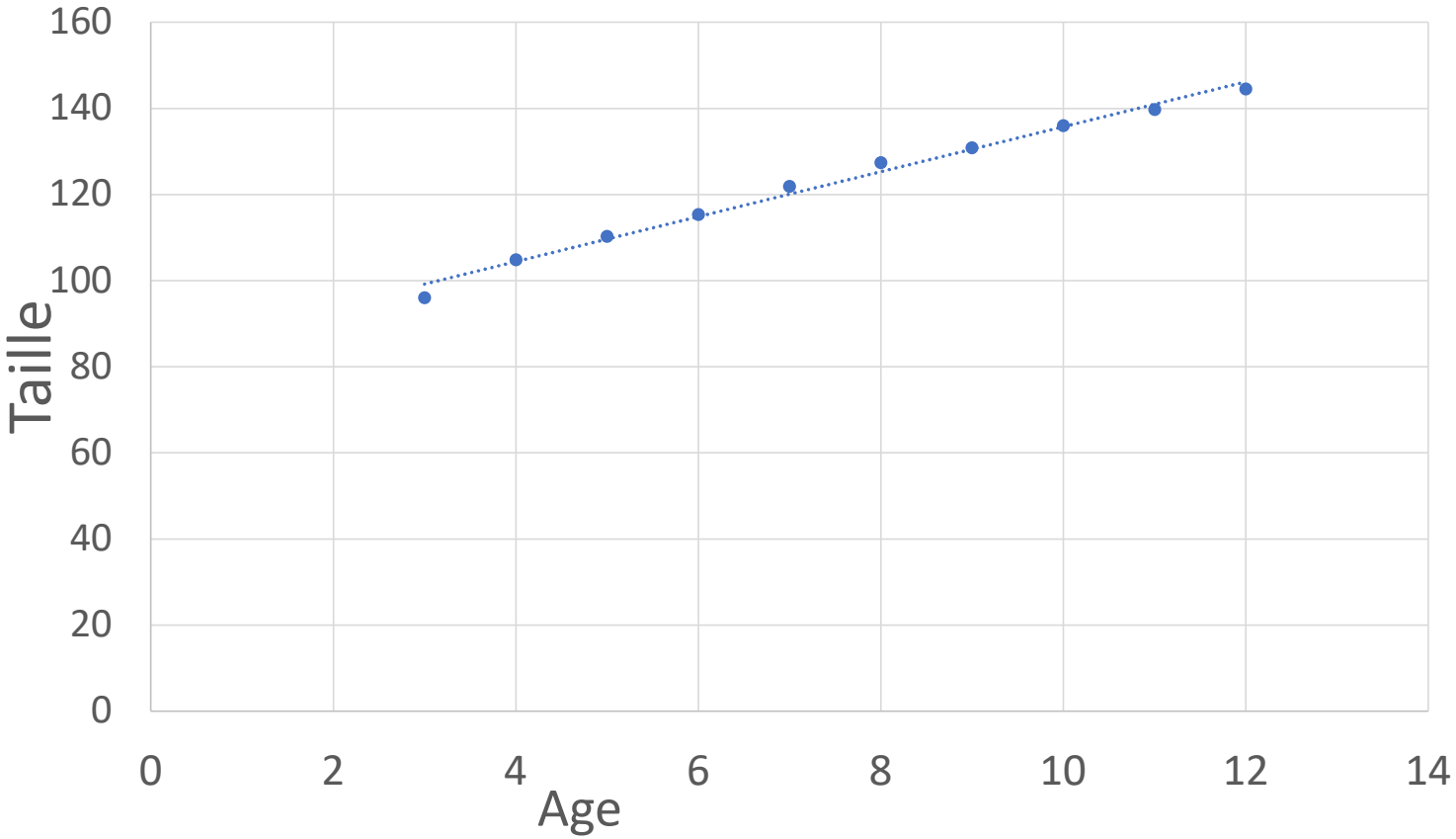
Droite de la méthode des moindres carrés

Droite de X en fonction de Y

$$\boldsymbol{D_Y(x): y = ax + b}$$

Interprétation de nuage des points

Les deux variable se varient dans le même sens quand X augmente Y augmente



$$\boldsymbol{D_Y(x): y = ax + b}$$

$$B \rightarrow a = 5{,}23$$
$$A \rightarrow b = 83{,}43$$
$$r = 0{,}9945$$

$$\boldsymbol{D_Y(x): y = 5{,}23x + 83{,}43}$$

2ème droite de la méthode des moindres carrés

***Droite de X en fonction de Y:* Notée par**

$$\boldsymbol{D_X(Y): x = \grave{a}y + \grave{b}}$$

Tels que:

$$\begin{cases} \boldsymbol{\grave{a}} = \dfrac{\sum_{i=1}^{N} x_i y_i \; - \bar{Y} \sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} y_i^2 - \bar{Y} \sum_{i=1}^{N} y_i} \\ \boldsymbol{\grave{b}} = \bar{X} - \grave{a}\bar{Y} \end{cases}$$

On reprend l'exemple précédant:

Droite de X en fonction de Y

$$\boldsymbol{D_X(Y): x = \grave{a}y + \grave{b}}$$

$$B \rightarrow \grave{a} = 0{,}189$$

$$A \rightarrow \grave{b} = -15{,}69$$

$$r = 0{,}9945$$

$$\boldsymbol{D_X(Y): x = 0{,}189y - 15{,}69}$$

Remarques

1. $a, à$ *et* r *ayant le même signe*

Les trois sont positives ou les trois sont négatives

2. $r^2 = aà$

Cas 2: Tableau de trois lignes

X=x_i	x_1	x_2	.....	x_{k-1}	x_k
Y=y_i	y_1	y_2	.....	y_{k-1}	y_k
n_i	n_1	n_2	.....	n_{k-1}	n_k

En adaptant les formules pour le cas 2

1) La covariance $Cov(X;Y) = \sigma_{XY} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i y_i \; - \bar{X}\bar{Y}$

2) La variance de X $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 \; - \bar{X}^2$

3) La variance de Y $\sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{k} n_i y_i^2 \; - \bar{Y}^2$

Alors $a = \frac{\sigma_{XY}}{\sigma_X^2}$

Cas 2: Tableau de trois lignes

Exemple

X=x_i	1	2	3	4	5
Y=y_i	10	15	17	3	18
n_i	1	3	5	7	4

$$\boldsymbol{D_Y(x): y = ax + b}$$

$$B \rightarrow a = -0{,}66$$
$$A \rightarrow b = 13{,}96$$
$$r = -0{,}1122$$

(Corrélation faible négative)

$$\boldsymbol{D_Y(x): y = -0{,}66 + 13{,}96}$$

Cas 3

Tableau de contingence

X=x_i / Y=y_i	y_1	$\cdots$	y_j	$\cdots$	y_l	*Totaux*
x_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1l}	$n_{1.}$
x_i	n_{i1}		n_{ij}		n_{il}	$n_{i.}$
x_k	n_{k1}		n_{kj}		n_{kl}	$n_{k.}$
Totaux	$n_{.1}$		$n_{.j}$		$n_{.l}$	$n_{..} = N$

En adaptant les formules pour le cas d'un tableau de contingence

1) La covariance $Cov(X;Y) = \sigma_{XY}$

$$= \frac{1}{N}\sum_{k}\sum_{l} n_{ij}x_i y_j \; - \bar{X}\bar{Y}$$

1) La variance de X $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 \; - \bar{X}^2$

2) La variance de Y $\sigma_Y^2 = \frac{1}{N}\sum_{j=1}^{l} n_j y_j^2 \; - \bar{Y}^2$

Alors $a = \frac{\sigma_{XY}}{\sigma_X^2}$

Y=y_i / X=x_i	1	4	5	Totaux
2	1	10	3	14
3	2	4	1	7
7	3	1	4	8
8	7	2	2	11
$Totaux$	13	17	10	40

Exemple: La saisie dans la calculatrice

k= 4 lignes et $l = 3$ colones, donc $k \times l = 4 \times 3 = 12\ valeurs$

X	Y	ni
2	1	1
2	4	10
2	5	3
3	1	2
3	4	4
3	5	1
7	1	3
7	4	1
7	5	4
8	1	7
8	4	2
8	5	2

$$\boldsymbol{D_Y(x): y = ax + b}$$

$$B \rightarrow a = -0{,}04539$$
$$A \rightarrow b = 3{,}3816$$
$$r = -0{,}77$$

(Corrélation moyenne négative)

$$\boldsymbol{D_Y(x): y = -0{,}04539x + 3{,}3816}$$