FORMATIONS AUX STATISTIQUES : MANUEL DU FORMATEUR DE NIVEAU 1

Janvier 1991

SCÉNARIO 6

Ajustement à une droite

INRA

DOCUMENTS "STAGIAIRE"

DÉPARTEMENT DE BIOMÉTRIE

formation permanente

Données concernant la croissance de critaux de glace. (C.S.I.R.O. : Commonwealth Scientific and Industrial Research Organisation) Australie. Journal of the Atmospheric Sciences, 33, 1976, 842-850.

Des cristaux de glace sont introduits dans une chambre froide, dont la température intérieure est maintenue à -5°c, et dont le taux d'humidité est maintenu constant. On observe la longueur des critaux (μm), en fonction du temps (entre 50 et 180 secondes après l'introduction des cristaux dans la chambre).

Chaque mesure représente une expérience complète : un cristal de glace est introduit dans une chambre. Les expériences sont faites sur plusieurs jours, elles sont randomisées dans le temps.

temps (secondes)	longueur du cristal (μm)
50	19
80	25
70	17
100	30
90	21
150	36
110	30
130	31
140	33
120	36

a et b sont estimés graphiquement

a =	
b =	

X	Y	Y ajusté	écarts Y-Y ajusté	$(\text{écarts})^2 = (\text{Y-Y ajusté})^2$
50	19			
80	25			
70	17			
100	30			
90	21			
150	36			
110	30			
130	31			
140	33			
120	36			
			Total :	

Equation : $Y = a + bX$

Y : variable expliquée a : ordonnée à l'origine
X : variable explicative b : pente

Calcul des paramètres a et b qui minimisent le critère des moindres carrés :

$$b = \frac{\Sigma_i \left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)}{\Sigma_i \left(X_i - \bar{X}\right)^2} = \frac{\Sigma_i X_i Y_i - n\bar{X}\bar{Y}}{\Sigma_i X_i^2 - n\bar{X}^2}]$$

$$a = \bar{Y} - b\bar{X}]$$

avec $\bar{Y}$ = moyenne des Y observés
$\bar{X}$ = moyenne des X observés
n = nombre d'observations.

	X	Y	X^2	Y^2	$X.Y$	Y ajusté $= a + bX$	écarts = $Y - Y$ ajusté	écarts^2 $= (Y - Y\text{ajustés})^2$
$\Sigma =$ Somme : Total	$\Sigma_i X_i$	$\Sigma_i Y_i$	$\Sigma_i X_i^2$		$\Sigma_i X_i Y_i$			Somme des carrés des écarts = critère
Total/n = moyenne	$\bar{X}$	$\bar{Y}$						

n =

Calcul des paramètres a et b qui minimisent le critère des moindres carrés :

	X	Y	X^2	Y^2	$X.Y$	Y ajusté $= a + bX$	écarts = $Y - Y$ ajusté	$(\text{écarts})^2$ $= (Y - Y\text{ajustés})^2$
	50	19	2500	361	950			
	80	25	6400	625	2000			
	70	17	4900	289	1190	21,22	-4,22	17.8
	100	30	10 000	900	3000	27,03	2,97	8,8
	90	21	8100	441	1890	25,09	-4,09	16,7
	150	36	22 500	1296	5400	36,7	-0,70	0,5
	110	30				28,96	1,04	1,1
	130	31	16 900	961	4030	32,83	-1,83	3,3
	140	33	19 600	1089	4620	34.77	-1,77	3,1
	120	36	14 400	1296	4320	30,90	5,10	26,0
Total	1040	278						
moyenne								

n =

b =

a =

critère des moindres carrés =

Le problème :

On teste la corrosion de 13 spécimens d'un alliage cuivre-nickel (90/10).
Chaque spécimen contient une quantité de fer connue.
On fait tourner la roue dans l'eau de mer, à 30 tours/seconde, pendant 60 jours. On mesure la corrosion. C'est la perte de poids en milligrammes par décimètre carré et par jour.
Les données récoltées sont les suivantes :

Pertes de Poids :
128, 124, 111, 104, 102, 130, 122, 92, 113, 84, 128, 91, 86.

Les quantités de fer correspondantes sont :
0.01, 0.48, 0.71, 0.95, 1.19, 0.01, 0.48, 1.44, 0.71, 1.96, 0.01, 1.44, 1.96.

Questions :
Quelle est la variable explicative?

Pour faire un graphique, quelle variable met-on en abscisse, quelle variable met-on en ordonnée?

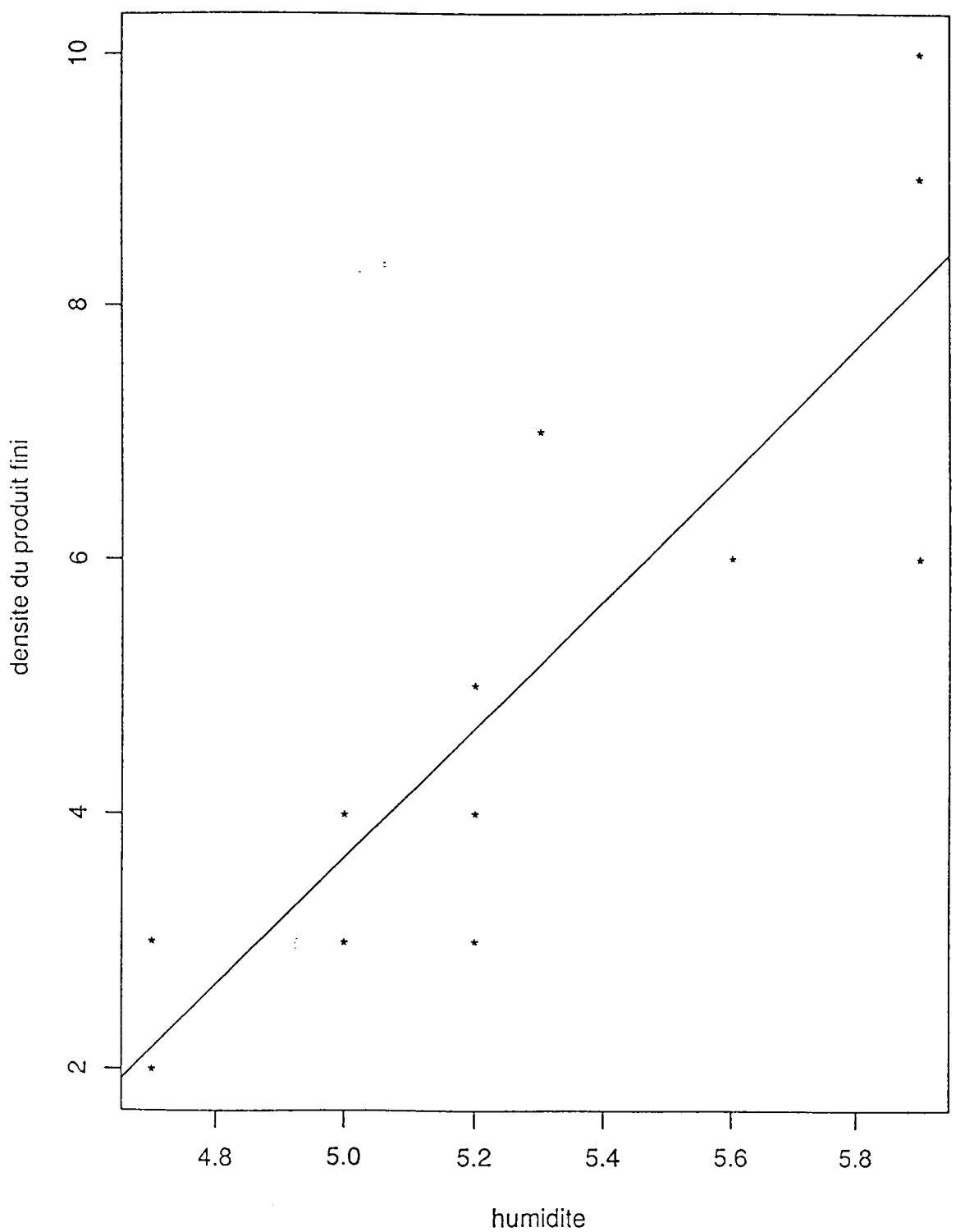
Sur cet exemple, on a calculé les paramètres qui minimisent la somme des carrés des écarts.
On trouve :

- pente = -24,1
- ordonnée à l'origine : 129,9

Donner l'équation de la droite.

Tracer la droite.

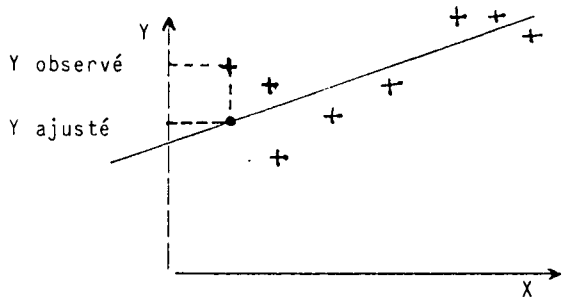
D'après ce graphique des observés et de la droite ajustée proposer une représentation graphique des résidus.



Scénario 6

VOCABULAIRE

Ajustement ou calcul d'une valeur ajustée (fitting, smoothing) : prenons l'exemple où nous voulons résumer un nuage de points par une droite. On ajuste une droite au nuage de points.

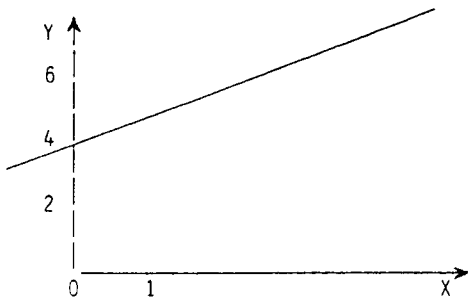


A chaque valeur Y observée correspond une valeur Y ajustée qui est en général différente. Cette valeur ajustée est déterminée par un calcul (l'équation de la droite dans notre exemple).

Critère des moindres carrés (least squares method) : prenons l'exemple où nous voulons synthétiser un nuage de points par une droite.

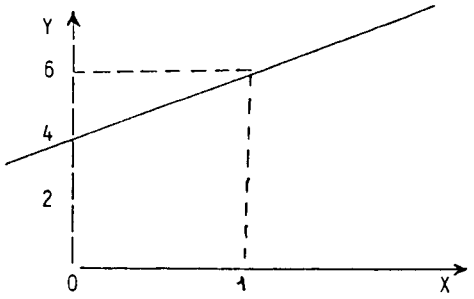
Le critère des moindres carrés est une méthode qui permet de définir la droite qui s'ajuste au mieux sur les Y observés, à partir du calcul des carrés des écarts entre les Y observés et les Y ajustés.

Ordonnée à l'origine (intercept) :



L'ordonnée à l'origine correspond à la valeur de Y quand $X = 0$. Sur le schéma ci-dessus, elle vaut 4.

Pente (slope)



La pente est déterminée par la variation de Y notée ΔY correspondant à une augmentation de $X(\Delta X)$ égale à une unité.

pente $= \frac{\Delta Y}{\Delta X}$

pente $= \Delta Y$ si $\Delta X = 1$. Sur le schéma ci-dessus, la pente est 2.

Régression linéaire (linear regression) : si on étudie l'influence du temps sur la longueur de cristaux de glace et si on suppose que la **relation** entre ces deux variables est **linéaire**, on peut alors rechercher l'équation d'une droite qui s'ajuste au mieux aux observations réalisées pour cette étude. Cette droite est appelée droite de régression linéaire.

Relation linéaire : une relation entre deux variables X et Y est linéaire lorsque la variation de l'une des deux variables est proportionnelle à celle de l'autre.

Résidu (residual) : le résidu correspond à l'écart entre Y observé et Y calculé pour une valeur de X donnée (voir **critère des moindres carrés**).

Variable expliquée et variable explicative : prenons l'exemple où l'on étudie l'influence du temps sur la longueur des cristaux de glace. On cherche à expliquer la longueur des cristaux de glace par le temps. La longueur des cristaux est appelée variable expliquée et le temps est la variable explicative.