

date 05/10

## Introduction à l'analyse de régression linéaire :

### Rappels :

**Variable aléatoire :** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, on appelle toute fonction mesurable  $X$  définie d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$  dans un espace mesurable  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une variable aléatoire réel :

$$X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$$\omega \mapsto X(\omega)$$

Soit  $X$  une V.a qui peut prendre les valeurs  $x_1, \dots, x_n$ , alors :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (n: \text{nombre des individus})$$

Si  $\bar{X} = 0$ , on dit que  $X$  est une V.a centrée

**La variance :** Pour une V.a  $X$  de moyenne  $\bar{X}$ , la variance est :

$$\begin{aligned} \text{Var}(X) &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{X} + \bar{X}^2) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{2\bar{X}}{n} \sum_{i=1}^n x_i + \frac{1}{n} (n\bar{X}^2) \end{aligned}$$

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{X}^2 + \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2$$

**L'écart type :**  $S(X) = \sqrt{\text{Var}(X)}$

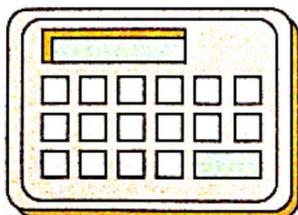
**La Covariance :** Soit  $X$  et  $Y$  de V.a qui peut prendre les valeurs  $(x_1, \dots, x_n)$  et  $(y_1, \dots, y_n)$  (respectivement)

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i y_i - x_i \bar{Y} + \bar{X} \bar{Y} - \bar{X} y_i)$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \frac{1}{n} \bar{Y} \sum_{i=1}^n x_i + \bar{X} \bar{Y} - \frac{1}{n} \bar{X} \sum_{i=1}^n y_i$$

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{X} \bar{Y} + \bar{X} \bar{Y} - \bar{X} \bar{Y}$$



$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{X} \bar{Y}$$

**Remarque :**  $\text{Cov}(X, X) = \text{Var} X$

• Si  $\text{Cov}(X, Y) = 0 \rightarrow X$  et  $Y$  sont indépendantes

• le Coefficient de corrélation

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad \text{et } -1 \leq \rho(X, Y) \leq 1$$

L'origine du mot régression du verbe scientifique britannique "Francis Galton" en 1885, lorsque il travaille sur l'hérédité, il veut comprendre la relation entre la taille des pères et celle de leurs fils. Il a remarqué que les

I) Fondements de la régression linéaire les enfants ne reprenaient pas exactement la taille de ses parents

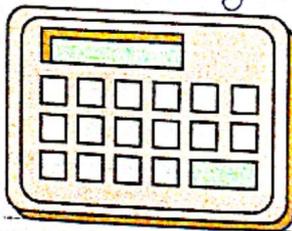
- Si le père était plus grand que la moyenne le fils avait tendance à être un peu plus petit mais plus proche de la moyenne

- Si le père était plus petit que la moyenne, le fils avait tendance à être un peu plus grand, et aussi plus proche de la moyenne

C'est ainsi que Galton a proposé l'idée de la régression vers la moyenne (regression toward mediocrity). Cette idée est importante car elle montre que les valeurs très élevées ou très basses ont tendance à se rapprocher de la moyenne quand on répète les observations plus tard. Cette découverte a permis de développer les méthodes de régression statistique qui servent aujourd'hui à analyser les relations entre différentes variables et à faire des prévisions.

## 1) Introduction et Régression linéaire simple

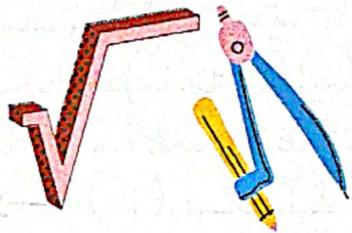
1.1) Introduction à la régression linéaire : La RL se classe parmi les méthodes d'analyse multivariées qui traitent des données quantitatives où l'objectif principale est de rechercher une liaison linéaire entre une variable  $y$  quantitative et une ou plusieurs variables  $x$  également quantitative.



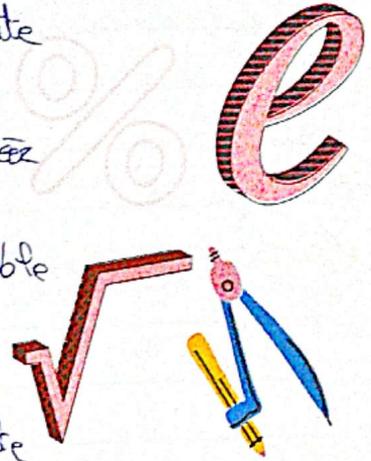
- C'est la méthode la plus utilisée sur :

• une méthode ancienne

• c'est l'outil de base de plupart des modélisations plus sophistiquées comme la régression logistique.



**Def:** R à simple si cherché à ajuster une droite  $y = ax + b$  qui permet de mesurer la relation entre une variable  $y$  appelée variable à expliquer (à regresser, réponse, variable dépendante) et une variable  $x$  appelée explicative (regresseur, variable indépendante)



**Tq:** (a) la pente de la droite  
(b) l'ordonnée du point d'intersection de la droite avec l'axe verticale  $x = 0$

**Exep:** prédire le prix d'une maison en fonction de sa surface la relation entre la taille et le prix

**Equation Générale:** la relation observée sur l'échantillon n'est pas exacte le nuage des points est étiré mais les points ne sont pas alignés pour rendre compte de cette situation on écrit la relation entre  $y$  et  $x$  sous la forme générale suivante:

$$y = ax + b + \epsilon$$

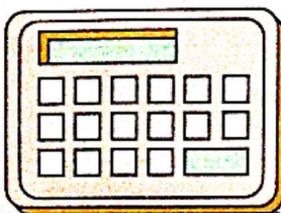
$y =$  droite + erreur

**Tq:** le terme d'erreur qui varie de façon aléatoire d'un individu à l'autre on applique cette équation générale aux observations de  $y$  et aux valeurs correspondantes de  $x$  on écrit le modèle sous forme  $y_i = ax_i + b + \epsilon_i$  pour  $i = 1, \dots, n$

Dans le modèle, les variables  $\epsilon_i$  ne sont pas observées et les coefficients  $a$  et  $b$  ne sont pas connus.

Condition sur les erreurs on pose des conditions sur les erreurs pour étudier le modèle

(1) les erreurs ~~pas~~ sont des variables indépendantes, et des variance  $\text{Var}(\epsilon) = \sigma^2$



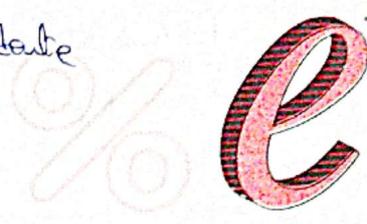
(2) Centrés ( $E(\epsilon) = 0$ )

**1.2 / Estimation des paramètres**

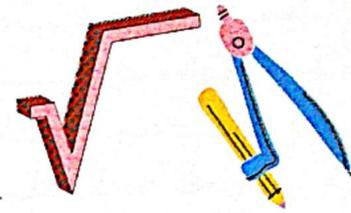
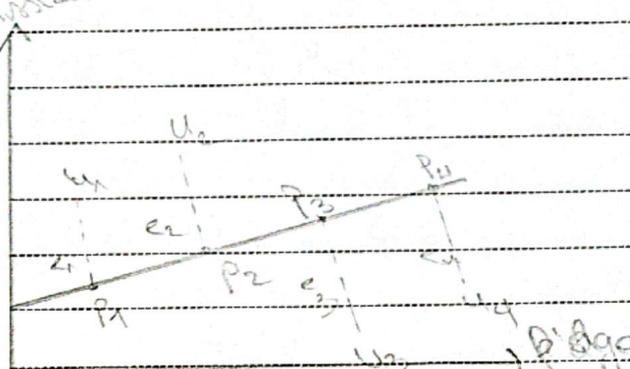
Exemples

Le modèle étant posé il faut estimer numériquement les paramètres du modèle, c.à.d. Calculer les

valeurs numériques des coefficients qui correspondante de mieux aux données pour déterminer la droite qui est la plus proches des points et la plus proche des points.



(a) Critère des moindres Carrés :



température en fonction de l'âge.

On projette les points  $u_1, \dots, u_4$  parallèlement à l'axe des y sur la droite on obtient les points  $P_1, P_2, P_3, P_4$ , comme dans le graphique. Le Critère retenu pour déterminer la droite D passant au plus près de tous les points sera tel que :

La somme des carrés des écarts (SCE) des points observés  $u_i$  à la droite solution soit minimum.

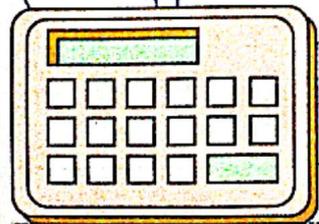
la droite solution sera appelée droite de régression de y sur x.

(b) Formules de calcul des coefficients estimés :

La somme de carrés de écart (SCE) est donner par :

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

La valeur de cette fonction S est minimum lorsque les dérivées par rapport à a et b s'annulent la solution est obtenue en résolvant



le système :

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \text{et} \\ \frac{\partial S}{\partial b} = 0 \end{cases}$$

~~Les dérivées s'annulent pour deux valeurs a et b.~~

Les dérivées par rapport à a et b sont :

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^n x_i (y_i - ax_i - b)$$

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

Les Dér. s'annulent pour deux valeurs a et b. Solution des deux équations à deux inconnues :

$$\begin{cases} \sum_{i=1}^n x_i (y_i - ax_i - b) = 0 & (1) \\ \sum_{i=1}^n (y_i - ax_i - b) = 0 & (2) \end{cases}$$

• L'équation (1) donne :  $\sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0$

• " (2) " :  $\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb = 0$

En dérivant par n :

$$\bar{y} - a\bar{x} - b = 0 \Rightarrow b = \bar{y} - a\bar{x}$$

On remarque que la droite solution passe par le Centre de gravité

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

En remplaçant b dans (1)

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

et faisant passer en facteur :  $\frac{1}{n} \sum x_i y_i - \bar{y}\bar{x} - a \left( \frac{1}{n} \sum x_i^2 - \bar{x}^2 \right) = 0$

$$\Rightarrow a = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

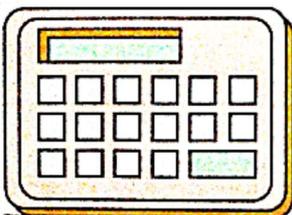
**Exemple :** on étudie la relation entre le nombre d'heures passé à étudier les maths et le résultat au test.

On suppose qu'il y a 8 étudiants.

• trouver la droite d'ajustement.

(1,2) (1,3) (1,4) (1,5) (1,6) (1,7) (1,8) (1,9)

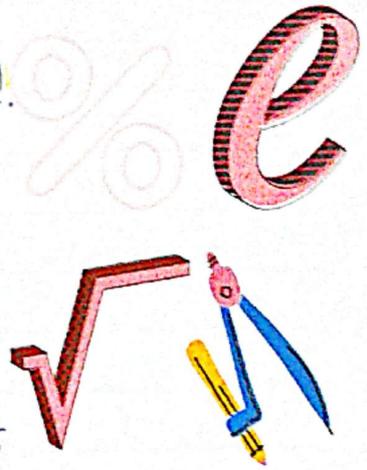
**Solution :**



date \_\_\_/\_\_\_

0 + a    n = 8

Hours (x)	Note out test (y)	x <sup>2</sup>	xy	amā a = $\frac{Cov(x,y)}{Var(x)}$
1	1	1	1	
2	11	4	22	b = $\bar{y} - a\bar{x}$
3	13	9	39	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
4	14	16	56	$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
5	15	25	75	
6	15,5	36	93	$\bar{x} = \frac{1}{8} (36) = 4,5$
7	17	49	119	
8	18	64	144	$\bar{y} = \frac{1}{8} (104,5) = 13,0625$
$\Sigma$	36	104,5	204	549



$$Cov(x,y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \Rightarrow Cov(x,y) = \frac{1}{8} (549) - (4,5 \times 13,0625)$$

$$Cov(x,y) = 9,84$$

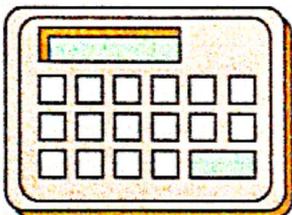
$$Var(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \Rightarrow Var(x) = \frac{1}{8} (204) - (4,5)^2$$

$$Var(x) = 5,25$$

$$a = \frac{9,84}{5,25} = 1,87 \quad ; \quad b = 13,0625 - (1,87 \times 4,5)$$

$$b = 4,65$$

$$\hat{y} = 1,87 \hat{x} + 4,65$$



date 12 / 11 / 20

## Q) Comment mesurer la qualité de l'ajustement?

### 1) Le Coefficient de détermination $R^2$

Afin d'avoir une idée globale de la qualité de l'ajustement linéaire, on définit  $R^2$  le Coefficient de détermination qui est le carré de Coefficient de corrélation  $R$ .

Il mesure la part de la variation totale  $y$  expliquée par le modèle de régression sur  $X$ .

La valeur ajustée résiduelle et somme des carrés des résidus

•  $\hat{y}_i = \hat{a}_i x_i + \hat{b}_i$ , s'appelle la valeur ajustée ou prédite de  $y$  par le modèle

•  $e_i = y_i - \hat{y}_i$ , s'appelle le résidu d'observation  $i$ , c'est l'écart entre la valeur de  $y$  observée sur l'individu  $i$  et la valeur prédite

• SCR la variance résiduelle : la somme des carrés des résidus est

$$SCR = \sum_{i=1}^n e_i^2$$

elle mesure la distance de la droite de régression aux points du nuage de points qui est minimale au sens le moindre carrés

• SCE : représente la variance expliquée par la régression (mesure de variation des valeurs ajustées autour de la moyenne  $\bar{y}$ ).

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

• SCT : représente la variance totale des observations  $y_i$  autour de leur moyenne  $\bar{y}$

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

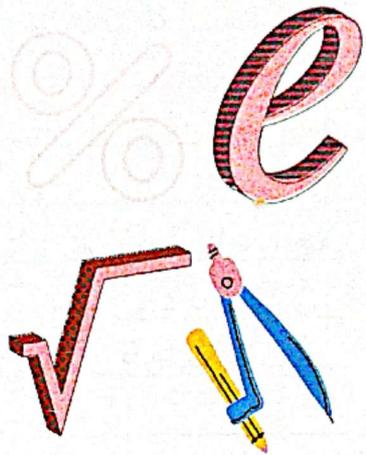
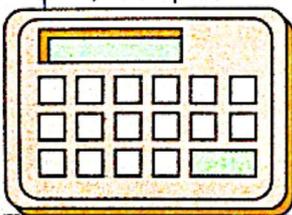
$$SCT = SCE + SCR$$

• Interprétation de  $R^2$  :  $R^2$  qui varie entre 0 et 1 ( $0 \leq R^2 \leq 1$ ).

mesure la proportion de variation totale de  $y$  autour de la moyenne expliquée par la régression, quand d'prise en compte par le modèle

plus  $R^2$  se rapproche de la valeur 1, on dit meilleure modèle aux données.

• Un  $R^2$  faible (s'approche à la valeur 0), signifie que le modèle a une faible pouvoir explicatif.



date 12/1/10

Travailler à partir de l'exemple précédente

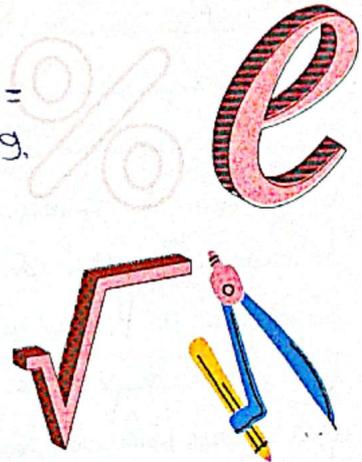
Calculer  $R^2 = \frac{\text{Cov}(\hat{y}, y)}{\sqrt{\text{Var}(\hat{y}) \text{Var}(y)}} = \frac{9,84}{\sqrt{5,25 \times 27,09}} = 0,734$

$R^2 = \frac{\text{SCE}}{\text{SCT}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{146,85}{199,97} = 0,734$

Le modèle d'ajustement est donné par :

$\hat{y} = 1,87x + 4,65$  ;  $\bar{y} = 13,063$

$R^2 = \frac{146,85}{199,97} = 0,734$  ;  $R^2 = 73\%$



X	Y	y	ec	(y - y)	(y - y) <sup>2</sup>	(y - y)	(y - y) <sup>2</sup>
1	1	6,52	-5,52	-6,543	42,81	-12,083	145,99
2	11	8,39	2,61	-4,673	21,83	-2,083	4,33
3	13	10,26	2,74	-2,803	7,85	0,83	0,688
4	14	12,13	1,87	-0,933	0,87	-0,917	0,84
5	15	14	1	0,937	0,877	1,917	3,67
6	15,5	15,87	-0,37	2,807	7,879	2,417	5,84
7	17	17,74	0,74	4,677	21,874	3,917	15,34
8	18	19,61	-1,61	6,547	42,86	4,917	24,17
36	104,5				146,85		199,97

### 113 Théorème de Gauss-Markov

a) Hypothèses essentielles : Pour que le théorème s'applique il faut :

- linéarité du modèle par rapport aux paramètres :  $y = \beta x + \varepsilon$

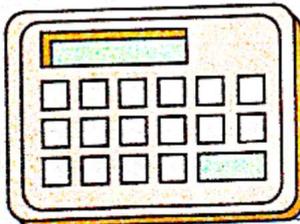
-  $E(\varepsilon) = 0$

- les erreurs ont des même variance  $\text{Var}(\varepsilon_i) = \sigma^2$  (Homoscedasticité),

et absence d'autocorrélation  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$  pour  $i \neq j$

Matriciellement :  $\text{Var}(\varepsilon_i) = \sigma^2 I_n$ , Tq :  $I_n$  est la matrice d'identité

Les estimateurs considérés doivent être linéaires en y et sans biais ( $E(\hat{\beta}) = \beta$ ), c.à.d. :  $\hat{\beta} = c_1 y_1 + \dots + c_n y_n$



149/10

date 19 / 10

## b. Théorème de Gauss-Markov (France)

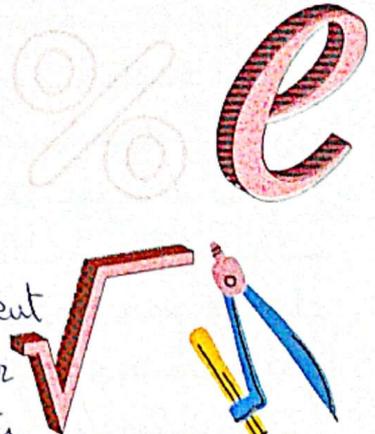
On observe  $n$  observations indépendantes

$y_i = a x_i + b + \varepsilon_i, i = 1, \dots, n$ , où les  $\varepsilon_i$  sont  
permutés tous les estimateurs linéaires et sans biais

des paramètres ( $a$  et  $b$ ), les estimateurs de  
moindre carré ont la variance minimale, on peut

résumer tout cela en disant que l'estimateur  
par moindre carré est le "BLUE" (Best Linear

Unbiased estimator), on peut écrire  $BLUE(\hat{a}, \hat{b}) = MC(\hat{a}, \hat{b})$



## c. Interprétation Géométrique

L'estimation de moindre carré correspond à la projection  
orthogonale du vecteur  $y$  sur l'espace colonne de  $X$  (espace engendré par

les vecteurs  $(x_i)$  et  $(1)$ ). Les résidus  $r = y - \hat{y}$  sont orthogonaux à

chaque colonne de  $X$ . Cette vision explique pourquoi les MC

## d. Inférence statistique pour les paramètres ?

L'inférence statistique et la régression linéaire sont deux concepts  
statistiques qui fonctionnent ensemble. La régression linéaire

modélise la relation entre des variables (prédiction), tandis que

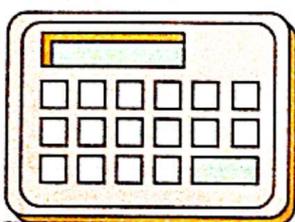
l'inférence statistique utilise des échantillons pour tirer des

conclusions sur une population plus large souvent en testant

## a. Inférence statistique

Déf: C'est le processus qui consiste à utiliser les données  
d'un échantillon ~~entière~~ entière, avec une mesure de l'incertitude.

- Méthodes de calcul



- teste d'hypothèse

- Intervalle de confiance

date 26/10

## b) Principe d'estimation

soit le modèle  $y_i = ax_i + b + \varepsilon_i$  ;  $\varepsilon_i \sim N(0, \sigma^2)$

$\varepsilon_i$  écrit sur la droite et le point  $i$ , on estime  $a$  et  $b$  telle que SCR est minimum ( $SCR = \sum_{i=1}^n \varepsilon_i^2$ )

Estimation de la pente =  $a$  : la pente  $a$  est donnée par la formule  $a = \frac{cov(x,y)}{var(x)}$   
pour estimer cette paramètre, il faut trouver un estimateur pour  $cov(x,y)$  et  $var(x)$

La variance de  $X$  est estimée par (dans le cas d'un échantillon de taille  $n$ ) :  $S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n(\bar{x})^2}{n-1}$

Le estimateur de la covariance  $cov(x,y)$  est donné par :

$$\hat{cov}(x,y) = S_{xy} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n-1} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{n-1}$$

Alors  $a$  est estimé par  $\hat{a} = \frac{S_{xy}}{S_x^2}$  ;  $\hat{a}$  = estimateur de  $a$

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

Estimation de l'origine =  $b$  :

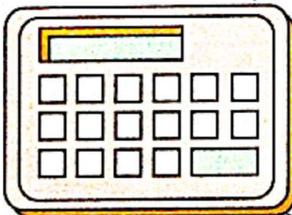
la droite passe par  $m_x$  et  $m_y$  ; où  $m_x = \bar{x}$  ;  $m_y = \bar{y}$   
( $\bar{x}, \bar{y}$  sont des estimateurs sans biais pour  $m_x$  et  $m_y$  respectivement)  
Alors  $b$  est estimé par  $\hat{b} = \bar{y} - \hat{a} \bar{x}$

Remarque : une fois les paramètres  $a$  et  $b$  sont estimés, on peut déduire les valeurs ajustées  $\hat{y}_i = \hat{a} x_i + \hat{b}$  pour les résidus obtenus

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

Estimation de Coefficient de Corrélation  $\rho$  :  $\hat{\rho} = \frac{S_{xy}}{S_x S_y}$

$$R = \rho(x,y) = \text{corr}(x,y) \text{ est estimé par } \hat{R} = \frac{S_{xy}}{S_x S_y}$$



## c) Intervalle de Confiance

Sous les hypothèses :  $E(\varepsilon_i) = 0$  ;  $\sigma^2 = var(\varepsilon_i)$  et

$cov(\varepsilon_i, \varepsilon_j) = 0$  si  $i \neq j$  ; on a

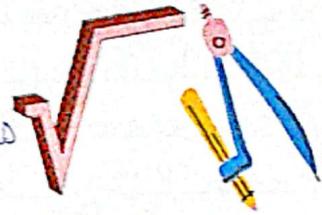
$\hat{a}$  est un estimateur sans biais ( $E(\hat{a}) = a$ ) et efficace

$\hat{\sigma}^2$  est un estimateur sans biais et efficace de  $\sigma^2$   
 $S_e^2 = \frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2 = \text{MSE}$  est un estimateur sans biais de  $\sigma^2$



on suppose de plus que les conditions suivantes sont vérifiées

- $\varepsilon_i \sim N(0, \sigma^2)$
- $\varepsilon_i$  sont mutuellement indépendants, alors sous les hypothèses précédentes on a:



l'intervalle de confiance pour  $\alpha$  et  $\frac{1}{2}$

$$I_C(\alpha) = \left[ \hat{\sigma}^2 \pm t_{\left(n-2; 1-\frac{\alpha}{2}\right)} S_e \sqrt{\frac{1}{\sum x_i^2 - n\bar{x}}} \right]$$

$$I_C(\beta) = \left[ \hat{\beta} \pm t_{\left(n-2; 1-\frac{\alpha}{2}\right)} S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2 - n\bar{x}}} \right]$$

où  $t_{\left(n-2; 1-\frac{\alpha}{2}\right)}$  est le quantile d'ordre  $1-\frac{\alpha}{2}$  de la loi student à  $(n-2)$  degrés de liberté.

2/11

### 1. Tests d'hypothèses

Pratiquement, les hypothèses portant sur "a" ont plus d'intérêt que celles portant sur "b".

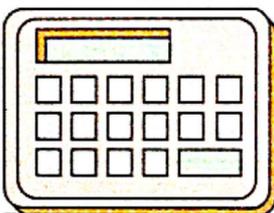
Il s'agit de tester les hypothèses

- $H_0: a=0 \equiv R=0$  (il n'y a pas de lien entre y et x)
- $H_1: a \neq 0 \equiv R \neq 0$  (il y a un lien entre y et x)

Accepter  $H_0$  implique que l'on conclut qu'il n'y a pas de relation linéaire entre x et y, Ceci peut signifier que la relation entre y et x n'est pas linéaire.

La variation de x influe peu ou pas sur la variation de y. Au contraire, rejeter  $H_0$  implique que l'on conclut que la variation de x influe sur la variation de y.

sous l'hypothèse  $H_0$ : la statistique



$$T_n = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S_e} \sim T_{(n-2)}$$

Pour une hypothèse alternative:

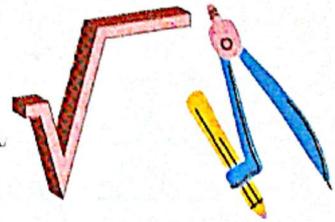
$H_1: a \neq 0$  bilatéral, on rejette  $H_0$  avec un risque  $\alpha$  si:  $|t| > t_{(n-2), 1-\frac{\alpha}{2}}$

où  $t$  est la réalisation de  $T_n$

• Test de Fisher

s'intéresse à la significativité global d'un modèle. Il s'agit de tester les hypothèses

$$\begin{cases} H_0: b = 0 \\ H_1: b \neq 0 \end{cases}$$



### 1.51 prédictions

On déduit prévoir à l'aide du modèle, la valeur de la variable  $y$  pour une valeur non observée

• soit  $x_{n+1}$  une nouvelle valeur, pour laquelle nous voulons prédire  $y_{n+1}$  d'après le modèle on a  $y = ax_{n+1} + b + \epsilon_{n+1}$

où  $y_{n+1}$  et  $\epsilon_{n+1}$  sont des variables aléatoires

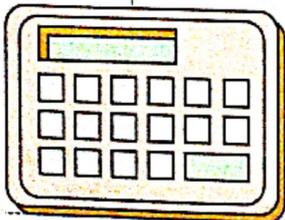
La prédiction naturelle est alors  $\hat{y}_{n+1} = \hat{\alpha}x_{n+1} + \hat{\beta}$ , et l'erreur de prédiction est définie par  $\hat{\epsilon}_{n+1} = y_{n+1} - \hat{y}_{n+1}$

### Intervalle de prédictions

Intervalle de prédictions est donc un intervalle dans lequel une future observation  $y_{n+1}$ , va tomber avec une certaine probabilité sous les hypothèses du modèle (incluant l'hypothèse de normalité) on a l'intervalle de prédictions pour  $y_{n+1}$  en  $x_{n+1}$  au niveau de confiance  $1 - \alpha$  est

$$I.P.(y_{n+1}) = \left[ \hat{y}_{n+1} \pm t_{(n-2), 1-\frac{\alpha}{2}} \cdot S_e \sqrt{1 + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} \right]$$

### • Comparaison entre I.C. et I.P.



• Les longueurs des deux types d'intervalles croissent lorsque  $x_{n+1}$  s'éloigne de  $\bar{x}$

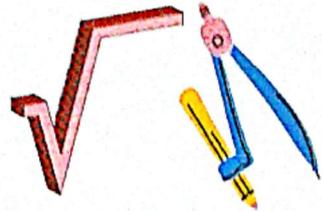
• L'I.C. de la droite de régression ne convient pas pour effectuer des prévisions puisqu'il concerne

date ..... / ..... / .....

La vraie réponse moyenne au point  $x = x_{n+1}$   
Soit une paramètre de la population et non  
une nouvelle observation

- L'IP au  $x_{n+1}$  est toujours plus grand que IC  
au  $x_{n+1}$  car il dépend de l'erreur associée aux  
futurs observations

- L'IP n'est valide que pour une  
nouvelle observation à la fois



Handwritten notes on a lined notebook page. The page contains several horizontal lines for writing, with the text from the previous blocks already written in cursive.

