TD Régression - M1 Info - Le Mans Université

Exercice 1 :

Un père a deux garçons, et s'inquiète de la croissance de son cadet qu'il trouve petit. Il décide donc de faire un modèle familial à partir des mesures de taille en fonction de l'âge de l'aîné : $n = 10$ $\bar{x} = 7,5$ $\bar{y} = 122,67$ $(\bar{x})^2 = 56,25$

	âge	3	4	5	6	7	8	9	10	11	12
x											
y	taille	96	104.8	110.3	115.3	121.9	127.4	130.8	136	139.7	144.5
xy											

1. Représenter les données sur un graphique et justifier l'utilisation d'un modèle de régression linéaire simple. Discuter les hypothèses nécessaires.

Il faut que l'erreur suive une loi normale cR. ε_i indep ε_j $Cov(\varepsilon_i, \varepsilon_j) = 0$

2. Estimer les coefficients de la régression et tracez sur le graphique la droite de régression estimée. $Cov(x,y) = 43,02$ $Var(x) = 8,25$

3. Calculer le R^2 et représenter les résidus. La régression semble-t'elle valable ?

Exercice 2 :

Slide (17) :

On s'intéresse à la relation entre les observations de deux variables x et y du tableau suivant :

x	-1	0	1	2	5
y	-1	3	2.5	5	2

1. Représenter graphiquement Y en fonction de X

2. On admet l'existence d'une relation linéaire de la forme $y = ax + b + \epsilon$ calculez les estimateurs des cœfficients a et b.

2. Maintenant on veut établir une régression polynomiale de la forme $y = ax^2 + bx + c + \epsilon$ entre les deux variables. Calculez les estimateurs des cœfficients a, b et c.

3. Calculer l'erreur quadratique moyenne de chaque modèle (moyenne des erreurs au carré $\sum \varepsilon_i^2$

4. Quel est le meilleur modèle pour représenter la relation entre x et y ? $\frac{1}{n}\left(\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2\right)$

en AL : si outliner, l'enlever

avec 5 dimensions, on passe par tout les points mais ce n'est pas bon : surapprentissage.

Exercice 3 :

Les données du tableaux ci-dessous montre la consommation de bonbons (Y) en millions de livres et la population (X), en millions, pour 17 pays en 1991.

Observations (i)	Pays	Population	Consommation
1	Australie	17,3	327,4
2	Autriche	7,7	179,5
3	Belgique/Luxembourg	10,4	279,4
4	Danemark	5,1	139,1
5	Finlande	5,0	92,5
6	France	56,9	926,7
7	Allemagne	79,7	2186,3
8	Irlande	3,5	96,8
9	Italie	57,8	523,9
10	Japon	124,0	935,9
11	Pays-bas	15,1	444,2
12	Norvège	4,3	119,7
13	Espagne	39,0	300,7
14	Suède	8,7	201,9
15	Suisse	6,9	194,7
16	Royaume-Uni	57,7	1592,9
17	États-Unis	252,7	5142,2

1. Représenter graphiquement Y en fonction de X
2. Soit le modèle : $Y = \beta_0 + X\beta_1 + \epsilon$.

   Estimer les paramètres β_0 et β_1 et écrire l'équation de la droite de regression
3. Dessiner sur le graphe la droite de regression obtenue
4. Calculer les résidus $\sum_i \epsilon_i$
5. Quel pourcentage de la variation totale est expliqué par la variable population ?] C'est quoi ? R^2

questions Bougares

Exercice 4 :

On s'intéresse dans un secteur de production à la relation entre les bénéfices réalisés par les entreprises et le budget annuel qu'elles consacrent à la publicité. 15 observations ont été réalisées :

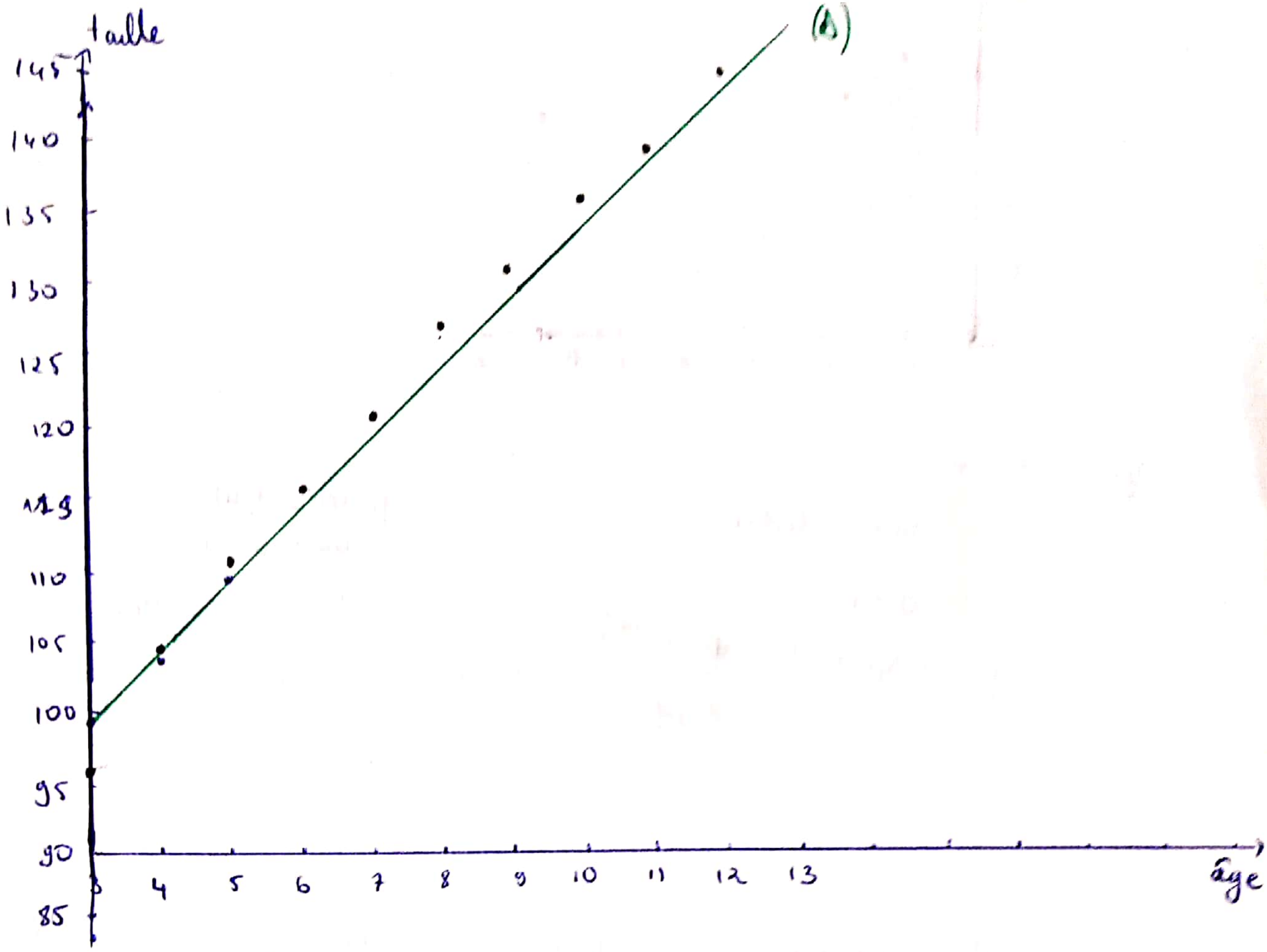
Budget de publicité	15	8	36	41	16	8	21	21	53	10	32	17	58	6	20
Bénéfices	48	43	77	89	50	40	56	62	100	47	71	58	102	35	60

1. On veut établir une régression linéaire entre les deux variables, qu'elle doit être la variable endogène ? un

   ↳ dépendante
2. On admet l'existence d'une relation linéaire de la forme $y_i = ax_i + b + \epsilon$ calculez les estimateurs des cœfficients a et b.
3. Précisez l'équation d'analyse de la variance, calculer ses valeurs et en déduire le cœfficient de détermination (R^2)

[TD1 : Exercice 1] Regression linéaire.

Exo 1



1-) Les points ne sont pas éparpillés, on peut faire passer une droite de regress $y = f(x)$
taille âge

$y = \beta_0 + \beta_1 x$

83,52 5,64 x

$\boxed{y = 83,415 + 5,234 x}$

2-) $\beta_1 = \frac{Cov(x,y)}{Var(x,y)} = \frac{43,02}{8,25} = 5,234$

$\beta_0 = \bar{y} - (5,234)\bar{x}$

$= 183,415$

$\cancel{\hat{y}_0 = 83,415}$

$\hat{y}_1 =$

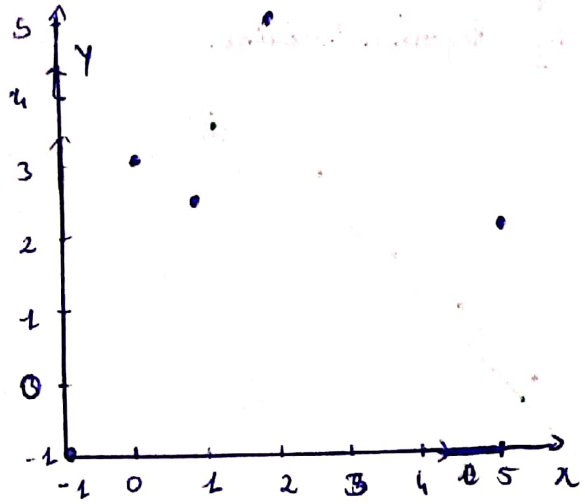
$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

3.) R^2 = pouvoir prédictif du modèle 0-1

~~1,164~~ 0,95 pouvoir prédictif fort

Exo 2

1.)



2.) $y = ax + b + \epsilon$

$var(x) = 4,24$

$0,35$

$\bar{y} = 0,35\bar{x} + b \quad b = \bar{y} - 0,35\bar{x}$

$= 1,81$

$cov = 1,48$

$var = 4,24$

$r = \frac{1,48}{4,24} = 0,349$

$y = 0,38 \quad b = 1,8.$

3.) $y = ax^2 + bx + c + \epsilon$

$$M = \begin{bmatrix} (-1)^2 & (0)^2 & 1^2 & 2^2 & 5^2 \\ -1 & 0 & 1 & 2 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(x,y)	ax^2+bx+c
$(-1,-1)$	$a(-1)^2 + b(-1) + c = -1$
$(0,3)$	$a(0)^2 + b(0) + c = +3$
$(1,2,5)$	a - - - $= 2,5$
$(2,5)$	- - - - $= 5$
$(5,2)$	- - - - - $= 2$

$$Y = \begin{bmatrix} -1 \\ 3 \\ 2,5 \\ 5 \\ 2 \end{bmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \underbrace{(M.M^T)^{-1}}_{\substack{3,5 \; 5\times3 \\ 3\times3}}.MY \; \epsilon$$

$$(MM^T)^{-1} = \begin{bmatrix} 53/3822 & -16/273 & -5/1274 \\ -16/273 & 23/78 & -9/182 \\ -5/1274 & -9/182 & 187/637 \end{bmatrix}$$

$$M^T = \begin{bmatrix} (-1)^2 & -1 & 1 \\ 0^2 & 0 & 1 \\ 1^2 & 1 & 1 \\ 2^2 & 2 & 1 \\ 5^2 & 5 & 1 \end{bmatrix}$$

$$MY = \begin{bmatrix} 143/2 \\ 47/2 \\ 23/2 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -549/1274 \\ 395/182 \\ 2464/1274 \end{bmatrix} = \begin{bmatrix} -0,43 \\ 2,17 \\ 1,93 \end{bmatrix}$$

$y = -0,43x^2 + 2,17x + 1,93$