

Niveau : L3 EMB

Le corrigé de TD₁

BERKAI née MECHERI Kheira

Corrigé 1.

1. Représentation du nuage de points

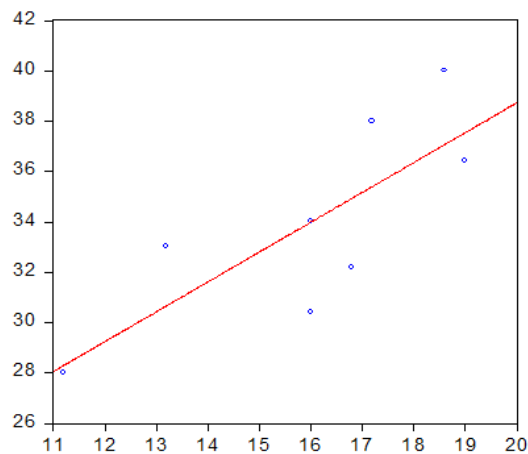


FIGURE 1: Le nuage de points

D'après la figure, nous constatons que le nuage de points ne donne pas une idée claire sur le type de relation qu'il existe entre les deux variables, et que la droite de régression ne passe pas par l'origine.

$\widehat{B}_0$ et $\widehat{B}_1$ sont les quantités qui minimisent les distances verticales entre les observations Y_t et la droite de régression théorique $Y_t = B_0 + B_1X_t$, et la droite de régression estimée est donnée par :

$$\widehat{Y}_t = \widehat{B}_0 + \widehat{B}_1 X_t.$$

Afin de calculer les estimateurs des coefficients de régression, nous réalisons le tableau suivant :

n	X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2	$\widehat{Y}_i$	$e_i = Y_i - \widehat{Y}_i$	e_i^2
1	11.2	28	313.6000	125.44	784.0000	28.309359	-0.309359	0.095703
2	13.2	33	435.6000	174.24	1089.0000	30.680460	2.319540	5.380265
3	16	30.4	486.4000	256.0000	924.1600		-3.599999	12.960002
4	16	34	544.0000	256.0000	1156.0000	34.000000	0.000000	0.000000
5	16.8	32.2	540.9600	282.2400	1036.8400	34.000000	-2.748440	7.553924
6	17.2	38	653.6000	295.8400	1444.0000	34.948440	2.577339	6.6426795
7	18.6	40	744.0000	345.9600	1600.0000	35.422660	2.917569	8.512211
8	19	36.4	691.6000	361.0000	1324.9600	37.556650	-1.156650	1.337840
$\sum$	128	272	4409.7600	2096.7200	9358.9600			42.48263

$$\bar{X} = \frac{128}{8} = 16 \text{ et } \bar{Y} = \frac{272}{8} = 34$$

$$\widehat{B}_1 = \frac{\sum_{t=1}^n X_t Y_t - n \bar{X} \bar{Y}}{\sum_{t=1}^n X_t^2 - n \bar{X}^2} = \frac{4409.76 - 8(16)(34)}{2096.72 - 8(16)^2} = 1.18555008$$

$$\widehat{B}_0 = \bar{Y} - \widehat{B}_1 \bar{X} = 34 - 1.18555008(16) = 15.031199$$

On écrit

$$\widehat{Y}_i = 15.031199 + 1.18555008 X_i \quad (1)$$

2. ♣ On a $\sum_{t=1}^n e_t^2 = 42.48263$. Et $\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^n e_t^2}{(n-2)} = \frac{42.48263}{8-2} = 7.080438$

$$\clubsuit \text{ } Var(\widehat{B}_1) = \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{t=1}^n (X_t - \bar{X})^2} = \frac{7.080438}{2096.7200 - 8(16)^2} = 0.145329$$

d'où

$$\widehat{\sigma}_{\widehat{B}_1} = 0.381220$$

$$\clubsuit \text{ } Var(\widehat{B}_0) = \widehat{\sigma}_\varepsilon^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{t=1}^n (X_t - \bar{X})^2} \right]$$

3.

4. Avant de donner le tableau d'analyse de la variance (ANOVA) , on doit d'abord calculer SCT et SCE

♣ On sait que : $R^2 = 0.978944$, $SCR = \sum_{t=1}^n e_t^2 = 79.03797$ et que

$$R^2 = 1 - \frac{SCR}{SCT}.$$

Alors

$$SCT = \frac{SCR}{(1 - R^2)} = \frac{79.03797}{(1 - 0.978944)} = 3753.70298 \quad (2)$$

♣ On a $SCT = SCE + SCR$, donc

$$SCE = SCT - SCR = 3753.70298 - 79.03797 = 3674.66501 \quad (3)$$

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1	$SCE = 3674.66501$	1	$\frac{SCE}{1} = 3674.66501$
Résiduelle	$SCR = 79.03797$	$n - 2 = 15 - 2 = 13$	$\frac{SCR}{n-2} = \frac{79.03797}{13} = 6.07984$
Totale	$SCT = 3753.70298$	$n - 1 = 15 - 1 = 14$	

Corrigé 2.

$$\bar{X} = 34.9; \quad \frac{1}{20} \sum_1^{20} (X_i - \bar{X})^2 = 28.29; \quad \bar{Y} = 18.34;$$

$$\frac{1}{20} \sum_1^{20} (Y_i - \bar{Y})^2 = 2.85; \quad \frac{1}{20} \sum_1^{20} (X_i - \bar{X})(Y_i - \bar{Y}) = 6.26.$$

1. On sait que : $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i, \quad i = \overline{1, 20}$

$$\hat{\beta} = \frac{\sum_1^{20} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_1^{20} (X_i - \bar{X})^2} = \frac{20 \cdot 6.26}{20(28.29)} = 0.22$$

$$\text{Et } \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = 18.34 - 0.22(34.9) = 10.66$$

Donc le modèle est :

$$\hat{Y}_i = 10.66 + 0.22 X_i, \quad i = \overline{1, 20} \quad (4)$$

2. $R^2 = \frac{SCE}{SCT}$

On sait que :

$$\begin{cases} SCE = \hat{\beta}^2 \sum_1^{20} (X_i - \bar{X})^2, \\ SCT = \sum_1^{20} (Y_i - \bar{Y})^2, \end{cases} \quad (5)$$

Donc

$$R^2 = \frac{\hat{\beta}^2 \sum_1^{20} (X_i - \bar{X})^2}{\sum_1^{20} (Y_i - \bar{Y})^2} = \frac{(0.22)^2 (28.29) 20}{20(2.85)} = 0.48 \quad (6)$$

D'après la valeur du R^2 , on déduit que la qualité de l'ajustement est moyenne.

$$3. SCR = SCT - SCE = 20(2.85 - (0.22)^2(28.29)) = 29.62$$

$$\hat{\sigma}_\varepsilon^2 = \frac{SCR}{n-2} \quad (7)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{29.62}{20-2} = 1.65$$

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^{20} (X_i - \bar{X})^2} \quad (8)$$

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{1.65}{20(28.29)} = 0.0029$$

$$\text{D'où } \hat{\sigma}_{\hat{\beta}} = 0.054$$

Donc, le test de Student pour β est définie par :

$$H_0 : \beta = 0 \text{ contre } H_1 : \beta \neq 0$$

Pour cela, on doit calculer $t_{c\beta}$

$$t_{c\beta} = \left| \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} \right| \quad (9)$$

Sous H_0 , on obtient :

$t_{c\beta} = \left| \frac{0.22}{0.054} \right| = 4.07 > t_T(18, 0.025) = 2.101$, $t_T(18, 0.025)$ est la valeur de la statistique de Student lu sur la table avec 18 est degré de liberté et 0.025 est le seuil de confiance.

Donc, la variable X_i est significative.

De la même manière, on va étudier la significativité de α

Le test de Student pour α est définie par :

$$H_0 : \alpha = 0 \text{ contre } H_1 : \alpha \neq 0$$

Pour cela, on doit calculer $t_{c\alpha}$

$$t_{c\alpha} = \left| \frac{\hat{\alpha} - \alpha}{\hat{\sigma}_{\hat{\alpha}}} \right| \quad (10)$$

Dans un premier temps, on doit calculer $\hat{\sigma}_{\hat{\alpha}}$

$$\hat{\sigma}_{\hat{\alpha}}^2 = \hat{\sigma}_{\varepsilon}^2 \left[\frac{\overline{X}^2}{\sum_1^{20} (X_i - \overline{X})^2} + 1 \right] \quad (11)$$

$$\hat{\sigma}_{\hat{\alpha}}^2 = 1.65 \left[\frac{(34.9)^2}{20(28.29)} + 1 \right] = 5.20$$

$$\hat{\sigma}_{\hat{\alpha}} = 2.28.$$

Alors, sous H_0 , on obtient :

$t_{c_{\alpha}} = \left| \frac{10.66}{2.28} \right| = 4.68 > t_T(18, 0.025) = 2.101$, $t_T(18, 0.025)$ est la valeur de la statistique de Student lu sur la table avec 18 est degré de liberté et 0.025 est le seuil de confiance.

D'après ce résultat, on conclut que α est significatif. Ce qui ne permis de dire que ce modèle est validé, mais il faut ajouter d'autre variables pour améliorer le R^2 .

Corrigé 3.

1. L'équation du modèle de régression est donnée par :

$$\hat{Y}_t = \hat{B}_0 + \hat{B}_1 X_t \quad (12)$$

Alors, d'après le tableau, les valeurs des paramètres sont données par "*coefficient*" et en écrit

$$\hat{Y}_t = -23.74045 + 0.765572 X_t \quad (13)$$

2. ♣ $1 \rightarrow \hat{\sigma}_{\hat{B}_0}$,

On sait que :

$$t_{c_{B_0}} = \frac{\hat{B}_0 - 0}{\hat{\sigma}_{\hat{B}_0}} \quad (14)$$

Donc :

$$\hat{\sigma}_{\hat{B}_0} = \frac{\hat{B}_0 - 0}{t_{c_{B_0}}} = \frac{-23.74045}{-11.58787} = 2.048733 \quad (15)$$

- ♣ $2 \rightarrow t_{c_{B_1}}$,

$$t_{c_{B_1}} = \frac{\hat{B}_1 - 0}{\hat{\sigma}_{\hat{B}_1}} = \frac{0.765572}{0.031140} = 24.58467 \quad (16)$$

♣ 3 $\rightarrow \hat{\sigma}_\varepsilon$, on sait que :

$$\hat{\sigma}_\varepsilon^2 = \frac{SCR}{n-2} \quad (17)$$

où $SCR = Sum \ squared \ resid = \sum_{t=1}^n e_t^2 = 79.03797$, d'ou

$\hat{\sigma}_\varepsilon^2 = \frac{79.03797}{15-2} = 6.079844$, et par conséquent :

$$\hat{\sigma}_\varepsilon = \sqrt{6.07984} = 2.465734$$

♣ 4 $\rightarrow F - statistic = F_c$, sa valeur est donnée par la formule suivante :

$$F_c = \frac{R^2/1}{(1-R^2)/(n-2)} = \frac{0.978944}{(1-0.978944)/(15-2)} = 604.40121 \quad (18)$$

3. Comme $F_c = 604.40121 > F_{\frac{\alpha}{2}}(1, n-2) = F_{0.05}(1, 15-2) = 4.67$ et $R^2 = 0.978944$ est très proche de 1, alors ce modèle est globalement significatif au seuil 5%.

4. les deux paramètres sont significatif au seuil 5% car la probabilité de significativité de chaque paramètre $Prod = 0.0000 < 0.025$.

Donc, la variable X explique bien le modèle et que ce modèle contient une constante.

5.

$$B_1 = \widehat{B}_1 \pm t_{0.025}(15-2)\hat{\sigma}_{\widehat{B}_1} = 0.765572 \pm 2.160(0.031140) \quad (19)$$

$$I_{C \ B_1} = [0.69831, 0.83283] \quad (20)$$

Puisque $0 \notin I_{C \ B_1}$, alors B_1 est significatif et la variable X explique bien le modèle.

6. Avant de donner le tableau d'analyse de la variance (ANOVA) , on doit d'abord calculer SCT et SCE

♣ On sait que : $R^2 = 0.978944$, $SCR = \sum_{t=1}^n e_t^2 = 79.03797$ et que

$$R^2 = 1 - \frac{SCR}{SCT}.$$

Alors

$$SCT = \frac{SCR}{(1-R^2)} = \frac{79.03797}{(1-0.978944)} = 3753.70298 \quad (21)$$

♣ On a $SCT = SCE + SCR$, donc

$$SCE = SCT - SCR = 3753.70298 - 79.03797 = 3674.66501 \quad (22)$$

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1	$SCE = 3674.66501$	1	$\frac{SCE}{1} = 3674.66501$
Résiduelle	$SCR = 79.03797$	$n - 2 = 15 - 2 = 13$	$\frac{SCR}{n-2} = \frac{79.03797}{13} = 6.07984$
Totale	$SCT = 3753.70298$	$n - 1 = 15 - 1 = 14$	