Niveau : L3 EMB

Le corrigé de TD₂

BERKAI née MECHERI Kheira

Corrigé 1.

1. La forme matricielle est par l'équation suivante :

$$\underset{(n=10,1)}{Y} = \underset{(n=10,k+1=3)}{X} \underset{(k+1=3,1)}{B} + \underset{(n=10,1)}{\varepsilon} \quad (1)$$

où

- Y est un vecteur aléatoire de dimension $n = 10$,
- X est une matrice de taille $n \times (k + 1) = 10(3) = 30$, connue, appelée matrice du plan d'expérience,
- B est le vecteur des paramètres inconnus du modèle,
- ε est le vecteur des erreurs.

Avec

$$\underset{(n=10,1)}{Y} = \begin{pmatrix} Y_1 = 40 \\ Y_2 = 44 \\ Y_3 = 46 \\ Y_4 = 48 \\ Y_5 = 52 \\ Y_6 = 58 \\ Y_7 = 60 \\ Y_8 = 68 \\ Y_9 = 74 \\ Y_{10} = 80 \end{pmatrix}; \quad \underset{(n=10,k+1=3)}{X} = \begin{pmatrix} 1 & X_{1\ 1} = 6 & X_{2\ 1} = 4 \\ 1 & X_{1\ 2} = 10 & X_{2\ 2} = 4 \\ 1 & X_{1\ 3} = 12 & X_{2\ 3} = 5 \\ 1 & X_{1\ 4} = 14 & X_{2\ 4} = 7 \\ 1 & X_{1\ 5} = 16 & X_{2\ 5} = 9 \\ 1 & X_{1\ 6} = 18 & X_{2\ 6} = 12 \\ 1 & X_{1\ 7} = 22 & X_{2\ 7} = 14 \\ 1 & X_{1\ 8} = 24 & X_{2\ 8} = 20 \\ 1 & X_{1\ 9} = 26 & X_{2\ 9} = 21 \\ 1 & X_{1\ 10} = 32 & X_{2\ 10} = 24 \end{pmatrix};$$

$$B_{(k+1=3,1)} = \begin{pmatrix} B_0 \\ B_1 \\ B_2 \end{pmatrix}; \quad \varepsilon_{(n=10,1)} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_{n=10} \end{pmatrix}.$$

2. Nous allons estimer le vecteur des paramètres B par la méthode Moindres carrés ordinaires (MCO), en supposant que : $Y = X B + \varepsilon$.

L'estimer du vecteur des paramètres B est donné par :

$$\hat{B} = (X'X)^{-1}X'Y. \quad (2)$$

Afin de caculer les estimateurs des coefficients de régression, nous devons calculer dans un premier temps $(X'X)^{-1}$.

$$(X'X) = \begin{pmatrix} n & \sum X_{1t} & \sum X_{2t} \\ 10 & 180 & 120 \\ \sum X_{1t} & \sum X_{1t}^2 & \sum X_{1t}X_{2t} \\ 180 & 3816 & 2684 \\ \sum X_{2t} & \sum X_{2t}X_{1t} & \sum X_{2t}^2 \\ 120 & 2684 & 1944 \end{pmatrix}$$

Où, les résultats utilisés sont donnés dans le tableau suivant :

n	Y_t	X_{1t}	X_{2t}	$X_{1t}X_{2t}$	X_{1t}^2	X_{2t}^2	$X_{1t}Y_t$	$X_{2t}Y_t$	Y_t^2
1	40	6	4	24	36	16	240	160	1600
2	44	10	4	40	100	16	440	176	1936
3	46	12	5	60	144	25	552	230	2116
4	48	14	7	98	196	49	672	336	2304
5	52	16	9	144	256	81	832	468	2704
6	58	18	12	216	324	144	1044	696	3364
7	60	22	14	308	484	196	1320	840	3600
8	68	24	20	480	576	400	1632	1630	4624
9	74	26	21	546	676	441	1924	1554	5476
10	80	32	24	786	1024	676	2560	1920	6400
Σ	570	180	120	2684	3816	1944	11216	7740	34124

$$\det(X'X) = \begin{vmatrix} 10 & 180 & 120 \\ 180 & 3816 & 2684 \\ 120 & 2684 & 1944 \end{vmatrix}$$

$$\det(X'X) = 10 \begin{vmatrix} 3816 & 2684 \\ 2684 & 1944 \end{vmatrix} - 180 \begin{vmatrix} 180 & 120 \\ 2684 & 1944 \end{vmatrix} + 120 \begin{vmatrix} 180 & 120 \\ 3816 & 2684 \end{vmatrix} = 157280$$

La comatrice de $(X'X)$

$$C_{(X'X)} = \begin{pmatrix} + \begin{vmatrix} 3816 & 2684 \\ 2684 & 1944 \end{vmatrix} & - \begin{vmatrix} 180 & 120 \\ 2684 & 1944 \end{vmatrix} & + \begin{vmatrix} 180 & 120 \\ 3816 & 2684 \end{vmatrix} \\ - \begin{vmatrix} 180 & 120 \\ 2684 & 1944 \end{vmatrix} & + \begin{vmatrix} 10 & 120 \\ 120 & 1944 \end{vmatrix} & - \begin{vmatrix} 10 & 180 \\ 120 & 2624 \end{vmatrix} \\ + \begin{vmatrix} 180 & 120 \\ 3816 & 2684 \end{vmatrix} & - \begin{vmatrix} 10 & 120 \\ 180 & 2684 \end{vmatrix} & + \begin{vmatrix} 10 & 180 \\ 180 & 3816 \end{vmatrix} \end{pmatrix}$$

$$C_{(X'X)} = \begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{C'_{(X'X)}}{\det(X'X)} = \frac{1}{157280} \begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix}$$

Et par conséquent

$$\hat{B} = \frac{C'_{(X'X)}}{\det(X'X)}(X'Y) = \frac{1}{157280} \begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix} \begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix}$$

Avec

$$(X'Y) = \begin{pmatrix} \sum Y_t = 570 \\ \sum X_{1t} Y_t = 11216 \\ \sum X_{2t} Y_t = 7740 \end{pmatrix}$$

$$\hat{B} = \frac{1}{157280} \begin{pmatrix} 5029920 \\ 102240 \\ 174560 \end{pmatrix} = \begin{pmatrix} \frac{5029920}{157280} \\ \frac{102240}{157280} \\ \frac{174560}{157280} \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} 31.98 \\ 0.65 \\ 1.11 \end{pmatrix} \quad (3)$$

Donc, le modèle s'écrit comme suit :

$$\widehat{Y}_t = 31.98 + 0.65X_{1t} + 1.11X_{2t}, \quad t = \overline{1, 10} \quad (4)$$

3. Pour le calcul de la variance des paramètres du modèle, on doit calculer la matrice variance covariance $\widehat{\Omega}_{\widehat{B}}$

$$\widehat{\Omega}_{\widehat{B}} = \widehat{\sigma}_\varepsilon^2 (X'X)^{-1} \quad (5)$$

Où

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n e_i^2}{n - (k + 1)}$$

Avec

$$\sum_{t=1}^n e_i^2 = \sum_{t=1}^n Y_t^2 - \widehat{B}' (X'Y) = 34124 - (30.98, 0.65, 1.11) \begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix} = 34124 - 34110.4 = 13.6$$

Et on déduit que :

$$\widehat{\sigma}_\varepsilon^2 = \frac{13.6}{10 - 7} = 1.94 \Rightarrow \widehat{\sigma}_\varepsilon = 1.39$$

La variance des paramètres du modèle se trouve sur la diagonale de la matrice variance covariance $\widehat{\Omega}_{\widehat{B}}$, c'est à dire :

$$\widehat{\sigma}_{\widehat{B}_0}^2 = 1.94 \times 1.363 = 2.64 \Rightarrow \widehat{\sigma}_{\widehat{B}_0} = 1.63$$

$$\widehat{\sigma}_{\widehat{B}_1}^2 = 1.94 \times 0.032 = 0.062 \Rightarrow \widehat{\sigma}_{\widehat{B}_1} = 0.25$$

$$\widehat{\sigma}_{\widehat{B}_2}^2 = 1.94 \times 0.036 = 0.07 \Rightarrow \widehat{\sigma}_{\widehat{B}_2} = 0.26$$

4. ♣ Le calcul du coefficient de détermination R^2

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{68.47737}{110.960000} = 0.617136. \quad (6)$$

♣ Le calcul du SCT

$$SCT = \sum_{t=1}^n Y_t^2 - n\bar{Y}^2 = 34124 - 10(57)^2 = 1634$$

Alors

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{13.6}{1634} = 0.99$$

♣ Le calcul du coefficient de détermination corrigé $\bar{R}^2$

$$\bar{R}^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2) = 1 - \frac{10-1}{10-3}(1-0.99) = 0.99 \quad (7)$$

D'après le résultat de R^2 et $\bar{R}^2$, on conclut que la qualité de l'ajustement est très importante.

5. Le test global

$H_0 : B_1 = B_2 = \dots = B_k = 0$, aucune variable n'explique la variable Y

$H_1 : \exists i \in \overline{1, k} / B_i \neq 0$

$$F_c = \frac{SCE/k}{SCR/(n-(k+1))} = \frac{R^2/k}{(1-R^2)/(n-(k+1))} \quad (8)$$

$$F_c = \frac{0.99/2}{(1-0.99)/(10-3)} = 346.5$$

On accepte H_1 , car $F_c = 346.5 > F_{\frac{\alpha}{2}}(k, n-(k+1)) = F_{0.05}(2, 7) = 4.74$, ce qui signifie que le modèle est globalement significatif au seuil de 5%

6. On utilise le test de Student pour la significativité de B_i , qui est définie par :

$H_0 : B_i = 0$ contre $H_1 : B_i \neq 0$

Pour cela, on doit calculer $t_{c B_i}$

$$t_{c B_i} = \left| \frac{\widehat{B}_i - B_i}{\widehat{\sigma}_{\widehat{B}_i}} \right| \quad (9)$$

On obtient alors les résultats suivants :

$$t_{c_{B_0}} = \frac{30.98 - 0}{1.63} = 19.62$$

$$t_{c_{B_1}} = \frac{0.65 - 0}{0.25} = 2.62$$

$$t_{c_{B_2}} = \frac{1.11 - 0}{0.26} = 4.27$$

On constate d'après la valeur de la Statistique de Student calculer pour chaque B_i qu'elle est supérieure à la valeur de la statistique de Student ($t_T(7, 0.025) = 2.365$) lu sur la table avec 7 est son degré de liberté et 0.025 est le seuil de confiance. On conclut donc que tous les paramètres sont significatifs et que les deux variables sont significatives. Ce qui nous permis de dire que ce modèle est validé.

7. Le calcul de la prévision à la date 11 sachant que $X_{1 \ 11} = 36$ et $X_{2 \ 11} = 27$

Du moment le modèle est validé, on peut donc calculer la prévision à la date 11 .

♣ La valeur ponctuelle ajustée de la prévision est donné par :

$$\hat{Y}_{n+h} = \hat{B}_0 + \hat{B}_1 X_{1 \ n+h} + \hat{B}_2 X_{2 \ n+h} + \dots \hat{B}_k X_{k \ n+h} \quad (10)$$

Donc

$$\hat{Y}_{11} = \hat{B}_0 + \hat{B}_1 X_{1 \ 11} + \hat{B}_2 X_{2 \ 11} \quad (11)$$

Donc

$$\hat{Y}_{11} = 30.98 + 0.65 X_{1 \ 11} + 1.11 X_{2 \ 11} = 30.98 + 0.65(36) + 1.11(27) = 85.35$$

♣ L'intervalle de prédiction est défini par :

$$Y_{11} = \hat{Y}_{11} \pm t_{0.025}(7) \sigma_{e_{n+h}} \quad (12)$$

Et $\sigma_{e_{n+h}}^2$ est la variance de l'erreur de prévision qui est donnée par :

$$\sigma_{e_{n+h}}^2 = \text{Var}(e_{n+h}) = \hat{\sigma}_\varepsilon^2 \left[X'_{n+h} (X'X)^{-1} X_{n+h} + 1 \right] \quad (13)$$

$$\sigma_{e_{n+h}}^2 = \hat{\sigma}_\varepsilon^2 \left[X'_{11} (X'X)^{-1} X_{11} + 1 \right]$$

Avec $X'_{11} = (1, 36, 27)$. Alors

$$\sigma_{e_{n+h}}^2 = 1.94 \left[\frac{1}{157280} (1, 36, 27) \begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix} \begin{pmatrix} 1 \\ 36 \\ 27 \end{pmatrix} + 1 \right]$$

$$\sigma_{e_{n+h}}^2 = 1.94 \left[\frac{1}{157280} (1, 36, 27) \begin{pmatrix} -107392 \\ 12120 \\ -7920 \end{pmatrix} + 1 \right] = 1.94 [0.823 + 1] = 3.537$$

♣ L'intervalle de prédiction est défini par :

$$Y_{11} = 85.35 \pm 2.62(1.881) = [81.02 ; 89.68]. \quad (14)$$

8. Le tableau de l'analyse de la variance

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1, X_2	$SCE = SCT - SCR$ $= 1634 - 13.6 = 1620.4$	2	$\frac{SCE}{2} = \frac{1620.4}{2} = 810.2$
Résiduelle	$SCR = 13.6$	$n - 3 = 10 - 3 = 7$	$\frac{SCR}{7} = \frac{13.6}{7} = 1.94$
Totale	$SCT = 1634$	$n - 1 = 10 - 1 = 9$	

Corrigé 2.

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

$$y_t = B_0 + B_1 x_{1t} + B_2 x_{2t} + \varepsilon_t, \quad i = \overline{1, 9} \quad (15)$$

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

$$\hat{Y}_t = \hat{B}_0 + \hat{B}_1 X_{1t} + \hat{B}_2 X_{2t} \quad (16)$$

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

On sait que :

$$t_{cB_0} = \frac{\hat{B}_0 - 0}{\hat{\sigma}_{\hat{B}_0}} \quad (17)$$

Donc :

$$\hat{\sigma}_{\hat{B}_0} = \frac{\hat{B}_0 - 0}{t_{cB_0}} = \frac{-437.136}{-7.555779} = 57.93097 \quad (18)$$

- ♣ 2 $\rightarrow \hat{B}_1$,

On sait que :

$$t_{cB_1} = \frac{\hat{B}_1 - 0}{\hat{\sigma}_{\hat{B}_1}} \quad (19)$$

Alors,

$$\hat{B}_1 = \hat{\sigma}_{\hat{B}_1} t_{cB_1} = 0.089665(3.753215) = 0.336530 \quad (20)$$

- ♣ 3 $\rightarrow t_{cB_2}$,

$$t_{cB_2} = \frac{\hat{B}_2 - 0}{\hat{\sigma}_{\hat{B}_2}} = \frac{0.410016}{0.196143} = 2.090387 \quad (21)$$

- ♣ 4 $\rightarrow \overline{R}^2$, Le calcul du coefficient de détermination corrigé

$$\overline{R}^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2) = 1 - \frac{9-1}{9-3}(1-0.978404) = 0.971206 \quad (22)$$

- ♣ 5 $\rightarrow SCR = \sum_{t=1}^n e_t^2$,

on sait que :

$$\hat{\sigma}_\varepsilon^2 = \frac{SCR}{n-k-1} \quad (23)$$

et par conséquent :

$$SCR = \text{Sum squared resid} = \sum_{t=1}^n e_t^2 = (n-k-1)\hat{\sigma}_\varepsilon^2 = (9-2-1)(23.07151)^2 = 3193.767 \quad (24)$$

où $\hat{\sigma}_\varepsilon = S.E \text{ of regression} = 23.07151$

♣ $6 \rightarrow F - statistic = F_c,$

sa valeur est donnée par la formule suivante :

$$F_c = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} = \frac{0.978404/2}{(1 - 0.978404)/(9 - 3)} = 135.9164 \quad (25)$$

3. D'après la probabilité $Prob(F - statistic) = 0.000010 < 0.05$, alors ce modèle est globalement significatif au seuil 5%.

4. On utilise le test de Student pour la significativité de B_i , qui est définie par :

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

D'après la valeur de la probabilité de la Statistique de Student calculer pour chaque B_0 et B_1 qu'elles sont inférieure à 5%, ce que veut dire que les deux paramètres sont significatifs au seuil 5% et que B_2 n'est pas significatif, car la probabilité de la Statistique de Student calculer de B_2 est supérieure à 5%.

Ce qui nous a permis de dire que ce modèle contient une constante et que la variable X_1 explique bien le modèle , mais ce n'est pas le cas de la variable X_2 . Pour cela, ce modèle n'est pas validé.

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

♣ On sait que : $R^2 = 0.978404$, $SCR = \sum_{t=1}^n e_t^2 = 3193.767$ et que $R^2 = 1 - \frac{SCR}{SCT}$.

Alors

$$SCT = \frac{SCR}{(1 - R^2)} = \frac{3193.767}{(1 - 0.978404)} = 147886.97 \quad (26)$$

♣ On a $SCT = SCE + SCR$, donc

$$SCE = SCT - SCR = 147886.97 - 3193.767 = 144693.203 \quad (27)$$

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1	$SCE = 144693.203$	2	$\frac{SCE}{2} = 72341.601$
Résiduelle	$SCR = 3193.767$	$n - 3 = 9 - 3 = 6$	$\frac{SCR}{n-3} = \frac{3193.767}{6} = 532.2945$
Totale	$SCT = 147886.97$	$n - 1 = 9 - 1 = 8$	

Corrigé 3

$$Y = \begin{pmatrix} 5 \\ 4 \\ 6 \\ 6 \\ 9 \\ 9 \end{pmatrix}, X = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 7 \\ 1 & 9 \end{pmatrix}, B = \begin{pmatrix} B_0 \\ B_1 \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

- b) En utilisant la forme matricielle, estimer les paramètres du modèle par la méthode des moindres carrés ordinaires.

$$\hat{B} = (X'X)^{-1}(X'Y)$$

$$X'X = \begin{pmatrix} 6 & 33 \\ 33 & 217 \end{pmatrix}, X'Y = \begin{pmatrix} 39 \\ 238 \end{pmatrix} \text{ (1 pts).}$$

$$\det(X'X) = 213 \neq 0 \text{ (1 pts).}$$

$$\hat{B} = \frac{1}{213} \begin{pmatrix} 609 \\ 141 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} 2.859 \\ 0.662 \end{pmatrix} \text{ (2 pts).}$$

- c) Comparer les résultats obtenus par les deux méthodes et que penser vous ?

D'après les deux méthodes, nous avons obtenu les mêmes résultats. Mais de préférence utilisé la régression simple c'est mieux, car elle est plus simple et rapide. **(1 pts)**.

- 1) Quelle est la qualité de cet ajustement ? commenter.

On calcule le R^2 : **(1.5 pts)**.

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

$$SCT = \sum Y_i^2 - n \bar{Y}^2 = 275 - 6(6.5)^2$$

$$SCT = 21.5$$

$$SCE = \hat{B}^2 \left[\sum X_i^2 - n \bar{X}^2 \right] = 0.662^2 \quad (35.5)$$

$$SCE = 15.558$$

$$R^2 = \frac{15.558}{21.5} = 0.724$$

Ce modèle est intéressant.

- 1) Ce modèle est-il globalement significatif au seuil de 2%?(1 pts).

$$F_C = \frac{SCE/1}{SCR/(n-1)} = \frac{15.558}{5.942/4} = 10.473$$

$F_C = 10.473 < F_T(1, 4) = 21.20$. Donc ce modèle n'est pas globalement significatif au seuil de 2%.

- 2) Est-ce que le paramètre β de ce modèle est significativement différent de (-1) au seuil 5%?(1.5 pts).

$$H_0 : B = -1 \quad \text{contre} \quad H_1 : B \neq -1$$

$$t_{CB} = \left| \frac{\hat{B} - B}{\hat{\sigma}_{\hat{B}}} \right|, \text{ SOUS } H_0, \text{ on a :}$$

$$t_{CB} = \left| \frac{0.662 + 1}{0.205} \right| = 8.110$$

Avec

$$\hat{\sigma}_{\hat{B}}^2 = \frac{\hat{\sigma}_{\varepsilon}^2}{\sum X_i^2 - n \bar{X}^2} = \frac{1.482}{35.5} = 0.042 \quad \text{et} \quad \hat{\sigma}_{\hat{B}} = 0.205$$

Où

$$\hat{\sigma}_{\varepsilon}^2 = \frac{SCR}{n - 2} = \frac{5.942}{4} = 1.486$$

$t_{CB} = 8.110 > t_T(0.025, 4) = 2.776$. Donc, on accepte H_1 , c'est-à-dire le paramètre B de ce modèle est significativement différent de (-1) au seuil 5%.

1) Donner un intervalle de confiance pour α au seuil 5%? Conclusion ? (1.5 pts).

$$\alpha = \hat{\alpha} \pm \hat{\sigma}_{\hat{\alpha}} t_T(0.025, 4) = 2.859 \pm 2.776 (1.230) = [-0.555; 6.273]$$

Où

$$\hat{\sigma}_{\hat{\alpha}}^2 = \hat{\sigma}_{\varepsilon}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum X_i^2 - n \bar{X}^2} \right] = \hat{\sigma}_{\varepsilon}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum X_i^2 - n \bar{X}^2} \right]$$