CORRECTION DU TD : MÉTHODES ÉCONOMÉTRIQUES

FPN : FILIÈRE : SEG S6

ANNÉE UNIVERSITAIRE : 2019-2020

Exercice 1

Une étude porte sur la relation entre les frais fournés dans un projet commercial et le volume de gains qu'il réalise. On a recueilli au cours des dix derniers mois les données suivantes :

Frais fournés X	4	2	2.5	2	3	1	2.5	5.5	3.5	4.5
Volume des gains Y	13.5	10	8	7.5	12	6	5.5	13.5	11.5	10.5

Nous nous référons au modèle linéaire : $y_i = a_0 + a_1 x_i + \varepsilon_i, \quad i = 1, \dots, 10.$

1. Calculer $\widehat{a}_1$ et $\widehat{a}_0$ et déduire la série des résidus :

Exercice 1

Une étude porte sur la relation entre les frais fournés dans un projet commercial et le volume de gains qu'il réalise. On a recueilli au cours des dix derniers mois les données suivantes :

Frais fournés X	4	2	2.5	2	3	1	2.5	5.5	3.5	4.5
Volume des gains Y	13.5	10	8	7.5	12	6	5.5	13.5	11.5	10.5

Nous nous référons au modèle linéaire : $y_i = a_0 + a_1 x_i + \varepsilon_i, \quad i = 1, \ldots, 10$.

1. Calculer $\widehat{a}_1$ et $\widehat{a}_0$ et déduire la série des résidus :

On a

$$\widehat{a}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 1.701,$$

et

$$\widehat{a}_0 = \overline{y} - \widehat{a}_1 \overline{x} = 4.611.$$

L'équation de la droite de régression est alors

$$y = 1.701x + 4.611.$$

Exercice 1

Une étude porte sur la relation entre les frais fournés dans un projet commercial et le volume de gains qu'il réalise. On a recueilli au cours des dix derniers mois les données suivantes :

Frais fournés X	4	2	2.5	2	3	1	2.5	5.5	3.5	4.5
Volume des gains Y	13.5	10	8	7.5	12	6	5.5	13.5	11.5	10.5

Nous nous référons au modèle linéaire : $y_i = a_0 + a_1 x_i + \varepsilon_i, \quad i = 1, \ldots, 10.$

1. Calculer $\widehat{a}_1$ et $\widehat{a}_0$ et déduire la série des résidus :

On a

$$\widehat{a}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 1.701,$$

et

$$\widehat{a}_0 = \overline{y} - \widehat{a}_1 \overline{x} = 4.611.$$

L'équation de la droite de régression est alors

$$y = 1.701x + 4.611.$$

On calcul alors les valeurs estimées

$$\widehat{y}_i = 1.701x_i + 4.611,\ i = 1, \ldots, 10,$$

puis la série des résidus $(y_i - \widehat{y}_i)_{1 \leq i \leq 10}$.

Les résultats sont donnés dans le tableau suivant (avec $\overline{y} = 9.8$) :

x_i	y_i	$(y_i - \overline{y})^2$	$\widehat{y}_i$	$(\widehat{y}_i - \overline{y})^2$	$(y_i - \widehat{y}_i)^2$
4	13.5	13.689	11.416	2.611	4.342
2	10	0.039	8.013	3.190	3.944
2.5	8	3.240	8.864	0.875	0.747
2	7.5	5.290	8.013	3.190	0.264
3	12	4.839	9.714	0.007	5.221
1	6	14.440	6.312	12.160	0.097
2.5	5.5	18.490	8.864	0.875	11.319
5.5	13.5	13.689	13.967	17.369	0.218
3.5	11.5	2.889	10.565	0.585	0.873
4.5	10.5	0.489	12.266	6.083	3.120
Total		77.1		46.949	30.150

2. Estimer la variance résiduelle et les écarts type de $\widehat{a}_1$ et $\widehat{a}_0$:

La variance résiduelle est

$$\widehat{\sigma}_{\varepsilon}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-2} = \frac{30.150}{8} = 3.768.$$

2. Estimer la variance résiduelle et les écarts type de $\widehat{a}_1$ et $\widehat{a}_0$:

La variance résiduelle est

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-2} = \frac{30.150}{8} = 3.768.$$

Les écarts type de $\widehat{a}_1$ et $\widehat{a}_0$ sont donnés par :

$$\widehat{\sigma}_{\widehat{a}_1}^2 = \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\widehat{\sigma}_\varepsilon^2}{n\sigma_x^2}$$

$$= \frac{3.768}{10 \times 1.622} = 0.232 \Rightarrow \widehat{\sigma}_{\widehat{a}_1} = 0.481.$$

$$\widehat{\sigma}_{\widehat{a}_0}^2 = \widehat{\sigma}_\varepsilon^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) = \widehat{\sigma}_\varepsilon^2\left(\frac{1}{n} + \frac{\overline{x}^2}{n\sigma_x^2}\right)$$

$$= 3.768 \times \left(\frac{1}{10} + \frac{3.05^2}{10 \times 1.622}\right) = 2.537 \Rightarrow \widehat{\sigma}_{\widehat{a}_0} = 1.593.$$

3. Calculer r_{xy} le coefficient de corrélation linéaire. Est-il significativement différent de 0 ?

3. Calculer r_{xy} le coefficient de corrélation linéaire. Est-il significativement différent de 0 ? On a

$$r_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.780.$$

Soit à tester l'hypothèse H_0 : "$r_{x,y} = 0$" contre l'hypothèse H_1 : "$r_{x,y} \neq 0$".

3. Calculer r_{xy} le coefficient de corrélation linéaire. Est-il significativement différent de 0 **?** On a

$$r_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.780.$$

Soit à tester l'hypothèse H_0 : " $r_{x,y} = 0$" contre l'hypothèse H_1 : " $r_{x,y} \neq 0$".
Nous savons que

$$\frac{\rho_{x,y}}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}} \rightsquigarrow \mathcal{T}_{n-2}$$

3. Calculer r_{xy} le coefficient de corrélation linéaire. Est-il significativement différent de 0 ? On a

$$r_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.780.$$

Soit à tester l'hypothèse H_0 : " $r_{x,y} = 0$" contre l'hypothèse H_1 : " $r_{x,y} \neq 0$".
Nous savons que

$$\frac{\rho_{x,y}}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}} \rightsquigarrow \mathcal{T}_{n-2}$$

Nous calculons alors la statistique de Student :

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}} = \frac{0.780}{\sqrt{\frac{1-(0.780)^2}{8}}} = 3.525.$$

3. Calculer r_{xy} le coefficient de corrélation linéaire. Est-il significativement différent de 0 ? On a

$$r_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.780.$$

Soit à tester l'hypothèse H_0 : " $r_{x,y} = 0$" contre l'hypothèse H_1 : " $r_{x,y} \neq 0$".
Nous savons que

$$\frac{\rho_{x,y}}{\sqrt{\frac{1-\rho^2_{x,y}}{n-2}}} \rightsquigarrow \mathcal{T}_{n-2}$$

Nous calculons alors la statistique de Student :

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{1-\rho^2_{x,y}}{n-2}}} = \frac{0.780}{\sqrt{\frac{1-(0.780)^2}{8}}} = 3.525.$$

Puisque $t^* > t_8^{0.025} = 2.306$ valeur lue dans une table de Student au seuil $\alpha = 5\%$ à $n-2 = 8$ degrés de liberté, nous rejetons l'hypothèse H_0.

3. Calculer r_{xy} le coefficient de corrélation linéaire. Est-il significativement différent de 0 **?** On a

$$r_{x,y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.780.$$

Soit à tester l'hypothèse H_0 : " $r_{x,y} = 0$" contre l'hypothèse H_1 : " $r_{x,y} \neq 0$".
Nous savons que

$$\frac{\rho_{x,y}}{\sqrt{\frac{1-\rho^2_{x,y}}{n-2}}} \rightsquigarrow \mathcal{T}_{n-2}$$

Nous calculons alors la statistique de Student :

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{1-\rho^2_{x,y}}{n-2}}} = \frac{0.780}{\sqrt{\frac{1-(0.780)^2}{8}}} = 3.525.$$

Puisque $t^* > t_8^{0.025} = 2.306$ valeur lue dans une table de Student au seuil $\alpha = 5\%$ à $n-2 = 8$ degrés de liberté, nous rejetons l'hypothèse H_0.
Le coefficient de corrélation est donc significativement différent de 0.

4. Tester à un seuil de 5% l'hypothèse H_0 : "$a_1 = 0$" contre H_1 : "$a_1 \neq 0$". Commenter. Donner un intervalle de confiance pour a_1 à 95% :

4. Tester à un seuil de 5% l'hypothèse $H_0 : '' a_1 = 0''$ contre $H_1 : '' a_1 \neq 0''$. Commenter. Donner un intervalle de confiance pour a_1 à 95% :

Nous Testons à un seuil de 5% l'hypothèse

$$H_0 : '' a_1 = 0'' \text{contre} \quad H_1 : '' a_1 \neq 0''.$$

Le ratio de Student empirique est $t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{1.701}{0.481} = 3.536$ est supérieure à $t_8^{0.025} = 2.306$.

4. Tester à un seuil de 5% l'hypothèse H_0 : " $a_1 = 0$" contre H_1 : " $a_1 \neq 0$". Commenter. Donner un intervalle de confiance pour a_1 à 95% :

Nous Testons à un seuil de 5% l'hypothèse

$$H_0 : \text{"} a_1 = 0\text{" contre} \quad H_1 : \text{"} a_1 \neq 0\text{"}.$$

Le ratio de Student empirique est $t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{1.701}{0.481} = 3.536$ est supérieure à $t_8^{0.025} = 2.306$.

Donc, a_1 est significativement différents de 0.

La variable exogène x contribue bien à expliquer la variable endogène y.

4. Tester à un seuil de 5% l'hypothèse H_0 : "$a_1 = 0$" contre H_1 : "$a_1 \neq 0$". Commenter. Donner un intervalle de confiance pour a_1 à 95% :
Nous Testons à un seuil de 5% l'hypothèse

$$H_0 : "a_1 = 0" \text{ contre } \quad H_1 : "a_1 \neq 0".$$

Le ratio de Student empirique est $t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{1.701}{0.481} = 3.536$ est supérieure à $t_8^{0.025} = 2.306$.
Donc, a_1 est significativement différents de 0.
La variable exogène x contribue bien à expliquer la variable endogène y.
Un intervalle de confiance pour a_1 à 95% est

$$\left[\widehat{a}_1 - \widehat{\sigma}_{\widehat{a}_1} \times t_{n-2}^{\alpha/2}; \widehat{a}_1 + \widehat{\sigma}_{\widehat{a}_1} \times t_{n-2}^{\alpha/2}\right]$$
$$= [1.701 - 0.481 \times 2.306; 1.701 + 0.481 \times 2.306] = [0.591; 2.81].$$

5. Tester à un seuil de 5% l'hypothèse H_0 : "$a_1 = -0.5$" contre H_1 : "$a_1 \neq -0.5$" :

5. Tester à un seuil de 5% l'hypothèse H_0 : " $a_1 = -0.5$" contre H_1 : " $a_1 \neq -0.5$" :

Pour l'hypothèse bilatérale H_0 : " $a_1 = -0.5$" contre H_1 : " $a_1 \neq -0.5$".

Le ratio de Student empirique est $t^* = \dfrac{|\widehat{a}_1 - a_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \dfrac{|1.701 - (-0.5)|}{0.481} = 4.575$

est supérieure à $t_8^{0.025} = 2.306$.
Donc, a_1 est significativement différents de -0.5.

5. Tester à un seuil de 5% l'hypothèse $H_0 : "a_1 = -0.5"$ contre $H_1 : "a_1 \neq -0.5"$:

Pour l'hypothèse bilatérale $H_0 : "a_1 = -0.5"$ contre $H_1 : "a_1 \neq -0.5"$.

Le ratio de Student empirique est $t^* = \dfrac{|\widehat{a}_1 - a_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \dfrac{|1.701 - (-0.5)|}{0.481} = 4.575$

est supérieure à $t_8^{0.025} = 2.306$.

Donc, a_1 est significativement différents de -0.5.

6. Calculer les sommes SCR, SCT et SCE puis R^2 le coefficient de détermination. Conclure :

5. Tester à un seuil de 5% l'hypothèse $H_0 : "a_1 = -0.5"$ contre $H_1 : "a_1 \neq -0.5"$:

Pour l'hypothèse bilatérale $H_0 : "a_1 = -0.5"$ contre $H_1 : "a_1 \neq -0.5"$.

Le ratio de Student empirique est $t^* = \dfrac{|\widehat{a}_1 - a_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \dfrac{|1.701 - (-0.5)|}{0.481} = 4.575$

est supérieure à $t_8^{0.025} = 2.306$.
Donc, a_1 est significativement différents de -0.5.

6. Calculer les sommes SCR, SCT et SCE puis R^2 le coefficient de détermination. Conclure : Les sommes sont calculées dans le tableau ci-dessus :

$$SCR = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = 30.150, \quad SCT = \sum_{i=1}^{n} (y_i - \overline{y})^2 = 77.1,$$

$$\text{et} \quad SCE = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 = 46.949.$$

5. Tester à un seuil de 5% l'hypothèse H_0 : "$a_1 = -0.5$" contre H_1 : "$a_1 \neq -0.5$" :

Pour l'hypothèse bilatérale H_0 : "$a_1 = -0.5$" contre H_1 : "$a_1 \neq -0.5$".

Le ratio de Student empirique est $t^* = \dfrac{|\widehat{a}_1 - a_1|}{\widehat{\sigma}_{\widehat{a}_1}} = \dfrac{|1.701 - (-0.5)|}{0.481} = 4.575$

est supérieure à $t_8^{0.025} = 2.306$.

Donc, a_1 est significativement différents de -0.5.

6. Calculer les sommes SCR, SCT et SCE puis R^2 le coefficient de détermination. Conclure : Les sommes sont calculées dans le tableau ci-dessus :

$$SCR = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = 30.150, \quad SCT = \sum_{i=1}^{n}(y_i - \overline{y})^2 = 77.1,$$

$$\text{et} \quad SCE = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 = 46.949.$$

Le coefficient de détermination est $R^2 = 1 - \dfrac{SCR}{SCT} = 1 - \dfrac{30.150}{77.1} = 0.609.$

On a $R^2 > 0.5$. On conclut alors que l'ajustement linéaire est bon.

7. Tracer le tableau d'analyse de la variance et faire un test de Fisher à un seuil de 5% :

7. Tracer le tableau d'analyse de la variance et faire un test de Fisher à un seuil de 5% :

Le tableau d'analyse de la variance :

Variation	Somme des carrés	Degré de liberté	Carrés moyens
x	$SCE = 46.949$	1	$\frac{SCE}{1} = 46.949$
Résidu	$SCR = 30.150$	8	$\frac{SCR}{8} = 3.768$
Total	$SCT = 77.1$	9	

Soit le test d'hypothèse H_0 : "$SCE = 0$" contre H_1 : "$SCE \neq 0$".

7. Tracer le tableau d'analyse de la variance et faire un test de Fisher à un seuil de 5% :

Le tableau d'analyse de la variance :

Variation	Somme des carrés	Degré de liberté	Carrés moyens
x	$SCE = 46.949$	1	$\frac{SCE}{1} = 46.949$
Résidu	$SCR = 30.150$	8	$\frac{SCR}{8} = 3.768$
Total	$SCT = 77.1$	9	

Soit le test d'hypothèse H_0 : " $SCE = 0$" contre H_1 : " $SCE \neq 0$".

La statistique de ce test est donnée par :

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{46.949}{3.768} = 12.159.$$

7. Tracer le tableau d'analyse de la variance et faire un test de Fisher à un seuil de 5% :

Le tableau d'analyse de la variance :

Variation	Somme des carrés	Degré de liberté	Carrés moyens
x	$SCE = 46.949$	1	$\frac{SCE}{1} = 46.949$
Résidu	$SCR = 30.150$	8	$\frac{SCR}{8} = 3.768$
Total	$SCT = 77.1$	9	

Soit le test d'hypothèse H_0 : " $SCE = 0$" contre H_1 : " $SCE \neq 0$".

La statistique de ce test est donnée par :

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{46.949}{3.768} = 12.159.$$

Puisque $F^{\alpha}_{1;n-2} = F^{0.05}_{1;8} = 5.32$, alors $F^* > F^{\alpha}_{1;n-2}$.

Donc, nous rejetons au seuil α l'hypothèse H_0 et donc la variable explicative est significative.

8. Déterminer au seuil 5%, un intervalle de confiance pour y_{11} relatif à la valeur prévisible $x_{11} = 5mDHs$:

8. Déterminer au seuil 5%, un intervalle de confiance pour y_{11} relatif à la valeur prévisible $x_{11} = 5mDHs$:

On a l'intervalle de prédiction $I_{y_{n+1}}$ relatif à la valeur prévisible x_{n+1} est donné par :

$$y_{n+1} = \widehat{y}_{n+1} \pm t_{n-2}^{\alpha/2} \widehat{\sigma}_{\varepsilon} \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} + 1}$$

où

$$x_{n+1} = 5mDH, \quad \overline{x} = 3.05, \quad \sum_{i=1}^{n} (x_i - \overline{x})^2 = 16.255, \quad \widehat{y}_{n+1} = 13.117$$

$$t_{n-2}^{\alpha/2} = t_8^{0.025} = 2.306, \quad \widehat{\sigma}_{\varepsilon}^2 = \frac{SCR}{n-2} = 3.768.$$

Donc,

$$I_{y_{11}} = [7.947; 18.286].$$

Exercice 2 :

Soient les résultats d'une estimation économétrique :

$$\widehat{y}_t = -4,2 + 1,6x_t, \quad t = 1,\ldots,10, \quad R^2 = 0,84, \quad \widehat{\sigma}_\varepsilon = 8,24.$$

1. Calculer les statistiques suivantes :

(a) SCR**, la somme des carrés des résidus :**

Exercice 2 :

Soient les résultats d'une estimation économétrique :

$$\widehat{y}_t = -4,2 + 1,6x_t, \quad t = 1,\dots,10, \quad R^2 = 0,84, \quad \widehat{\sigma}_\varepsilon = 8,24.$$

1. Calculer les statistiques suivantes :

(a) *SCR*, **la somme des carrés des résidus :**

$$\widehat{\sigma}_\varepsilon^2 = \frac{SCR}{n-2} \Rightarrow SCR = (n-2)\widehat{\sigma}_\varepsilon^2 = 8 \times 8,24^2 = 543,18.$$

Exercice 2 :
Soient les résultats d'une estimation économétrique :

$$\widehat{y}_t = -4,2 + 1,6x_t, \quad t = 1, \ldots, 10, \quad R^2 = 0,84, \quad \widehat{\sigma}_\varepsilon = 8,24.$$

1. Calculer les statistiques suivantes :

(a) *SCR*, **la somme des carrés des résidus :**

$$\widehat{\sigma}^2_\varepsilon = \frac{SCR}{n-2} \Rightarrow SCR = (n-2)\widehat{\sigma}^2_\varepsilon = 8 \times 8,24^2 = 543,18.$$

(b) *SCT*, **la somme des carrés totaux :**

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \Rightarrow SCT = \frac{SCR}{1-R^2} = \frac{543,18}{1-0,84} = 3394,88.$$

Exercice 2 :

Soient les résultats d'une estimation économétrique :

$$\widehat{y}_t = -4,2 + 1,6x_t, \quad t = 1,\ldots,10, \quad R^2 = 0,84, \quad \widehat{\sigma}_\varepsilon = 8,24.$$

1. Calculer les statistiques suivantes :

(a) SCR**, la somme des carrés des résidus :**

$$\widehat{\sigma}^2_\varepsilon = \frac{SCR}{n-2} \Rightarrow SCR = (n-2)\widehat{\sigma}^2_\varepsilon = 8 \times 8,24^2 = 543,18.$$

(b) SCT**, la somme des carrés totaux :**

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \Rightarrow SCT = \frac{SCR}{1-R^2} = \frac{543,18}{1-0,84} = 3394,88.$$

(c) SCE**, la somme des carrés expliqués :**

$$SCT = SCE + SCR \Rightarrow SCE = SCT - SCR = 3394,88 - 543,18 = 2851,70.$$

Exercice 2 :

Soient les résultats d'une estimation économétrique :

$$\widehat{y}_t = -4,2 + 1,6x_t, \quad t = 1,\ldots,10, \quad R^2 = 0,84, \quad \widehat{\sigma}_\varepsilon = 8,24.$$

1. Calculer les statistiques suivantes :

(a) SCR**, la somme des carrés des résidus :**

$$\widehat{\sigma}_\varepsilon^2 = \frac{SCR}{n-2} \Rightarrow SCR = (n-2)\widehat{\sigma}_\varepsilon^2 = 8 \times 8,24^2 = 543,18.$$

(b) SCT**, la somme des carrés totaux :**

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \Rightarrow SCT = \frac{SCR}{1-R^2} = \frac{543,18}{1-0,84} = 3394,88.$$

(c) SCE**, la somme des carrés expliqués :**

$$SCT = SCE + SCR \Rightarrow SCE = SCT - SCR = 3394,88 - 543,18 = 2851,70.$$

(d) F^* **la valeur de la statistique du Fisher empirique :**

$$F^* = (n-2)\frac{R^2}{1-R^2} = 8 \times \frac{0,84}{1-0,84} = 42.$$

2. Sachant que le carré de t de Student t^* et la statistique de Fisher F^* sont égaux, calculer l'écart type $\widehat{\sigma}_{\widehat{a}_1}$ du coefficient $\widehat{a}_1$:

2. Sachant que le carré de t de Student t^* et la statistique de Fisher F^* sont égaux, calculer l'écart type $\widehat{\sigma}_{\widehat{a}_1}$ du coefficient $\widehat{a}_1$:

$$(t^*)^2 = F^* = 42 \Rightarrow t^* = \sqrt{F^*} = 6,48,$$

et puisque

$$t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}}, \quad \text{alors} \quad \widehat{\sigma}_{\widehat{a}_1} = \frac{|\widehat{a}_1|}{t^*} = \frac{1,6}{6,48} = 0,24.$$

2. Sachant que le carré de t de Student t^* et la statistique de Fisher F^* sont égaux, calculer l'écart type $\widehat{\sigma}_{\widehat{a}_1}$ du coefficient $\widehat{a}_1$:

$$(t^*)^2 = F^* = 42 \Rightarrow t^* = \sqrt{F^*} = 6,48,$$

et puisque

$$t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}}, \quad \text{alors} \quad \widehat{\sigma}_{\widehat{a}_1} = \frac{|\widehat{a}_1|}{t^*} = \frac{1,6}{6,48} = 0,24.$$

3. Le coefficient de la variable x est-il significativement supérieur à 1 ? On donne $t_8^{0,05} = 2,306$:

2. Sachant que le carré de t de Student t^* et la statistique de Fisher F^* sont égaux, calculer l'écart type $\widehat{\sigma}_{\widehat{a}_1}$ du coefficient $\widehat{a}_1$:

$$(t^*)^2 = F^* = 42 \Rightarrow t^* = \sqrt{F^*} = 6,48,$$

et puisque

$$t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}}, \quad \text{alors} \quad \widehat{\sigma}_{\widehat{a}_1} = \frac{|\widehat{a}_1|}{t^*} = \frac{1,6}{6,48} = 0,24.$$

3. Le coefficient de la variable x est-il significativement supérieur à 1 ?
On donne $t_8^{0,05} = 2,306$:

Nous testons à un seuil de 5% l'hypothèse

$$H_0 : "a_1 = 1" \quad \text{contre} \quad H_1 : "a_1 > 1".$$

2. Sachant que le carré de t de Student t^* et la statistique de Fisher F^* sont égaux, calculer l'écart type $\widehat{\sigma}_{\widehat{a}_1}$ du coefficient $\widehat{a}_1$:

$$(t^*)^2 = F^* = 42 \Rightarrow t^* = \sqrt{F^*} = 6,48,$$

et puisque

$$t^* = \frac{|\widehat{a}_1|}{\widehat{\sigma}_{\widehat{a}_1}}, \quad \text{alors} \quad \widehat{\sigma}_{\widehat{a}_1} = \frac{|\widehat{a}_1|}{t^*} = \frac{1,6}{6,48} = 0,24.$$

3. Le coefficient de la variable x est-il significativement supérieur à 1 ?
On donne $t_8^{0,05} = 2,306$:
Nous testons à un seuil de 5% l'hypothèse

$$H_0 : \text{"}\, a_1 = 1\text{"} \quad \text{contre} \quad H_1 : \text{"}\, a_1 > 1\text{"}.$$

Sous H_0, nous avons

$$t^*_{\widehat{a}_1} = \frac{|\widehat{a}_1 - 1|}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{1,6-1}{0,24} = 2,5$$

qui est supérieur à $t_8^{0,05} = 2,306$.
Donc, nous rejetons H_0, et alors a_1 est significativement supérieur à 1.

Exercice 3 :

Soit le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$, **où les résultats observés sont :**

t	y	x_1	x_2
1	4	2	5
2	5	7	8
3	8	6	9
4	12	8	9
5	15	9	10

Exercice 3 :

Soit le modèle : $y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$, **où les résultats observés sont :**

t	y	x_1	x_2
1	4	2	5
2	5	7	8
3	8	6	9
4	12	8	9
5	15	9	10

1. Mettre le modèle sous forme matricielle en spécifiant bien les dimensions de chacune des matrices :

$$Y = \underbrace{\begin{pmatrix} 4 \\ 5 \\ 8 \\ 12 \\ 15 \end{pmatrix}}_{(5,1)}, X = \underbrace{\begin{pmatrix} 1 & 2 & 5 \\ 1 & 7 & 8 \\ 1 & 6 & 9 \\ 1 & 8 & 9 \\ 1 & 9 & 10 \end{pmatrix}}_{(5,3)}, a = \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}}_{(3,1)}, \varepsilon = \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}}_{(5,1)}.$$

2. Estimer les paramètres :

2. Estimer les paramètres :

Nous savons que $\widehat{a} = (X'X)^{-1}X'Y$. Calculons $X'X$ puis $(X'X)^{-1}$.

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 7 & 6 & 8 & 9 \\ 5 & 8 & 9 & 9 & 10 \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 \\ 1 & 7 & 8 \\ 1 & 6 & 9 \\ 1 & 8 & 9 \\ 1 & 8 & 10 \end{pmatrix}$$

2. Estimer les paramètres :

Nous savons que $\widehat{a} = (X'X)^{-1}X'Y$. Calculons $X'X$ puis $(X'X)^{-1}$.

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 7 & 6 & 8 & 9 \\ 5 & 8 & 9 & 9 & 10 \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 \\ 1 & 7 & 8 \\ 1 & 6 & 9 \\ 1 & 8 & 9 \\ 1 & 8 & 10 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 5 & 32 & 41 \\ 32 & 234 & 282 \\ 41 & 282 & 351 \end{pmatrix} \Rightarrow (X'X)^{-1} = \frac{1}{det(X'X)} \left(com(X'X)\right)'$$

$$(X'X)^{-1} = \begin{pmatrix} 10.875 & 1.375 & -2.375 \\ 1.375 & 0.308 & -0.408 \\ -2.375 & -0.408 & 0.608 \end{pmatrix}$$

2. Estimer les paramètres :

Nous savons que $\widehat{a} = (X'X)^{-1}X'Y$. Calculons $X'X$ puis $(X'X)^{-1}$.

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 7 & 6 & 8 & 9 \\ 5 & 8 & 9 & 9 & 10 \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 \\ 1 & 7 & 8 \\ 1 & 6 & 9 \\ 1 & 8 & 9 \\ 1 & 8 & 10 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 5 & 32 & 41 \\ 32 & 234 & 282 \\ 41 & 282 & 351 \end{pmatrix} \Rightarrow (X'X)^{-1} = \frac{1}{det(X'X)} \left(com(X'X)\right)'$$

$$(X'X)^{-1} = \begin{pmatrix} 10.875 & 1.375 & -2.375 \\ 1.375 & 0.308 & -0.408 \\ -2.375 & -0.408 & 0.608 \end{pmatrix}$$

$$\text{On a} \quad X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 7 & 6 & 8 & 9 \\ 5 & 8 & 9 & 9 & 10 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 8 \\ 12 \\ 15 \end{pmatrix} = \begin{pmatrix} 44 \\ 322 \\ 390 \end{pmatrix}.$$

2. Estimer les paramètres :

Nous savons que $\widehat{a} = (X'X)^{-1}X'Y$. Calculons $X'X$ puis $(X'X)^{-1}$.

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 7 & 6 & 8 & 9 \\ 5 & 8 & 9 & 9 & 10 \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 \\ 1 & 7 & 8 \\ 1 & 6 & 9 \\ 1 & 8 & 9 \\ 1 & 8 & 10 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 5 & 32 & 41 \\ 32 & 234 & 282 \\ 41 & 282 & 351 \end{pmatrix} \Rightarrow (X'X)^{-1} = \frac{1}{det(X'X)} \left(com(X'X)\right)'$$

$$(X'X)^{-1} = \begin{pmatrix} 10.875 & 1.375 & -2.375 \\ 1.375 & 0.308 & -0.408 \\ -2.375 & -0.408 & 0.608 \end{pmatrix}$$

On a $$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 7 & 6 & 8 & 9 \\ 5 & 8 & 9 & 9 & 10 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 8 \\ 12 \\ 15 \end{pmatrix} = \begin{pmatrix} 44 \\ 322 \\ 390 \end{pmatrix}.$$

Alors, $$\widehat{a} = (X'X)^{-1}X'Y = \begin{pmatrix} 10.875 & 1.375 & -2.375 \\ 1.375 & 0.308 & -0.408 \\ -2.375 & -0.408 & 0.608 \end{pmatrix} \begin{pmatrix} 44 \\ 322 \\ 390 \end{pmatrix}$$

$$\widehat{a} = \begin{pmatrix} -5 \\ 0.533 \\ 1.266 \end{pmatrix}.$$

$$\Rightarrow \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \widehat{a}_2 \end{pmatrix} = \begin{pmatrix} -5 \\ 0.533 \\ 1.266 \end{pmatrix}$$

3. Calculer $\widehat{\sigma}_{\varepsilon}$ et $\widehat{\sigma}_{\widehat{a}}$:

3. Calculer $\widehat{\sigma}_{\varepsilon}$ et $\widehat{\sigma}_{\widehat{a}}$:

$$\text{On sait que} \quad \widehat{\sigma}_{\varepsilon}^2 = \frac{ee'}{n-k-1}, \quad \text{où} \quad e = Y - \widehat{Y} = Y - X\widehat{a}$$

3. Calculer $\widehat{\sigma}_\varepsilon$ et $\widehat{\sigma}_{\widehat{a}}$:

$$\text{On sait que} \quad \widehat{\sigma}_\varepsilon^2 = \frac{ee'}{n-k-1}, \quad \text{où} \quad e = Y - \widehat{Y} = Y - X\widehat{a}$$

$$\Rightarrow e = \begin{pmatrix} 4 \\ 5 \\ 8 \\ 12 \\ 15 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 5 \\ 1 & 7 & 8 \\ 1 & 6 & 9 \\ 1 & 8 & 9 \\ 1 & 9 & 10 \end{pmatrix} \begin{pmatrix} -5 \\ 0.533 \\ 1.266 \end{pmatrix}$$

$$\Rightarrow e = \begin{pmatrix} 4 \\ 5 \\ 8 \\ 12 \\ 15 \end{pmatrix} - \begin{pmatrix} 2.4 \\ 8.866 \\ 9.6 \\ 10.666 \\ 12.466 \end{pmatrix} = - \begin{pmatrix} -1.6 \\ 3.866 \\ 1.6 \\ -1,333 \\ -2.533 \end{pmatrix}.$$

t	y_t	$\widehat{y}_t$	e_t	e_t^2
1	4	2.399	-1.6	2.56
2	5	8.866	3.866	14.951
3	8	9.599	1.6	2.56
4	12	10.666	$-1,333$	1.777
5	15	12.466	-2.533	6.417
Somme			0	28.266

On a $\quad \widehat{\sigma}_{\varepsilon}^{2} = \dfrac{ee'}{n-k-1} = \dfrac{\sum_{t=1}^{5} e_t^2}{5-2-1} = \dfrac{28.266}{2} = 14.133.$

La matrice des variances−covariances estimées est donnée par :

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

$$\text{On a} \quad \widehat{\sigma}_{\varepsilon}^{2} = \frac{ee'}{n-k-1} = \frac{\sum_{t=1}^{5} e_t^2}{5-2-1} = \frac{28.266}{2} = 14.133.$$

La matrice des variances−covariances estimées est donnée par :

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

$$\Rightarrow \widehat{\Omega}_{\widehat{a}} = 14.133 \times \begin{pmatrix} 10.875 & 1.375 & -2.375 \\ 1.375 & 0.308 & -0.408 \\ -2.375 & -0.408 & 0.608 \end{pmatrix}$$

$$= \begin{pmatrix} 153.696 & 19.432 & -33.565 \\ 19.432 & 4.357 & -5.770 \\ -33.565 & -5.770 & 8.597 \end{pmatrix}.$$

$$\text{On a} \quad \widehat{\sigma}_{\varepsilon}^{2} = \frac{ee'}{n-k-1} = \frac{\sum_{t=1}^{5} e_t^2}{5-2-1} = \frac{28.266}{2} = 14.133.$$

La matrice des variances$-$covariances estimées est donnée par :

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

$$\Rightarrow \widehat{\Omega}_{\widehat{a}} = 14.133 \times \begin{pmatrix} 10.875 & 1.375 & -2.375 \\ 1.375 & 0.308 & -0.408 \\ -2.375 & -0.408 & 0.608 \end{pmatrix}$$

$$= \begin{pmatrix} 153.696 & 19.432 & -33.565 \\ 19.432 & 4.357 & -5.770 \\ -33.565 & -5.770 & 8.597 \end{pmatrix}.$$

Les variances des coefficients de régression se trouvent sur la première diagonale :

$$\widehat{\sigma}_{\widehat{a}_0}^{2} = 153.696 \Rightarrow \widehat{\sigma}_{\widehat{a}_0} = 12.397,$$

$$\widehat{\sigma}_{\widehat{a}_1}^{2} = 4.357 \Rightarrow \widehat{\sigma}_{\widehat{a}_1} = 2.087,$$

$$\widehat{\sigma}_{\widehat{a}_2}^{2} = 8.597 \Rightarrow \widehat{\sigma}_{\widehat{a}_2} = 2.932.$$

4. Trouver les intervalles de confiance pour a_i, $i = 0,1,2$ à un seuil de 95% :
Les intervalles de confiance pour a_i, $i = 0,1,2$ à un seuil de 95% sont tels que :

$$IC_{a_i} = \left[\widehat{a}_i - \widehat{\sigma}_{\widehat{a}_i} \times t_{n-k-1}^{\alpha/2} ; \widehat{a}_i + \widehat{\sigma}_{\widehat{a}_i} \times t_{n-k-1}^{\alpha/2}\right],$$

avec $\alpha = 1 - 95\% = 0.05$. Donc,

4. Trouver les intervalles de confiance pour a_i, $i = 0,1,2$ à un seuil de 95% :
Les intervalles de confiance pour a_i, $i = 0,1,2$ à un seuil de 95% sont tels que :

$$IC_{a_i} = \left[\widehat{a}_i - \widehat{\sigma}_{\widehat{a}_i} \times t_{n-k-1}^{\alpha/2}; \widehat{a}_i + \widehat{\sigma}_{\widehat{a}_i} \times t_{n-k-1}^{\alpha/2}\right],$$

avec $\alpha = 1 - 95\% = 0.05$. Donc,

$$IC_{a_0} = [-5 - 12.397 \times 4.3027; -5 + 12.397 \times 4.3027] = [-58.340; 48.340],$$
$$IC_{a_1} = [0.533 - 2.087 \times 4.3027; 0.533 + 2.087 \times 4.3027] = [-8.446; 9.512],$$
$$IC_{a_2} = [1.266 - 2.932 \times 4.3027; 1.266 + 2.932 \times 4.3027] = [-11.349; 13.881].$$

5. Calculer le coefficient de détermination R^2 puis tester à un seuil de 5% l'hypothèse

$$H_0 : {}'' a_i = 0'' \quad \textbf{contre} \quad H_1 : {}'' a_i \neq 0'', \quad i = 0, 1$$

5. Calculer le coefficient de détermination R^2 puis tester à un seuil de 5% l'hypothèse

$$H_0 : "\, a_i = 0" \quad \textbf{contre} \quad H_1 : "\, a_i \neq 0", \quad i = 0, 1$$

Le coefficient de détermination R^2 est :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^{5} e_t^2}{\sum_{t=1}^{5} (y_t - \overline{y})^2}.$$

5. Calculer le coefficient de détermination R^2 puis tester à un seuil de 5% l'hypothèse

$$H_0 : "a_i = 0" \quad \textbf{contre} \quad H_1 : "a_i \neq 0", \quad i = 0, 1$$

Le coefficient de détermination R^2 est :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^{5} e_t^2}{\sum_{t=1}^{5}(y_t - \overline{y})^2}.$$

Nous avons $ee' = \sum_{t=1}^{5} e_t^2 = 28.266$ et $\sum_{t=1}^{5}(y_t - \overline{y})^2 = 86.8$, donc

$$R^2 = 1 - \frac{28.266}{86.8} = 0.674.$$

5. Calculer le coefficient de détermination R^2 puis tester à un seuil de 5% l'hypothèse

$$H_0 : "a_i = 0" \quad \textbf{contre} \quad H_1 : "a_i \neq 0", \quad i = 0, 1$$

Le coefficient de détermination R^2 est :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^{5} e_t^2}{\sum_{t=1}^{5}(y_t - \overline{y})^2}.$$

Nous avons $ee' = \sum_{t=1}^{5} e_t^2 = 28.266$ et $\sum_{t=1}^{5}(y_t - \overline{y})^2 = 86.8$, donc

$$R^2 = 1 - \frac{28.266}{86.8} = 0.674.$$

La significativité de a_1 et a_2 :

Nous testons à un seuil de 5% l'hypothèse

$$H_0 : "a_i = 0" \quad \text{contre} \quad H_1 : "a_i \neq 0", \quad i = 0, 1.$$

5. Calculer le coefficient de détermination R^2 puis tester à un seuil de 5% l'hypothèse

$$H_0 : '' a_i = 0'' \quad \textbf{contre} \quad H_1 : '' a_i \neq 0'', \quad i = 0, 1$$

Le coefficient de détermination R^2 est :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^{5} e_t^2}{\sum_{t=1}^{5}(y_t - \overline{y})^2}.$$

Nous avons $ee' = \sum_{t=1}^{5} e_t^2 = 28.266$ et $\sum_{t=1}^{5}(y_t - \overline{y})^2 = 86.8$, donc

$$R^2 = 1 - \frac{28.266}{86.8} = 0.674.$$

La significativité de a_1 et a_2 :

Nous testons à un seuil de 5% l'hypothèse

$$H_0 : '' a_i = 0'' \quad \text{contre} \quad H_1 : '' a_i \neq 0'', \quad i = 0, 1.$$

Sous H_0, nous avons $t^*_{\widehat{a}_i} = \dfrac{|\widehat{a}_i|}{\widehat{\sigma}_{\widehat{a}_i}} = \begin{cases} \frac{|0.533|}{2.087} = 0.255; & \text{i=1,} \\ \frac{|1.266|}{2.932} = 0.431; & \text{i=2.} \end{cases}$

qui sont inférieurs à $t_{n-k-1}^{\alpha/2} = t_2^{0.025} = 4.3027$.

Donc, nous acceptons H_0, et alors a_1 et a_2 ne sont pas significativement différents de 0.

6. Calculer les sommes SCR, SCT et SCE :

6. Calculer les sommes SCR, SCT **et** SCE **:**

- $SCR = \sum_{t=1}^{5}(y_t - \widehat{y}_t)^2 = \sum_{t=1}^{5} e_t^2 = 28.266$
- $SCT = \sum_{t=1}^{5}(y_t - \overline{y})^2 = 86.8$
- $SCT = SCE + SCR \Rightarrow SCE = SCT - SCR = 86.8 - 28.266 = 58.434$

6. Calculer les sommes *SCR*, *SCT* et *SCE* :

- $SCR = \sum_{t=1}^{5}(y_t - \widehat{y}_t)^2 = \sum_{t=1}^{5} e_t^2 = 28.266$
- $SCT = \sum_{t=1}^{5}(y_t - \overline{y})^2 = 86.8$
- $SCT = SCE + SCR \Rightarrow SCE = SCT - SCR = 86.8 - 28.266 = 58.434$

7. Tracer le tableau d'analyse de la variance et faire un test de Fisher à un seuil de 5% :

Variation	SC	DDL	CM
x_1, x_2	$SCE = 58.434$	2	29.217
Résidu	$SCR = 28.266$	2	14.133
Total	$SCT = 86.8$	4	

On a $F^* = \frac{SCE/2}{SCR/2} = \frac{29.217}{14.133} = 2.067$ et $F_{(2;2)}^{0,05} = 19$.
Puisque $F^* < F_{(2;2)}^{0.05}$, nous acceptons H_0 et le modèle n'est pas globalement explicatif par x_1 et x_2.

6. Calculer les sommes SCR, SCT et SCE :

- $SCR = \sum_{t=1}^{5}(y_t - \widehat{y}_t)^2 = \sum_{t=1}^{5} e_t^2 = 28.266$
- $SCT = \sum_{t=1}^{5}(y_t - \overline{y})^2 = 86.8$
- $SCT = SCE + SCR \Rightarrow SCE = SCT - SCR = 86.8 - 28.266 = 58.434$

7. Tracer le tableau d'analyse de la variance et faire un test de Fisher à un seuil de 5% :

Variation	SC	DDL	CM
x_1, x_2	$SCE = 58.434$	2	29.217
Résidu	$SCR = 28.266$	2	14.133
Total	$SCT = 86.8$	4	

On a $F^* = \frac{SCE/2}{SCR/2} = \frac{29.217}{14.133} = 2.067$ et $F^{0,05}_{(2;2)} = 19$.
Puisque $F^* < F^{0.05}_{(2;2)}$, nous acceptons H_0 et le modèle n'est pas globalement explicatif par x_1 et x_2.

8. Déterminer un intervalle de confiance de l'erreur à un seuil de 5% :

Il est calculé à partir de la formule $IC = \left[\frac{(n-k-1)\widehat{\sigma}^2_\varepsilon}{\chi^2_1}; \frac{(n-k-1)\widehat{\sigma}^2_\varepsilon}{\chi^2_2}\right]$.

Pour 2 degrés de liberté, on a $\chi^2_{0.025} = 7,38$ et $\chi^2_{0.975} = 0,05$. Donc,

$$IC = \left[\frac{2 \times 14.133}{7.38}; \frac{2 \times 14.133}{0.05}\right] = [3.830; 565.320].$$

Exercice 4 :

La quantité $y_i, \quad i = 1, \ldots, 5$ en kg, de 5 biens produit dans un atelier peut être expliquer par X_1 le nombre d'heure de travail dans une journée suivant le modèle linéaire $\mathcal{L}_1$: $Y = \alpha_0 + \alpha_1 X_1 + \varepsilon$. Les résultats sont donnés dans le tableau suivant :

$X_1 : x_{1i}$ en h	4	4.5	5	5.5	6
$Y : y_i$ en kg	25	30	35	40	40

Exercice 4 :

La quantité y_i, $i = 1, \ldots, 5$ en kg, de 5 biens produit dans un atelier peut être expliquer par X_1 le nombre d'heure de travail dans une journée suivant le modèle linéaire $\mathcal{L}_1$: $Y = \alpha_0 + \alpha_1 X_1 + \varepsilon$. Les résultats sont donnés dans le tableau suivant :

X_1 : x_{1i} en h	4	4.5	5	5.5	6
Y : y_i en kg	25	30	35	40	40

1. Donner les expressions puis les valeurs numériques de $\widehat{\alpha}_1$ et $\widehat{\alpha}_0$ du modèle $\mathcal{L}_1$ et déduire la série des résidus $(y_i - \widehat{y}_i)^2_{1 \leq i \leq 5}$ associée à ce modèle :

$$\widehat{\alpha}_1 = \frac{Cov(X_1; Y)}{Var(X_1)} = 8, \quad \widehat{\alpha}_0 = \overline{Y} - \widehat{\alpha}_1 \overline{X_1} = -6.$$

X_1 : x_{1i} en h	4	4.5	5	5.5	6
Y : y_i en kg	25	30	35	40	40
$(y_i - \widehat{y}_i)^2$	1	0	1	4	4

Exercice 4 :

La quantité $y_i, \quad i=1,\dots,5$ en kg, de 5 biens produit dans un atelier peut être expliquer par X_1 le nombre d'heure de travail dans une journée suivant le modèle linéaire $\mathcal{L}_1 : Y=\alpha_0+\alpha_1 X_1+\varepsilon$. Les résultats sont donnés dans le tableau suivant :

$X_1 : x_{1i}$ en h	4	4.5	5	5.5	6
$Y : y_i$ en kg	25	30	35	40	40

1. Donner les expressions puis les valeurs numériques de $\widehat{\alpha}_1$ et $\widehat{\alpha}_0$ du modèle $\mathcal{L}_1$ et déduire la série des résidus $(y_i-\widehat{y}_i)^2_{1\leq i\leq 5}$ associée à ce modèle :

$$\widehat{\alpha}_1=\frac{Cov(X_1;Y)}{Var(X_1)}=8, \quad \widehat{\alpha}_0=\overline{Y}-\widehat{\alpha}_1\overline{X_1}=-6.$$

$X_1 : x_{1i}$ en h	4	4.5	5	5.5	6
$Y : y_i$ en kg	25	30	35	40	40
$(y_i-\widehat{y}_i)^2$	1	0	1	4	4

2. Estimer la variance résiduelle et les écarts type de $\widehat{\alpha}_1$ et $\widehat{\alpha}_0$:

- $\widehat{\sigma}^2_\varepsilon=\frac{SCR}{n-2}=\frac{1}{n-2}\sum_{i=1}^5(y_i-\widehat{y}_i)^2=\frac{10}{3}=3.33,$
- $\widehat{\sigma}^2_{\widehat{\alpha}_1}=\frac{\widehat{\sigma}^2_\varepsilon}{\sum_{i=1}^n(x_i-\overline{x})^2}=\frac{\widehat{\sigma}^2_\varepsilon}{n\times Var(x)}=\frac{3.33}{5\times 0.5}\approx 1.33\Rightarrow \widehat{\sigma}_{\widehat{\alpha}_1}\approx 1.15,$
- $\widehat{\sigma}^2_{\widehat{\alpha}_0}=\widehat{\sigma}^2_\varepsilon\left(\frac{1}{n}+\frac{\overline{x}^2}{\sum_{i=1}^n(x_i-\overline{x})^2}\right)=\widehat{\sigma}^2_\varepsilon\left(\frac{1}{n}+\frac{\overline{x}^2}{n\times Var(x)}\right)=34\Rightarrow \widehat{\sigma}_{\widehat{\alpha}_0}\approx 5.83.$

3. Déterminer au seuil de 5%, un intervalle de confiance pour y_6 relatif à l'heure prévisible $x_{1,6} = 7h$:

L'intervalle de prédiction $I_{y_{n+1}}$ relatif à la valeur prévisible x_{n+1} est donné par :

$$y_{n+1} = \widehat{y}_{n+1} \pm t_{n-2}^{\alpha/2} \widehat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} + 1}.$$

Avec

$$x_{n+1} = x_{1,6} = 7, \quad \overline{x} = 5, \quad \sum_{i=1}^{n} (x_i - \overline{x})^2 = 2.5, \quad \widehat{y}_{n+1} = \widehat{y}_6 = 50$$

$$t_{n-2}^{\alpha/2} = t_3^{0,025} = 3.182, \quad \widehat{\sigma}_\varepsilon = \sqrt{3.33} \approx 1.82.$$

Donc,

$$I_{y_6} = [40.31; 59.69].$$

4. On souhaite savoir si $X_2 : x_{2i}, \quad i = 1, \ldots, 5,$ le nombre des ouvriers a aussi une influence sur la production de ces biens dans une même journée.
On considère alors le modèle linéaire $\mathcal{L}_2$: $Y = a_0 + a_1 X_1 + a_2 X_2 + \varepsilon$.
Soit $X = (\mathbb{I} \quad X_1 \quad X_2)$ la matrice associée. Nous avons obtenu les résultats suivants :

$$X'X = \begin{pmatrix} ? & 25 & ? \\ ? & 127.3 & 130 \\ 27 & ? & 157 \end{pmatrix}, \quad (X'X)^{-1} = \begin{pmatrix} 207.83 & ? & ? \\ -27.66 & 3.73 & ? \\ -12.83 & 1.66 & 0.83 \end{pmatrix}.$$

4. On souhaite savoir si $X_2 : x_{2i}, \quad i = 1, \ldots, 5$, le nombre des ouvriers a aussi une influence sur la production de ces biens dans une même journée.
On considère alors le modèle linéaire $\mathcal{L}_2$: $Y = a_0 + a_1 X_1 + a_2 X_2 + \varepsilon$.
Soit $X = (\mathbb{I} \quad X_1 \quad X_2)$ la matrice associée. Nous avons obtenu les résultats suivants :

$$X'X = \begin{pmatrix} ? & 25 & ? \\ ? & 127.3 & 130 \\ 27 & ? & 157 \end{pmatrix}, \quad (X'X)^{-1} = \begin{pmatrix} 207.83 & ? & ? \\ -27.66 & 3.73 & ? \\ -12.83 & 1.66 & 0.83 \end{pmatrix}.$$

(a) Donner les valeurs manquantes :
Le premier élément de la matrice $X'X$ et $n = 5$ le nombre des données. La matrice $X'X$ est symétrique, donc les valeurs manquantes sont les symétriques des valeurs qui existent par rapport à la diagonale et de même pour $(X'X)^{-1}$. On conclut

4. On souhaite savoir si $X_2 : x_{2i}, \quad i = 1, \ldots, 5,$ le nombre des ouvriers a aussi une influence sur la production de ces biens dans une même journée.
On considère alors le modèle linéaire $\mathcal{L}_2$: $Y = a_0 + a_1 X_1 + a_2 X_2 + \varepsilon$.
Soit $X = (\mathbb{I} \quad X_1 \quad X_2)$ la matrice associée. Nous avons obtenu les résultats suivants :

$$X'X = \begin{pmatrix} ? & 25 & ? \\ ? & 127.3 & 130 \\ 27 & ? & 157 \end{pmatrix}, \quad (X'X)^{-1} = \begin{pmatrix} 207.83 & ? & ? \\ -27.66 & 3.73 & ? \\ -12.83 & 1.66 & 0.83 \end{pmatrix}.$$

(a) Donner les valeurs manquantes :

Le premier élément de la matrice $X'X$ et $n = 5$ le nombre des données. La matrice $X'X$ est symétrique, donc les valeurs manquantes sont les symétriques des valeurs qui existent par rapport à la diagonale et de même pour $(X'X)^{-1}$. On conclut

$$X'X = \begin{pmatrix} \color{red}{5} & 25 & \color{red}{27} \\ \color{red}{25} & 127.3 & 130 \\ 27 & \color{red}{130} & 157 \end{pmatrix}, (X'X)^{-1} = \begin{pmatrix} 207.83 & \color{red}{-27.66} & \color{red}{-12.83} \\ -27.66 & 3.73 & \color{red}{1.66} \\ -12.83 & 1.66 & 0.83 \end{pmatrix}.$$

(b) Calculer le coefficient de corrélation linéaire empirique entre X_1 et X_2 :

(b) Calculer le coefficient de corrélation linéaire empirique entre X_1 et X_2 :

$$r_{X_1,X_2} = \frac{Cov(X_1,X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sum_{i=1}^{n}\left(x_{1i}-\overline{X_1}\right)\left(x_{2i}-\overline{X_2}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{1i}-\overline{X_1}\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_{2i}-\overline{X_2}\right)^2}} = \frac{\sum_{i=1}^{n}(x_{1i}x_{2i})-n\overline{X_1}\,\overline{X_2}}{\sqrt{\sum_{i=1}^{n}x_{1i}^2-n\overline{X_1}^2}\sqrt{\sum_{i=1}^{n}x_{2i}^2-n\overline{X_2}^2}}.$$

Or, $X'X = \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{pmatrix}$

$$\Leftrightarrow \begin{pmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 \end{pmatrix} = \begin{pmatrix} 5 & 25 & 27 \\ 25 & 127.3 & 130 \\ 27 & 130 & 157 \end{pmatrix}.$$

(b) Calculer le coefficient de corrélation linéaire empirique entre X_1 et X_2 :

$$r_{X_1,X_2} = \frac{Cov(X_1,X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sum_{i=1}^{n}\left(x_{1i}-\overline{X_1}\right)\left(x_{2i}-\overline{X_2}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{1i}-\overline{X_1}\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_{2i}-\overline{X_2}\right)^2}} = \frac{\sum_{i=1}^{n}(x_{1i}x_{2i})-n\overline{X_1X_2}}{\sqrt{\sum_{i=1}^{n}x_{1i}^2-n\overline{X_1}^2}\sqrt{\sum_{i=1}^{n}x_{2i}^2-n\overline{X_2}^2}}.$$

Or, $X'X = \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{pmatrix}$

$$\Leftrightarrow \begin{pmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 \end{pmatrix} = \begin{pmatrix} 5 & 25 & 27 \\ 25 & 127.3 & 130 \\ 27 & 130 & 157 \end{pmatrix}.$$

Par identification des coefficients,

- $\overline{X_1} = \frac{1}{n}\sum x_{1i} = \frac{1}{5}\sum x_{1i} = \frac{25}{5} = 5,$
- $\overline{X_2} = \frac{1}{n}\sum x_{2i} = \frac{27}{5} = 5.4,$
- $\sum x_{1i}x_{2i} = 130, \quad \sum_{i=1}^{n} x_{1i}^2 = 127.3, \quad \sum_{i=1}^{n} x_{2i}^2 = 157.$

Donc,

$$r_{X_1,X_2} = \frac{130 - 5\times 5\times 5.4}{\sqrt{127.3 - 5\times 5^2}\sqrt{157 - 5\times 5.4^2}} \approx -0.29.$$

(c) Estimer $a = (a_0, a_1, a_2)$:

On sait que $\widehat{a} = (X'X)^{-1}X'Y$. Donc

$$\widehat{a} = \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \widehat{a}_2 \end{pmatrix} = \begin{pmatrix} 207.83 & -27.66 & -12.83 \\ -27.66 & 3.73 & 1.66 \\ -12.83 & 1.66 & 0.83 \end{pmatrix} \begin{pmatrix} 170 \\ 870 \\ 875 \end{pmatrix} = \begin{pmatrix} 40.65 \\ -4.6 \\ -10.65 \end{pmatrix}.$$

(d) Les écarts type se calculent à partir de la matrice variance-covariance :

- $$e = Y - \widehat{Y} = Y - X\widehat{a} = \begin{pmatrix} 25 \\ 30 \\ 35 \\ 40 \\ 40 \end{pmatrix} - \begin{pmatrix} 24.5 \\ 31 \\ 35 \\ 39 \\ 40.5 \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ 0 \\ 1 \\ -0.5 \end{pmatrix}$$

  $$\Rightarrow \widehat{\sigma}_{\varepsilon}^2 = \frac{ee'}{n-k-1} = \frac{\sum_{i=1}^5 e_i^2}{2} = \frac{2.5}{2} = 1.25.$$

  $$\widehat{\Omega}_{\widehat{a}} = \widehat{\sigma}_{\varepsilon}^2 (X'X)^{-1} = 1.25 \begin{pmatrix} 207.83 & -27.66 & -12.83 \\ -27.66 & 3.73 & 1.66 \\ -12.83 & 1.66 & 0.83 \end{pmatrix}.$$

(c) Estimer $a = (a_0, a_1, a_2)$:
On sait que $\widehat{a} = (X'X)^{-1}X'Y$. Donc

$$\widehat{a} = \begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \widehat{a}_2 \end{pmatrix} = \begin{pmatrix} 207.83 & -27.66 & -12.83 \\ -27.66 & 3.73 & 1.66 \\ -12.83 & 1.66 & 0.83 \end{pmatrix} \begin{pmatrix} 170 \\ 870 \\ 875 \end{pmatrix} = \begin{pmatrix} 40.65 \\ -4.6 \\ -10.65 \end{pmatrix}.$$

(d) Les écarts type se calculent à partir de la matrice variance-covariance :

- $e = Y - \widehat{Y} = Y - X\widehat{a} = \begin{pmatrix} 25 \\ 30 \\ 35 \\ 40 \\ 40 \end{pmatrix} - \begin{pmatrix} 24.5 \\ 31 \\ 35 \\ 39 \\ 40.5 \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ 0 \\ 1 \\ -0.5 \end{pmatrix}$

  $\Rightarrow \widehat{\sigma}_{\varepsilon}^2 = \frac{ee'}{n-k-1} = \frac{\sum_{i=1}^{5} e_i^2}{2} = \frac{2.5}{2} = 1.25.$

$$\widehat{\Omega}_{\widehat{a}} = \widehat{\sigma}_{\varepsilon}^2 (X'X)^{-1} = 1.25 \begin{pmatrix} 207.83 & -27.66 & -12.83 \\ -27.66 & 3.73 & 1.66 \\ -12.83 & 1.66 & 0.83 \end{pmatrix}.$$

Ses éléments diagonaux donnent :

- $\widehat{\sigma}_{\widehat{a}_0}^2 = 259.78 \Rightarrow \widehat{\sigma}_{\widehat{a}_0} \approx 16.11,$
- $\widehat{\sigma}_{\widehat{a}_1}^2 = 4.66 \Rightarrow \widehat{\sigma}_{\widehat{a}_1} \approx 2.16,$
- $\widehat{\sigma}_{\widehat{a}_2}^2 = 1.03 \Rightarrow \widehat{\sigma}_{\widehat{a}_2} \approx 1.01.$

(e) Tester à un seuil de 5% l'hypothèse nulle $''H_0 : a_2 = 0''$ **:**

(e) Tester à un seuil de 5% l'hypothèse nulle $"H_0 : a_2 = 0"$ **:**
Sous H_0, nous avons

$$t^*_{\widehat{a}_2} = \frac{|\widehat{a}_2|}{\widehat{\sigma}_{\widehat{a}_2}} = \frac{10.65}{1.01} \approx 10.54$$

qui est supérieure à $t^{\alpha/2}_{n-k-1} = t^{0,025}_{2} = 4.3$.
Donc, nous rejetons H_0, et alors a_2 est significativement différente de 0.

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$:

Exercice 5 :
On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$: **1. Compléter le tableau d'analyse de variance correspondant :**

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$: **1. Compléter le tableau d'analyse de variance correspondant :**

Variation	SC	DDL	CM
X_1, X_2	1504,4	2	$752,2$
Résidu	176,4	$n-3$	$9,8$
Total	$1680,8$	$n-1$	$\times\times\times$

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$: **1. Compléter le tableau d'analyse de variance correspondant :**

Variation	SC	DDL	CM
X_1, X_2	1504,4	2	$752,2$
Résidu	176,4	$n-3$	$9,8$
Total	$1680,8$	$n-1$	$\times\times\times$

2. Trouver la valeur de n **:**

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$ **: 1. Compléter le tableau d'analyse de variance correspondant :**

Variation	SC	DDL	CM
X_1, X_2	1504,4	2	$752,2$
Résidu	176,4	$n-3$	$9,8$
Total	$1680,8$	$n-1$	$\times\times\times$

2. Trouver la valeur de n **:** On sait que $CM = \frac{SC}{DDL}$, alors pour la somme des carrés résiduelles on a

$$CMR = \frac{SCR}{DDL} = \frac{SCR}{n-3} \Leftrightarrow n = \frac{SCR}{CMR} + 3 = \frac{176,4}{9,8} + 3 = 21.$$

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$: **1. Compléter le tableau d'analyse de variance correspondant :**

Variation	SC	DDL	CM
X_1, X_2	1504,4	2	$752,2$
Résidu	176,4	$n-3$	$9,8$
Total	$1680,8$	$n-1$	$\times\times\times$

2. Trouver la valeur de n **:** On sait que $CM = \frac{SC}{DDL}$, alors pour la somme des carrés résiduelles on a

$$CMR = \frac{SCR}{DDL} = \frac{SCR}{n-3} \Leftrightarrow n = \frac{SCR}{CMR} + 3 = \frac{176,4}{9,8} + 3 = 21.$$

3. Tester l'hypothèse nulle H_0 : "$\alpha_1 = \alpha_2 = 0$" **au niveau** 95%. **On donne** $F^{0,95}_{2;18} = 3,55$ **:**

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$ **: 1. Compléter le tableau d'analyse de variance correspondant :**

Variation	SC	DDL	CM
X_1, X_2	1504,4	2	$752,2$
Résidu	176,4	$n-3$	$9,8$
Total	$1680,8$	$n-1$	$\times\times\times$

2. Trouver la valeur de n **:** On sait que $CM = \frac{SC}{DDL}$, alors pour la somme des carrés résiduelles on a

$$CMR = \frac{SCR}{DDL} = \frac{SCR}{n-3} \Leftrightarrow n = \frac{SCR}{CMR} + 3 = \frac{176,4}{9,8} + 3 = 21.$$

3. Tester l'hypothèse nulle H_0 : "$\alpha_1 = \alpha_2 = 0$" **au niveau** 95%**. On donne** $F_{2;18}^{0,95} = 3,55$ **:**

$$\text{On a} \quad F^* = \frac{SCE/2}{SCR/18} = \frac{752,2}{9,8} = 76,75, \quad \text{et } F_{(2;18)}^{0,95} = 3,55.$$

Exercice 5 :

On considère le modèle de régression linéaire multiple $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon$: **1. Compléter le tableau d'analyse de variance correspondant :**

Variation	SC	DDL	CM
X_1, X_2	1504,4	2	$752,2$
Résidu	176,4	$n-3$	$9,8$
Total	$1680,8$	$n-1$	$\times\times\times$

2. Trouver la valeur de n **:** On sait que $CM = \frac{SC}{DDL}$, alors pour la somme des carrés résiduelles on a

$$CMR = \frac{SCR}{DDL} = \frac{SCR}{n-3} \Leftrightarrow n = \frac{SCR}{CMR} + 3 = \frac{176,4}{9,8} + 3 = 21.$$

3. Tester l'hypothèse nulle H_0 : "$\alpha_1 = \alpha_2 = 0$" **au niveau** 95%**. On donne** $F^{0,95}_{2;18} = 3,55$ **:**

$$\text{On a} \quad F^* = \frac{SCE/2}{SCR/18} = \frac{752,2}{9,8} = 76,75, \quad \text{et } F^{0,95}_{(2;18)} = 3,55.$$

Puisque $F^* > F^{0,95}_{(2;18)}$, nous rejetons H_0 et le modèle est globalement explicatif.

4. Quel est le R^2 du modèle ? Proposer une interprétation du résultat :

$$\text{On a} \quad R^2 = \frac{SCE}{SCT} = \frac{1504,4}{1680,8} = 0,89.$$

Ce résultat nous assure que notre régression est bien présentée par ce modèle linéaire.

5. Donner une estimation de σ^2, la variance de ε :

$$\text{On a} \quad \widehat{\sigma}_{\varepsilon}^2 = \frac{ee'}{n-k-1} = \frac{SCR}{n-3} = \frac{176,4}{18} = 9,8 \Rightarrow \widehat{\sigma}_{\varepsilon} = \sqrt{9,8} = 3,13.$$