Feuille de Travaux Dirigés n° 4
Correction de Régression linéaire multiple

Exercice IV.1. Question 1. La somme des carrés dûe à la régression pour l'ensemble des trois variables est égale à :

$$981,326 + 190,232 + 129,431 = 1300,989.$$

On peut également calculer la somme ainsi :

$$1743,281 - 442,292 = 1300,989.$$

Question 2. La proportion de la variation dans le niveau d'anxiété est égale à :

$$R^2 = \frac{SC_{reg}}{SC_{tot}} = \frac{1300,989}{1743,281} = 0,746,$$

ou encore 74, 6%.

Question 3. Pour répondre à cette question, il faudrait s'assurer que les trois hypothèses du modèle sont vérifiées. Malheureusement on ne pourra pas le faire ici puisque nous ne connaissons pas les valeurs des observations. Donc nous allons supposer que les trois hypothèses sont vérifiées mais dans la pratique il faudrait les vérifier ABSOLUMENT.

Pour conclure que dans l'ensemble les trois variables ont un effet significatif sur le niveau d'anxiété, il faut faire **un test de Fisher**. Le modèle est :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon,$$

où ε est la variable résiduelle sur laquelle les trois hypothèses sont faites.
L'hypothèse nulle :

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : \exists j = 1,\ 2,\ \text{ou } 3,\ \beta_j \neq 0.$$

Calculons la statistique du test de Fisher observée qui est égale à :

$$F_{obs} = \frac{SC_{reg}/ddl}{SC_{res}/ddl} = \frac{1300,989/3}{442,292/(22-3-1=18)} \simeq 17,649.$$

Le quantile de la loi de Fisher critique lu dans la table des quantiles de la loi de Fisher à 95% est égal à :

$$F_{c,3,18} = 3,160.$$

La statistique du test de Fisher observée est plus grande que le quantile de la loi de Fisher critique, à 95%. Donc on est dans la zone de rejet de l'hypothèse nulle $\mathcal{H}_0$. Donc on décide d'accepter l'hypothèse alternative $\mathcal{H}_1$, c'est-à-dire :

$$\exists j = 1,\ 2,\ \text{ou } 3,\ \beta_j \neq 0.$$

Question 4.

Source de variation	Somme des carrés	*ddl*
Régression due à X_1	$981,326$	**1**
Résiduelle	$\mathbf{761,955}$	**20**
Totale	$\mathbf{1743,281}$	**21**

Question 5. Même remarque qu'à la question 3 de cet exercice.

a) Le modèle est :

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon.$$

L'hypothèse nulle :

$$\mathcal{H}_0 : \beta_1 = 0$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : \beta_1 \neq 0.$$

Calculons la statistique du test de Fisher observée qui est égale à :

$$F_{obs} = \frac{SC_{reg}/ddl}{SC_{res}/ddl} = \frac{981,326/1}{761,955/(22-1-1=20)} = 25,758.$$

Le quantile de la loi de Fisher critique lu dans la table des quantiles de la loi de Fisher à 95% est égal à :

$$F_{c,1,20} = 4,350.$$

La statistique du test de Fisher observée est plus grande que le quantile de la loi de Fisher critique. Donc on est dans la zone de rejet de l'hypothèse nulle $\mathcal{H}_0$. Donc on décide d'accepter l'hypothèse alternative $\mathcal{H}_1$, c'est-à-dire :

$$\beta_1 \neq 0.$$

b) Le modèle est :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon.$$

L'hypothèse nulle :

$$\mathcal{H}_0 : \beta_2 = 0$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : \beta_2 \neq 0.$$

Calculons la statistique du test de Fisher observée qui est égale à :

$$F_{obs} = \frac{SC_{reg}/ddl}{SC_{res}/ddl} = \frac{190,232/1}{571,723/(22-2-1=19)} = 6,332.$$

Le quantile de la loi de Fisher critique lu dans la table des quantiles de la loi de Fisher à 95% est égal à :

$$F_{c,1,19} = 4,380.$$

La statistique du test de Fisher observée est plus grande que le quantile de la loi de Fisher critique. Donc on est dans la zone de rejet de l'hypothèse nulle $\mathcal{H}_0$. Donc on décide d'accepter l'hypothèse alternative $\mathcal{H}_1$, c'est-à-dire :

$$\beta_2 \neq 0.$$

c) Le modèle est :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon.$$

L'hypothèse nulle :

$$\mathcal{H}_0 : \beta_3 = 0$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : \beta_3 \neq 0.$$

Calculons la statistique du test de Fisher observée qui est égale à :

$$F_{obs} = \frac{SC_{reg}/ddl}{SC_{res}/ddl} = \frac{129,431/1}{442,292/(22-3-1=18)} \simeq 5,267.$$

Le quantile de la loi de Fisher critique lu dans la table des quantiles de la loi de Fisher à 95% est égal à :

$$F_{c,1,18} = 4,410.$$

La statistique du test de Fisher observée est plus grande que le quantile de la loi de Fisher critique. Donc on est dans la zone de rejet de l'hypothèse nulle $\mathcal{H}_0$. Donc on décide d'accepter l'hypothèse alternative $\mathcal{H}_1$, c'est-à-dire :

$$\beta_3 \neq 0.$$

Question 6. La valeur du coefficient R^2 associée à l'estimation du modèle spécifié en 5.a) est égale à :

$$R^2 = \frac{SC_{reg}}{SC_{tot}} = \frac{981,326}{1743,281} = 0,563.$$

La valeur du coefficient R^2 associée à l'estimation du modèle spécifié en 5.b) est égale à :

$$R^2 = \frac{SC_{reg}}{SC_{tot}} = \frac{1171,558}{1743,281} = 0,672.$$

La valeur du coefficient R^2 associée à l'estimation du modèle spécifié en 5.c) est égale à :

$$R^2 = \frac{SC_{reg}}{SC_{tot}} = \frac{1300,989}{1743,281} = 0,746.$$

Question 7. Le modèle qui semble le mieux adapté est le modèle 5.c) car ce modèle a le plus grand coefficient de détermination R^2.

Remarque : Pour l'instant à cette étape, le cours de choix de modèle n'a pas été fait, donc on ne calcule pas le R^2 adjusté pour voir quel serait le modèle le mieux approprié. Et si on appliquait le cours de choix de modèle, on calculerait le coefficient R^2 ajusté du second modèle, c'est-à-dire celui en 5.b) et le coefficient R^2 ajusté du troisième modèle, c'est-à-dire celui en 5.c)

Exercice IV.2. Avant de lire le corrigé de cet exercice, il serait préférable d'ouvrir le projet MINITAB afin de vérifier toutes les hypothèses du modèle.

Question 1. Quel pourcentage de variation dans la résistance à la rupture est expliquée par chacune des régressions ?

Pour la régression de la résistance à la rupture (Y) en fonction de l'épaisseur (X_1) :

$$R^2_{Y,X_1} = \frac{SC_{reg}}{SC_{tot}} = \frac{980,635}{1420,667} = 0,690.$$

Pour la régression de la résistance à la rupture (Y) en fonction de la densité (X_2) :

$$R^2_{Y,X_2} = \frac{SC_{reg}}{SC_{tot}} = \frac{643,565}{1420,667} = 0,453.$$

Pour la régression de la résistance à la rupture (Y) en fonction de l'épaisseur (X_1) et de la densité (X_2) :

$$R^2_{Y,X_1,X_2} = \frac{SC_{reg}}{SC_{tot}} = \frac{1204,858}{1420,667} = 0,848.$$

Question 2. Pour chaque régression, le tableau est le suivant :

	Carré moyen résiduel	Écart-type des résidus
Régression avec X_1	$44,003$	$6,633$
Régression avec X_2	$77,710$	$8,815$
Régression avec X_1, X_2	$23,979$	$4,897$

Question 3. Le tableau d'analyse de variance pour la régression comportant les deux variables explicatives est le suivant :

Source de variation	*ddl*	Somme des carrés	Carrés moyens	F_{obs}
Régression(X_1, X_2)	**2**	**1204,858**	**602,429**	**25,123**
Résiduelle	**9**	**215,809**	**23,979**	
Totale	**11**	**1420,667**		

Question 4. Tester au seuil de signification $\alpha = 5\%$, l'hypothèse nulle $\mathcal{H}_0 : \beta_1 = \beta_2 = 0$ contre l'hypothèse alternative $\mathcal{H}_1$: au moins un des $\beta \neq 0$. Quelle est votre conclusion ?

C'est pour cette question qu'il est important de regarder les résultats du projet MINITAB.

Calculons la statistique du test de Fisher observée qui est égale à :

$$F_{obs} = \frac{SC_{reg}/ddl}{SC_{res}/ddl} = \frac{1204,858/2}{215,809/(12-2-1=9)} \simeq 25,123.$$

Le quantile de loi de Fisher critique lu dans la table des quantiles de la loi de Fisher à 95% est égal à :

$$F_{c,2,9} = 4,260.$$

La statistique du test de Fisher observée est plus grande que le quantile de loi de Fisher critique. Donc on est dans la zone de rejet de l'hypothèse nulle $\mathcal{H}_0$. Donc on décide d'accepter l'hypothèse alternative $\mathcal{H}_1$, c'est-à-dire :

$$\exists j = 1, \text{ ou } 2, \quad \beta_j \neq 0.$$

Question 5. Dans le cas du modèle de régression ne comportant que l'épaisseur du matériau comme variable explicative, déterminer un intervalle de confiance à 95% pour β_1.

C'est également pour cette question qu'il est important de regarder les résultats du projet MINITAB.

Un intervalle de confiance à 95% **pour** β_1**, d'après le complément du cours 1 est égal à :**

$$[6,036 - 2,228 * 1,279; 6,036 + 2,228 * 1,279] = [3,187; 8,885].$$

Question 6. Avec l'intervalle de confiance calculé à la question 5.), peut-on affirmer, au seuil de signification $\alpha = 5\%$, que la régression est significative entre la résistance à la rupture et l'épaisseur du matériau ? Justifier votre conclusion ?

C'est aussi pour cette question qu'il est important de regarder les résultats du projet MINITAB.

La régression est significative entre la résistance à la rupture et l'épaisseur du matériau si le test de Student qui teste si $\beta_1 = 0$ **n'est pas vérifié.** Calculons la statistique du test de Student observée :

$$t_{obs} = \frac{6,036}{1,279} = 4,720.$$

Le quantile de la loi de Student critique lu dans une table des quantiles de la loi de Student à 95% est égal à :

$$t_{c,0,95} = 2,228$$

La statistique du test de Student observée est plus grande que le quantile de la loi de Student critique. Par conséquent on est dans la zone de rejet de l'hypothèse nulle $\mathcal{H}_0$. Donc on décide d'accepter l'hypothèse alternative $\mathcal{H}_1$. **Donc la régression est significative entre la résistance à la rupture et l'épaisseur du matériau.**

Remarque : On peut répondre plus rapidement en disant que l'intervalle de confiance calculé à la question précédente ne contient pas 0. Par conséquent **la régression est significative entre la résistance à la rupture et l'épaisseur du matériau.**

Question 7. Quel est l'apport marginal de X_2 lorsqu'elle est introduite à la suite de X_1 ?
L'apport marginal de la variable explicative X_2 lorsqu'elle est introduite à la suite de la variable explicative X_1 est égal à :

$$1204,858 - 980,635 = 224,223.$$

Question 8. Est-ce que la contribution marginale de la variable « densité du matériau », lorsqu'elle est introduite à la suite de la variable « épaisseur du matériau »est significative au seuil $\alpha = 5\%$? Utiliser les deux façons équivalentes d'effectuer ce test.

C'est aussi pour cette question qu'il est important de regarder les résultats du projet MINITAB.

Pour répondre à la question : « Est-ce que la contribution marginale de la variable « densité du matériau », lorsqu'elle est introduite à la suite de la variable « épaisseur du matériau »est significative au seuil $\alpha = 5\%$? », **il suffit de faire un test, soit un test de Fisher, soit un test de Student.**

« F partiel »	F_c	t_{obs}	t_c
9, 350	**5, 120**	**3, 058**	**2, 262**

Remarque : Pour obtenir la valeur 9, 350, on calcule la statistique du test de Fisher :

$$F_{obs} \text{ partiel } = \frac{224,223/1}{215,809/9} = 9,350.$$

Pour obtenir la valeur 5, 120, on lit dans une table des quantiles de la loi de Fisher :

$$F_{c,1,9} = 5,120.$$

Pour obtenir la valeur 3, 058, on calcule la statistique du test de Student :

$$t_{obs} = \frac{11,072}{3,621} = 3,058.$$

Pour obtenir la valeur 2, 262, on lit dans une table des quantiles de la loi de Student :

$$t_{c,9} = 2,262.$$

Question 9. On veut obtenir diverses estimations et prévisions de la résistance à la rupture. Quelle est, en moyenne, la résistance à la rupture de jouets dont l'épaisseur du matériau utilisé et la densité du matériau sont ceux indiqués dans le tableau suivant ?

C'est aussi pour cette question qu'il est important de regarder les résultats du projet MINITAB.

Epaisseur X_1	Densité X_2
4	$3,8$
3	$3,4$
4	$2,9$

Estimation de la résistance moyenne	Écart-type de l'estimation
$31,612$	$2,10$
$22,278$	$1,43$
$21,647$	$2,57$

Question 10. Entre quelles valeurs peut se situer la résistance moyenne à la rupture, pour des jouets dont l'épaisseur du matériau est $X_1 = 4$ et de densité $X_2 = 3,8$, si l'entreprise utilise un niveau de confiance à 95% ?

C'est pour cette question qu'il est important de regarder les résultats du projet MINITAB.

Un intervalle de confiance à 95% **est égal à :**

$$[31,612 - 4,751\,;\, 31,612 + 4,751] = [26,861\,;\, 36,363].$$

Question 11. Quelle est la marge d'erreur dans l'estimation effectuée à la question 10. ?

La marge d'erreur dans l'estimation effectuée à la question 10. est égale à

$$36,363 - 26,861 = 2 \times 4,751.$$

Exercice IV.3. Question 1. Complétons le tableau d'ANOVA :

Source de variation	Somme des carrés	*ddl*	Carrés moyens	F_{obs}
Régression	$1\,504,4$	**2**	$\mathbf{752,2}$	$\mathbf{38,37}$
Résiduelle	$\mathbf{176,4}$	**9**	$19,6$	
Totale	$1\,680,8$	**11**		

Question 2. Pour répondre à cette question, il faudrait s'assurer que les trois hypothèses du modèle sont vérifiées. Malheureusement on ne pourra pas le faire ici puisque nous ne connaissons pas les valeurs des observations. Donc nous allons supposer que les trois hypothèses sont vérifiées mais dans la pratique il faudrait les vérifier ABSOLUMENT.

Testons l'hypothèse nulle

$$\mathcal{H}_0 : \quad \beta_1 = \beta_2 = 0$$

contre l'hypothèse alternative

$$\mathcal{H}_1 : \quad \exists j = 1, \text{ ou } 2, \quad \beta_j \neq 0.$$

On a trouvé d'après le tableau d'ANOVA :

$$F_{obs} = 38,37.$$

On lit dans la table des quantiles de la loi de Fisher, à 95%, pour $\nu_1 = 2$ et $\nu_2 = 9$:

$$F_{c,2,9} = 4,26.$$

Comme $F_{obs} > F_c$, on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ et d'accepter l'hypothèse alternative $\mathcal{H}_1$, c'est-à-dire :

$$\exists j = 1 \text{ ou } 2, \quad \beta_j \neq 0.$$

Remarque : À cette étape, et avec un test de Fisher, on ne sait pas dire qu'elle est la ou les variable(s) qu'il faut conserver dans le modèle.

Question 3. Calculons le coefficient de détermination R^2 du modèle :

$$R^2 = \frac{SC_{reg}}{SC_{tot}} = \frac{1\,504,4}{1\,680,8} = 0,895.$$

Question 4. Donnons une estimation de la variance de la variable résiduelle ε :

$$s^2 = \frac{||y - \hat{y}||^2}{n - p} = 19,6.$$

Exercice IV.4. Cet exercice a été traité avec MINITAB, car il demande beaucoup de calculs qui ne sont pas réalisables avec une simple calculette. Le corrigé est mis à part. Il faut alors consulter le document intitulé « EXERCICE4-TD4.MPJ ».

Exercice IV.5. Pour répondre aux deux questions qui vont suivre, il faudrait s'assurer que les trois hypothèses du modèle sont vérifiées. Malheureusement on ne pourra pas le faire ici puisque nous ne connaissons pas les valeurs des observations. Donc nous allons supposer que les trois hypothèses sont vérifiées mais dans la pratique il faudrait les vérifier ABSOLUMENT.

Question 1. Est-ce que la régression est significative dans son ensemble ? Utiliser $\alpha = 0,05$.

Pour cela, on fait un test de Fisher. On teste l'hypothèse nulle

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

contre l'hypothèse alternative

$$\mathcal{H}_1 : \exists j = 1, 2, 3 \text{ ou } 4, \quad \beta_j \neq 0.$$

On calcule la statistique du test de Fisher :

$$F_{obs} = \frac{21\,392,50}{95,07} = 225,03.$$

On lit dans une table des quantiles de la loi de Fisher, à 95%, avec $\nu_1 = 4$ et $\nu_2 = 15$

$$F_{c,4,15} = 3,06.$$

Comme $F_{obs} > F_c$, on décide de rejetter l'hypothèse nulle $\mathcal{H}_0$ et d'accepter l'hypothèse alternative $\mathcal{H}_1$. Donc la régression est très significative dans son ensemble.

Question 2. Est-ce que l'affirmation de votre collègue est vraisemblable au seuil de signification $\alpha = 0,05$? Effectuer le test approprié.
Pour cela, on fait un test de Fisher. On teste l'hypothèse nulle

$$\mathcal{H}_0 : \beta_3 = \beta_4 = 0$$

contre l'hypothèse alternative

$$\mathcal{H}_1 : \exists j = 3 \text{ ou } 4, \quad \beta_j \neq 0.$$

On calcule une statistique du test de Fisher partiel :

$$\begin{aligned} F_{obs} &= \frac{(SC(H_1)_{reg} - SC(H_0)_{reg})/(p-1-k)}{(SC(H_1)_{res})/(n-p)} \\ &= \frac{85\,570 - 62\,983}{1\,426} \times \frac{20-5}{5-1-2} = 118,79. \end{aligned}$$

On lit dans une table des quantiles de la loi de Fisher, à 95%, avec $\nu_1 = 2$ et $\nu_2 = 15$

$$F_{c,2,15} = 3,68.$$

Comme $F_{obs} > F_c$, on décide de rejetter l'hypothèse nulle $\mathcal{H}_0$ et d'accepter l'hypothèse alternative $\mathcal{H}_1$. L'affirmation de notre collègue n'est donc pas vraisemblable au seuil de signification $\alpha = 0,05$.

Exercice IV.6. Question 1. Dans quelle proportion, notée P, la variation non expliquée par X_1 est réduite avec l'ajout de X_2 dans l'équation de régression ?

Il faut d'abord calculer la proportion de la variation non expliquée par X_1. Elle est égale à :

$$(1 - 0,548) \times 100 = 45,2\%.$$

Ensuite il faut calculer la proportion de la variation non expliquée par X_1 et par X_2. Elle est égale à :

$$(1 - 0,683) \times 100 = 31,7\%.$$

Ensuite on résoud une équation à une inconnue :

$$45,2 - (45,2 \times P) = 31,7\%.$$

En résolvant cette équation, on obtient :

$$P = 29,86\%.$$

Donc la proportion P cherchée est égale à $29,86\%$.

Question 2. Déterminer la somme des carrés résiduelle lorsque les variables explicatives X_1 et X_2 sont dans l'équation de régression.

La somme de carrés résiduelle lorsque les variables explicatives X_1 et X_2 sont dans l'équation de régression est égale à :

$$SC_{res} = \mathrm{s}^2 \times (n - p) = 1,6352^2 \times (20 - 3) = 45,45.$$

Question 3. Quelle est la somme de carrés de régression attribuable à X_3 lorsqu'on ajoute cette variable à la suite de X_1 et X_2 ?

Pour répondre à cette question, introduisons quelques notations.

Accroissement de la variation expliquée par l'ajout de la variable explicative X_3 à la suite de la variable explicative X_1 et de la variable explicative X_2 :

$$SC_{reg}(X_1, X_2, X_3) - SC_{reg}(X_1, X_2) = SC_{reg}(X_3|X_1, X_2),$$

soit dans une proportion de

$$\frac{SC_{reg}(X_1, X_2, X_3) - SC_{reg}(X_1, X_2)}{SC_{res}(X_1, X_2)} = \frac{SC_{reg}(X_3|X_1, X_2)}{SC_{res}(X_1, X_2)} = r^2_{Y_{3.1,2}}$$

qui peut également s'écrire, si on divise chaque membre par SC_{tot}

$$\frac{\dfrac{SC_{reg}(X_1, X_2, X_3)}{SC_{tot}} - \dfrac{SC_{reg}(X_1, X_2)}{SC_{tot}}}{\dfrac{SC_{res}(X_1, X_2)}{SC_{tot}}} = \frac{R^2_{Y.1,2,3} - R^2_{Y.1,2}}{1 - R^2_{Y.1,2}} = r^2_{Y_{3.1,2}}.$$

Cette formule donne le coefficient de détermination partielle entre la variable expliquée Y et la variable explicative X_3, étant donné que les variables explicatives X_1 et X_2 sont déjà dans l'équation de régression.

On peut maintenant calculer la somme de carrés de régression attribuable à la variable explicative X_3 lorsqu'on ajoute cette variable à la suite des variables explicatives X_1 et X_2. On a :

$$\begin{aligned} SC_{reg}(X_3|X_1,X_2) &= r^2_{Y3.1,2} \times SC_{res}(X_1,X_2) \\ &= \frac{R^2_{Y.1,2,3} - R^2_{Y.1,2}}{1 - R^2_{Y.1,2}} \times SC_{res}(X_1,X_2) \\ &= \frac{0,940 - 0,683}{1 - 0,683} \times (1,6352)^2(20-3) \\ &= 36,8523. \end{aligned}$$

Question 4. Quelle est la somme des carrés de régression attribuable à X_4 lorsqu'on ajoute cette variable à la suite de X_1, X_2 et X_3 ?

On procède de la même manière que précedemment. On a :

$$\begin{aligned} SC_{reg}(X_4|X_1,X_2,X_3) &= r^2_{Y4.1,2,3} \times SC_{res}(X_1,X_2,X_3) \\ &= \frac{R^2_{Y.1,2,3,4} - R^2_{Y.1,2,3}}{1 - R^2_{Y.1,2,3}} \times SC_{res}(X_1,X_2,X_3) \\ &= \frac{0,959 - 0,940}{1 - 0,940} \times (0,7349)^2(20-4) \\ &= 2,7364. \end{aligned}$$

Exercice IV.7. Question 1. Déterminer les coefficients de détermination multiple suivants :

$$R^2_{Y,2}, R^2_{Y,2,3}, R^2_{Y,2,3,4}, R^2_{Y,2,3,4,1}.$$

Comme on s'appuie sur les mêmes données de l'exercice précédent, la somme des carrés totale ne change pas. Calculons d'abord cette somme des carrés totale avec les données de l'exercice précédent. On a

$$R^2_{Y.1,2,3,4} = 1 - \frac{SC_{res}}{SC_{tot}} = 0,959$$

On en déduit que :

$$\frac{SC_{res}}{SC_{tot}} = 1 - 0,959 = 0,041$$

ou encore

$$SC_{tot} = \frac{SC_{res}}{0,041} = \frac{0,6281^2 \times 15}{0,041} = 144,333.$$

Et par conséquent, on en déduit que :

$$\begin{aligned}
R^2_{Y.2} &= 1 - \frac{2,8116^2 \times 18}{144,333} = 0,014. \\
R^2_{Y.2,3} &= 1 - \frac{2,5486^2 \times 17}{144,333} = 0,235. \\
R^2_{Y.2,3,4} &= 1 - \frac{2,4935^2 \times 16}{144,333} = 0,311. \\
R^2_{Y.2,3,4,1} &= 1 - \frac{0,6281^2 \times 15}{144,333} = 0,959.
\end{aligned}$$

Question 2. Déterminer la somme des carrés résiduelle lorsque les variables explicatives X_2 et X_3 sont dans l'équation de régression.

Déterminons la somme des carrés résiduelle lorsque les variables explicatives X_2 et X_3 sont dans l'équation de régression.

$$SC_{res} = 2,5486^2 \times (20-3) = 110,421.$$

Question 3. Quelle est la somme des carrés de régression attribuable à X_4 lorsqu'on ajoute cette variable à la suite de X_2 et X_3 ?

On procède de la même manière que dans l'exercice précedent. On a :

$$\begin{aligned}
SC_{reg}(X_4|X_2,X_3) &= r^2_{Y4.2,3} \times SC_{res}(X_2,X_3) \\
&= \frac{R^2_{Y.2,3,4} - R^2_{Y.2,3}}{1 - R^2_{Y.2,3}} \times SC_{res}(X_2,X_3) \\
&= \frac{0,311 - 0,235}{1 - 0,235} \times (2,5486)^2(20-3) \\
&= 10,970.
\end{aligned}$$

Question 4. Quelle est la somme des carrés de régression attribuable à X_1 lorsqu'on ajoute cette variable à la suite de X_2, X_3 et X_4 ?

On procède de la même manière que précedemment. On a :

$$\begin{aligned}
SC_{reg}(X_1|X_2,X_3,X_4) &= r^2_{Y1.2,3,4} \times SC_{res}(X_2,X_3,X_4) \\
&= \frac{R^2_{Y.2,3,4,1} - R^2_{Y.2,3,4}}{1 - R^2_{Y.2,3,4}} \times SC_{res}(X_2,X_3,X_4) \\
&= \frac{0,959 - 0,311}{1 - 0,311} \times (2,2435)^2(20-4) \\
&= 75,740.
\end{aligned}$$