T. D. n° 7
Régression linéaire multiple

L'exercice 1 *est issu du livre « Probabilités, Statistique et technique de régression » de Gérald Baillargeon, Les éditions SMG et l'exercice* 2 *est issu du livre « Analyse de régression appliquée » de Yadolah Dodge, Édition Dunod.*

Exercice 1. L'entreprise CITRON fabrique un matériau en matière plastique qui est utilisé dans la fabrication de jouets. Le département de contrôle de qualité de l'entreprise a effectué une étude qui a pour but d'établir dans quelle mesure la résistance à la rupture (en kg/cm^2) de cette matière plastique pouvait être affectée par l'épaisseur du matériau ainsi que la densité de ce matériau. Une étude réalisée au préalable a montré que la résistance du jouet suit une loi normale. Douze essais ont été effectués et les résultats sont présentés dans le tableau ci-dessous.

Essai numéro	Résistance à la rupture Y_i	Épaisseur du matériau X_{i_1}	Densité X_{i_2}
1	$37,8$	4	$4,0$
2	$22,5$	4	$3,6$
3	$17,1$	3	$3,1$
4	$10,8$	2	$3,2$
5	$7,2$	1	$3,0$
6	$42,3$	6	$3,8$
7	$30,2$	4	$3,8$
8	$19,4$	4	$2,9$
9	$14,8$	1	$3,8$
10	$9,5$	1	$2,8$
11	$32,4$	3	$3,4$
12	$21,6$	4	$2,8$

Diverses analyses ont été effectuées sur ordinateur et nous les résumons dans les trois tableaux suivants :

Régression de la résistance à la rupture Y en fonction de l'épaisseur X_1

Coefficients $\widehat{\beta}_j$	Erreurs-types $s(\widehat{\beta}_j)$
$\widehat{\beta}_0 = 3,523$	
$\widehat{\beta}_1 = 6,036$	$s(\widehat{\beta}_1) = 1,279$

Source de variation	Somme des carrés	*ddl*
Régression X_1	$980,63$	1
Erreur	$440,03$	10

Régression de la résistance à la rupture Y en fonction de la densité X_2

Coefficients $\widehat{\beta}_j$	Erreurs-types $s(\widehat{\beta}_j)$
$\widehat{\beta}_0 = -36,373$	
$\widehat{\beta}_2 = +17,464$	$s(\widehat{\beta}_2) = 6,069$

Source de variation	Somme des carrés	*ddl*
Régression X_2	$643,57$	1
Erreur	$777,10$	10

Régression de la résistance à la rupture Y en fonction de l'épaisseur X_1 et de la densité X_2

Coefficients $\widehat{\beta}_j$	Erreurs-types $s(\widehat{\beta}_j)$
$\widehat{\beta}_0 = -30,081$	
$\widehat{\beta}_1 = +04,905$	$s(\widehat{\beta}_1) = 1,014$
$\widehat{\beta}_2 = +11,072$	$s(\widehat{\beta}_2) = 3,621$

Source de variation	Somme des carrés	*ddl*
Régression (X_1, X_2)	$1204,86$	2
Erreur	$215,81$	9

1. Quel pourcentage de variation dans la résistance à la rupture est expliquée par chacune des régressions ?
2. Pour chaque régression linéaire, compléter le tableau suivant :

	Carré moyen des résidus	Écart-type des résidus
Régression due à X_1		
Régression due à X_2		
Régression due à (X_1, X_2)		

3. Compléter le tableau d'analyse de la variance suivant pour la régression comportant les deux variables explicatives.

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F_{obs}
Régression due à (X_1, X_2)				
Résiduelle				
Totale				

4. Tester au seuil de signification $\alpha = 5\%$, l'hypothèse nulle

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

contre l'hypothèse alternative

$$\mathcal{H}_1 : \text{ au moins un des } \beta \neq 0.$$

Quelle est votre conclusion ?
5. Quel est l'apport marginal de la variable « densité du matériau » notée X_2 lorsqu'elle est introduite à la suite de la variable « épaisseur du matériau » notée X_1 ?
6. Est-ce que l'apport marginal de la variable « densité du matériau »notée X_2, lorsqu'elle est introduite à la suite de la variable « épaisseur du matériau » notée X_1, est significative au seuil de signification $\alpha = 5\%$? Pour répondre à cette question, utiliser le test de Student ou le test de Fisher partiel.
7. Nous voulons obtenir diverses estimations 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 celles indiquées dans le

tableau suivant ?

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

Estimation de la résistance moyenne	Écart-type de l'estimation
	$2,10$
	$1,43$
	$2,57$

8. 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 la densité du matériau est $X_2 = 3,8$, si l'entreprise utilise un niveau de confiance de 95% ?
9. Quelle est la marge d'erreur dans l'estimation effectuée à la question 8. ?
10. Nous désirons un intervalle de prédiction de la résistance à la rupture pour un jouet ayant comme épaisseur du matériau $X_1 = 4$ et densité du matériau égale à $X_2 = 3,9$. Quel est cet intervalle au niveau de confiance de 95% ?

Exercice 2. Les données proposées dans cet exercice présentent le taux de décès par crise cardiaque chez les hommes de 55 à 59 ans dans différents pays.
Les variables sont les suivantes :
— Y : 100 [log(nombre de décès par crise cardiaque pour 100 000 hommes de 55 à 59 ans) -2] ;
— X_1 : téléphones par millier d'habitants ;
— X_2 : calories grasses en pourcentage du total des calories ;
— X_3 : calories provenant de protéines animales en pourcentage du total des calories.
Une étude réalisée au préalable a montré que le taux de décès par crise cardiaque chez les hommes de 55 à 59 suit une loi normale.

1. Régresser Y sur X_1 et tester la signification de cette régression linéaire simple.
2. Trouver l'équation de la régression linéaire multiple de Y sur X_1 et X_2.
3. Effectuer un test conjoint d'hypothèse nulle $\mathcal{H}_0 : \beta_1 = \beta_2 = 0$.
4. Tester si l'adjonction de la variable X_2 à l'équation trouvée à la question 1. a significativement amélioré l'estimation.
5. Construire la régression linéaire multiple de Y sur X_1, X_2 et X_3.
6. Donner les limites de l'intervalle de confiance à 95% pour β_3 dans cette équation.
7. Donner les limites de l'intervalle de confiance à 95% pour $\widehat{Y}$ au point $X_1 = 221$, $X_2 = 39$ et $X_3 = 7$.
8. Tester si X_2 et X_3 ensemble apportent quelque chose à la régression linéaire simple de Y sur X_1.
9. Régresser X_1 sur X_2 et X_3.
10. Donner les limites de l'intervalle de confiance à 95% pour les coefficients de la régression linéaire de X_1 sur X_3.

Observation i	Pays	X_1 x_{1i}	X_2 x_{2i}	X_3 x_{3i}	Y x_{4i}
1	Australie	124	33	8	81
2	Autriche	49	31	6	55
3	Canada	181	38	8	80
4	Ceylan	4	17	2	24
5	Chili	22	20	4	78
6	Danemark	152	39	6	52
7	Finlande	75	30	7	88
8	France	54	29	7	45
9	Allemagne	43	35	6	50
10	Irlande	41	31	5	69
11	Israël	17	23	4	66
12	Italie	22	21	3	45
13	Japon	16	8	3	24
14	Mexique	10	23	3	43
15	Pays-Bas	63	37	6	38
16	Nouvelle-Zélande	170	40	8	72
17	Norvège	125	38	6	41
18	Portugal	12	25	4	38
19	Suède	221	39	7	52
20	Suisse	171	33	7	52
21	Grande-Bretagne	97	38	6	66
22	États-Unis	254	39	8	89