Feuille de Travaux Pratiques sous Minitab n$^{\text{o}}$ 3

Les exercices 1, 2, 5, 6 *sont issus du livre "Probabilités, Statistique et technique de régression" de Gérald Baillargeon, Les éditions SMG et les exercices* 3, 4 *sont issus du livre "Analyse de régression appliquée" de Yadolah Dodge, Edition Dunod.*

Exercice III.1. Un bureau de conseil en ressources humaines a effectué une étude sur le niveau d'anxiété Y mesuré sur une échelle de 1 à 50 de cadres d'entreprises au cours d'une période de 2 semaines. On veut examiner si les facteurs suivants peuvent influencer sur le niveau d'anxiété des cadres :
- X_1 : pression artérielle systolique
- X_2 : test évaluant les capacités managériales
- X_3 : niveau de satisfaction du poste occupé.

Le tableau d'analyse de variance indique l'apport de chaque variable introduite dans l'ordre indiqué et ceci pour 22 cadres.

Source de variation	Somme des carrés	*ddl*
Régression X_1	$981,326$	1
Régression X_2	$190,232$	1
Régression X_3	$129,431$	1
Résiduelle	$442,292$	18
Totale	$1743,281$	21

1. Quelle est la somme des carrés due à la régression pour l'ensemble des 3 variables explicatives ?
2. Quelle proportion de la variation dans le niveau d'anxiété est expliquée par les 3 variables explicatives ?
3. Peut-on conclure que dans l'ensemble les 3 variables explicatives ont un effet significatif sur le niveau d'anxiété ? Utiliser un seuil de signification $\alpha = 0,05$. Préciser les hypothèses que l'on veut tester.
4. Si on ne tient compte que de la variable explicative X_1, quel serait alors le tableau d'analyse de variance correspondant ?

Source de variation	Somme des carrés	*ddl*
Régression due à X_1	981.326	
Résiduelle		
Totale		

5. Tester les hypothèses suivantes au seuil $\alpha = 0,05$ en utilisant un rapport F approprié :

   a) $H_0 : \beta_1 = 0$ dans le modèle $Y = \beta_0 + \beta_1 X_1 + \varepsilon$;

   b) $H_0 : \beta_2 = 0$ dans le modèle $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$;

   c) $H_0 : \beta_3 = 0$ dans le modèle $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$.

6. Quelle est la valeur du R^2 associée à l'estimation de chaque modèle spécifié en 5. ?
7. Lequel des trois modèles semble le mieux approprié pour expliquer les fluctuations du niveau d'anxiété des cadres d'entreprises ?

Exercice III.2. L'entreprise SINTRON fabrique un matériau en matière plastique qui est utilisé dans la fabrication de jouets. Le département de contrôle de la 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. Douze essais ont été effectués et les résultats sont présentés dans le tableau ci-dessous.

Essai Numéro	Résistance à la rupture Y_i	Epaisseur du matériau X_{i_1}	Densité X_{i_2}
1	37.8	4	4
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 $\hat{\beta}_j$	Erreurs-types $s(\hat{\beta}_j)$
$\hat{\beta}_0 = 3,5226$	
$\hat{\beta}_1 = 6,0359$	$s(\hat{\beta}_1) = 1,2786$

Source de variation	Somme des carrés	*ddl*
Régression X_1	$980,635$	1
Résiduelle	$440,032$	10

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

Coefficients $\hat{\beta}_j$	Erreurs-types $s(\hat{\beta}_j)$
$\hat{\beta}_0 = -36,3726$	
$\hat{\beta}_2 = 17,4644$	$s(\hat{\beta}_2) = 6,0687$

Source de variation	Somme des carrés	*ddl*
Régression X_2	$643,565$	1
Résiduelle	$777,102$	10

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

Coefficients $\hat{\beta}_j$	Erreurs-types $s(\hat{\beta}_j)$
$\hat{\beta}_0 = -30,0809$	
$\hat{\beta}_1 = 4,9047$	$s(\hat{\beta}_1) = 1,037$
$\hat{\beta}_2 = 11,0721$	$s(\hat{\beta}_2) = 3,621$

Source de variation	Somme des carrés	*ddl*
Régression (X_1, X_2)	$1204,858$	2
Résiduelle	$215,809$	9

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

	Carré moyen résiduel	Écart-type des résidus
Régression avec X_1		
Régression avec X_2		
Régression avec X_1, X_2		

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

Source de variation	Somme des carrés	*ddl*	Carrés moyens	F
Régression avec X_1, X_2				
Résiduelle				
Totale				

4. Tester au seuil de signification $\alpha = 0,05$, l'hypothèse
$$H_0 : \beta_1 = \beta_2 = 0$$
contre l'hypothèse alternative
$$H_1 : \text{ au moins un des } \beta \neq 0.$$
Quelle est votre conclusion ?
5. Dans le cas du modèle de régression ne comportant que l'épaisseur du matériau comme variable explicative, déterminer un intervalle de confiance à $0,95$ pour β_1.
6. Avec l'intervalle de confiance calculé en 5., peut-on affirmer, au seuil de signification $\alpha = 0,05$, que la régression est significative entre la résistance à la rupture et l'épaisseur du matériau ? Jusitifier votre conclusion ?
7. Quel est l'apport marginal de X_2 lorsqu'elle est introduite à la suite de X_1 ?
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 $\alpha = 0,05$? Utiliser les deux façons équivalentes d'effectuer ce test.

“F partiel”	F_{th}	t_{obs}	t_{th}

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 ?

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
	2,1
	1,43
	2,57

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 de $0,95$?
11. Quelle est la marge d'erreur dans l'estimation effectuée en 10. ?
12. On désire un intervalle de précision de la résistance à la rupture pour un jouet ayant comme épaisseur de matériau et de densité ceux précisés en 10. Quel est cet intervalle au niveau $0,95$?

Exercice III.3. On utilise le modèle de régression linéaire multiple :

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

1. Compléter le tableau d'analyse de variance suivant :

Source de variation	Somme des carrés	*ddl*	Carrés moyens	F
Régression	1 504,4			
Résiduelle			19,6	
Totale	1 680,8			

2. Tester l'hypothèse

$$H_0 : \beta_1 = \beta_2 = 0.$$

3. Quel est le R^2 du modèle ?
4. Donner une estimation de la variance de ε.

Exercice III.4. Les données proposées dans cet exercice présentent le taux de décès par attaque cardiaque ches 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.

1. Régresser Y sur X_1 et tester la signification de cette régression simple.
2. Trouver l'équation de la régression multiple de Y sur X_1 et X_2.
3. Effectuer un test conjoint, d'hypothèse nulle $\beta_1 = \beta_2 = 0$.
4. Tester si l'adjonction de la variable X_2 à l'équation trouvée en 1. a significativement amélioré l'estimation.
5. Construire la régression multiple de Y sur X_1, X_2 et X_3.
6. Donner les limites de l'intervalle de confiance à $0,95$ pour β_3 dans cette équation.
7. Donner les limites de l'intervalle de confiance à $0,95$ pour $\hat{Y}$ 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 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 $0,95$ pour les coefficients de la régression 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	Israel	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

Exercice III.5. Dans une étude de régression multiple comportant quatre variables explicatives X_1, X_2, X_3, X_4, on a obtenu le tableau d'analyse de variance suivant, et ceci pour 20 observations.

Source de variation	Somme des carrés	*ddl*	Carrés moyens
Régression (X_1, X_2, X_3, X_4)	85 570	4	21 392,5
Résiduelle	1 426	15	95,067
Totale	86 996	19	

1. Est-ce que la régression est significative dans son ensemble ? Utiliser $\alpha = 0,05$.
2. Une de vos collègues mentionne que les variables X_3 et X_4 sont inutiles dans le modèle de régression. Une autre analyse de régression ne comportant cette fois que les variables explicatives X_1 et X_2 conduit au tableau d'analyse de variance suivant.

Source de variation	Somme des carrés	*ddl*	Carrés moyens
Régression (X_1, X_2)	62 983	2	31 491,5
Résiduelle	24 013	17	1 412,53
Totale	86 996	19	

Est-ce que l'affirmation de votre collègue est vraisemblable au seuil de signification $\alpha = 0,05$? Effectuer le test approprié.

Exercice III.6. Dans une étude de régression multiple comportant quatre variables explicatives et 20 observations, on a introduit, dans l'ordre, les variables X_1, X_2, X_3, X_4.

Variables explicatives dans l'équation	Variable additionnelle	Écart-type des résidus	Proportion de la variation expliquée
Aucune		$s_Y = 2,7456$	
X_1	X_1	1,8968	0,548
X_1, X_2	X_2	1,6352	0,683
X_1, X_2, X_3	X_3	0,7349	0,940
X_1, X_2, X_3, X_4	X_4	0,6281	0,959

1. Dans quelle proportion la variation non expliquée par X_1 est réduite avec l'ajout de X_2 dans l'équation de régression ?
2. Quelle est la corrélation partielle entre Y et X_3 après s'être débarrassé de l'influence des variables explicatives X_1 et X_2 sur Y et X_3 ? On considère que le coefficient de régression $\hat{\beta}_3$ est négatif.
3. Déterminer la somme de carrés résiduelle lorsque les variables explicatives X_1 et X_2 sont dans l'équation de régression.
4. 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 ?
5. Quelle est la somme de carrés de régression attribuable à X_4 lorsqu'on ajoute cette variable à la suite de X_1, X_2 et X_3 ?

Exercice III.7. Supposons que dans l'exercice précédent, les variables ont été introduites dans l'ordre suivant : X_2, X_3, X_4, X_1. On a alors le tableau suivant :

Variables explicatives dans l'équation	Variable additionnelle	Écart-type des résidus
Aucune		$s_Y = 2,7456$
X_2	X_2	2,8116
X_2, X_3	X_3	2,5486
X_2, X_3, X_4	X_4	2,4935
X_2, X_3, X_4, X_1	X_1	0,6281

1. Déterminer les coefficients de détermination multiple suivants : $R^2_{Y,2}$, $R^2_{Y,2,3}$, $R^2_{Y,2,3,4}$, $R^2_{Y,2,3,4,1}$.
2. Déterminer la somme de carrés résiduelle lorsque les variables explicatives X_2 et X_3 sont dans l'équation de régression.
3. Déterminer les coefficients de détermination partielle suivants : $r^2_{Y,3,2}$, $r^2_{Y,4,2,3}$, $r^2_{Y,1,2,3,4}$.
4. Quelle est la corrélation partielle entre Y et X_3 après s'être débarrassé de l'influence de X_2 sur Y et X_3. On suppose que le coefficient de β_3 peut être négatif.
5. Quelle est la somme de carrés de régression attribuable à X_4 lorsqu'on ajoute cette variable à la suite de X_2 et X_3.
6. Quelle est la somme de carrés de régression attribuable à X_1 lorsqu'on ajoute cette variable à la suite de X_2, X_3 et X_4.