Feuille de Travaux Pratiques sous Minitab n° 2

Exercice II.1. La taille d'un athlète peut jouer un rôle important dans ses résultats en saut en hauteur. Les données utilisées ici présentent donc la taille et la performance de 20 champions du monde.

Observation	Nom	Taille	Performance
1	Jacobs (EU)	1,73	2,32
2	Noji (EU)	1,73	2,31
3	Conway (EU)	1,83	2,40
4	Matei (Roum)	1,84	2,40
5	Austin (EU)	1,84	2,40
6	Ottey (Jam)	1,78	2,33
7	Smith (GB)	1,84	2,37
8	Carter (EU)	1,85	2,37
9	McCants (EU)	1,85	2,37
10	Sereda (URSS)	1,86	2,37
11	Grant (GB)	1,85	2,36
12	Paklin (URSS)	1,91	2,41
13	Annys (Bel)	1,87	2,36
14	Sotomayor (Cuba)	1,96	2,45
15	Sassimovitch (URSS)	1,88	2,36
16	Zhu Jianhua (Chine)	1,94	2,39
17	Brumel (URSS)	1,85	2,28
18	Sjoeberg (Suède)	2,00	2,42
19	Yatchenko (URSS)	1,94	2,35
20	Povarnitsine (URSS)	2,01	2,40

1. À partir de l'échantillon proposé, utiliser la méthode des moindres carrés pour estimer les paramètres de la régression linéaire :

$$(\text{Performance}) = \beta_0 + \beta_1.(\text{Taille}) + \varepsilon$$

2. Compléter le tableau d'analyse de la variance (dit aussi tableau d'ANOVA) :

Source de variation	Degrés de liberté	Somme des carrés	Moyenne des carrés	F_c
Régression Résiduelle				
Totale				

3. Quel pourcentage de la variation totale des performances est expliqué par la variable taille ? Que pensez-vous de ce résultat ? Que faudrait-il faire en tant que chargé de cette étude ?

Exercice II.2. On dispose des données suivantes au sujet de 2 variables d'intérêt X et Y :

x_i	7	9	9	10	13	17	19	20	21	25
y_i	5	4	6	4	1	2	0	1	1	0

$$\sum x_i = 150; \quad \sum y_i = 24; \quad \sum x_i y_i = 253; \quad \sum x_i^2 = 2\,596; \quad \sum y_i^2 = 100.$$

On se réfère au modèle linéaire :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

1. Estimer les paramètres β_0 et β_1 par la méthode des moindres carrés.
2. Pour chacun de ces 2 paramètres, trouver un intervalle de confiance avec un niveau de confiance de 99%.
3. Soit $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ $(i = 1, \cdots, n)$ où β_0 et β_1 sont les estimateurs de β_0 et β_1 obtenus en 1). Démontrer que l'on a $\sum \hat{y}_i = \sum y_i$, par deux méthodes (mathématique, et avec Minitab).
4. Calculer les intervalles de confiance pour les $\mu_Y(X)$.
5. Représenter graphiquement les points $(x_i; y_i)$, la droite de régression et l'ensemble des intervalles de confiance pour les $\mu_Y(X)$.

Exercice II.3. Exemples :

Source de variation	Degrés de liberté	Somme des carrés	Moyenne des carrés	F_c	F_t	R^2
Régression résiduelle	1 12	$501,76$ $794,90$	$501,76$ $66,24$	$7,575$	$4,75$	$0,387$
Totale	13	$1\,296,66$				

Source de variation	Degrés de liberté	Somme des carrés	Moyenne des carrés	F_c	F_t	R^2
Régression résiduelle	1 4	$34,186$ $3,148$	$34,186$ $0,787$	$43,44$	$7,71$	$0,916$
Totale	5	$37,333$				

Source de variation	Degrés de liberté	Somme des carrés	Moyenne des carrés	F_c	F_t	R^2
Régression résiduelle	1 48	$179,76$ $309,56$	$179,76$ $6,45$	$27,87$	$4,08$	$0,367$
Totale	49	$489,32$				

1. En comparant dans les exemples ci-dessus, les liens entre les valeurs F_c, F_t et R^2, quelles sont, selon vous, les meilleures régressions ?
2. Avant de calculer le R^2, en n'utilisant que les valeurs F_c et F_t, quelle règle pourrait-on énoncer pour réperer rapidement une bonne analyse de régression ?

Exercice II.4. Soient les données présentées dans le tableau ci-dessous. Il s'agit du nombre de calories consommées par jour et du pourcentage de population agricole dans 11 pays.

Observation i	Pays	% Population agricole	Calories par jour et par personne
1	Suisse	$4,0$	3 432
2	France	$5,7$	3 273
3	Suède	$4,9$	3 049
4	USA	$3,0$	3 642
5	Ex-URSS	$14,8$	3 394
6	Chine	$69,6$	2 628
7	Inde	$63,8$	2 204
8	Brésil	$26,2$	2 643
9	Pérou	$38,3$	2 192
10	Algérie	$24,7$	2 687
11	Ex-Zaire	$65,7$	2 159

1. Représenter graphiquement Y en fonction de X.
2. Estimer les paramètres β_0 et β_1 du modèle :
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$
3. Construire le tableau d'analyse de la variance correspondant à cette régression.
4. Construire un intervalle de confiance à 95% autour de la droite de régression.
5. Représenter sur le graphique de la question (1) la droite de régression et l'intervalle de confiance calculé à la question (4).

Exercice II.5. Un certain composant électronique est fabriqué une fois par mois par l'entreprise Micro-Systèmes. La quantité fabriquée varie avec la demande du marché. Dans le but de planifier la production et d'établir certaines normes sur le nombre d'hommes-minutes exigés pour la production de différents lots de ce composant électronique, le responsable de la productoin a relevé l'information suivante pour 15 cédules de production. Le nombre d'hommes-minutes est identifié par Y et la quantité fabriquée par X.

Y_i	150	192	264	371	300	358	192	134	242	238	226	302	340	182	169
X_i	35	42	64	88	70	85	40	30	55	60	51	72	80	44	39

1. Quelle serait la première étape à franchir avant d'aborder tout calcul préliminaire ?
2. Le responsable de la production envisage d'utiliser le modèle linéaire simple comme modèle prévisionnel. Spécifier ce modèle et bien identifier chacune des composantes du modèle dans le contexte de ce problème.
3. Déterminer l'équation de régression.
4. D'après l'équation de régression, si le nombre d'unités à fabriquer augment de 10, quelle sera l'augmentation correspondante du nombre moyen d'hommes-minutes requis ?
5. En l'absence de l'information que nous donne la quantité à fabriquer, quelle serait une bonne estimation du nombre d'hommes-minutes requis ?

6. Quelle correction peut-il apporter à son estimation du nombre moyen d'hommes-minutes requis en introduisant la connaissance de X dans son analyse ?
7. Calculer $s(\beta_1)$ et tester les hypothèses suivantes avec la loi de Student.
$$\mathrm{H}_0 : \beta_1 = 0, \quad \mathrm{H}_1 : \beta_1 \neq 0.$$
8. Calculer la variation qui est expliquée par la droite de régression et la variation qui est inexpliquée par la droite.
9. Déterminer le pourcentage de variation qui est expliqué par la droite de régression.
10. Calculer une estimation du nombre moyen d'hommes-minutes requis pour : $X_h = 42; X_h = 57; X_h = 72$.
11. Pour quelle quantité X_n, l'estimation du nombre moyen d'hommes-minutes requis serait-elle la plus précise ?
12. Entre quelles valeurs peut se situer le vrai nombre moyen d'hommes-minutes requis pour les lots dont la quantité a été déterminée par k ? Utiliser un niveau de confiance de 95%.
13. Quelle est la marge d'erreur dans l'estimation effectuée en 12.