Feuille de Travaux Pratiques sous Minitab n$^{\text{o}}$ 1

Exercice I.1. On donne les couples d'observations suivants :

x_i	18	7	14	31	21	5	11	16	26	29
y_i	55	17	36	85	62	18	33	41	63	87

a) Tracer le diagramme de dispersion des couples $(x_i; y_i)$. Peut-on soupçonner une liaison linéaire entre ces 2 variables ?

b) Déterminer pour ces observations la droite des moindres carrés. Indiquer sur la feuille les ordonnées des y_i calculés par la droite des moindres carrés correspondant aux différentes valeurs des x_i. Tracer ensuite la droite sur le même graphique.

c) Quelle est une estimation plausible de Y à $x_i = 21$?

d) Quel est l'écart entre la valeur observée de Y à $x_i = 21$ et la valeur estimée avec la droite des moindres carrés ? Comment appelle-t-on cet écart ?

e) Est-ce que la droite des moindres carrés obtenue en b) passe par le point $(\overline{x}; \overline{y})$? Peut-on généraliser cette conclusion à n'importe laquelle droite de régression ?

Exercice I.2. La société de Transport Laviolette veut établir une politique d'entretien des camions de sa flotte. Tous sont de même modèle et utilisés à des transports semblables. La direction de la société est d'avis qu'une liaison statistique entre le coût direct de déplacements (cents par km) et l'espace de temps écoulé depuis la dernière inspection de ce camion serait utile. On a donc recueilli un certain nombre de données sur ces 2 variables. On veut utiliser la régression comme outil statistique.

Coût direct	10	18	24	22	27	13	10	24	25	8	16
Nombre de mois	3	7	10	9	11	6	5	8	7	4	6
Coût direct	20	28	22	19	18	26	14	20	26	30	12
Nombre de mois	9	12	8	10	9	11	6	8	10	12	5

a) Quelle variable devrait-on identifier variable dépendante (Y) et laquelle devrait-on identifier variable explicative (X) ?

b) Tracer le diagramme de dispersion de ces observations. Est ce que le nuage de points suggère une forme de liaison particulière ?

c) Les calculs "à la main" conduisent à :
$\sum x_i = 176$, $\sum y_i = 432$, $\sum x_i y_i = 3\,768$, $\sum x_i^2 = 1\,546$, $\sum y_i^2 = 9\,348$.
Comment retrouver ces valeurs avec Minitab ? En clair, quelles fonctions de Minitab utilisez-vous ? Ensuite calculez en utilisant les formules du cours les coefficients $\hat{\beta}_0$ et $\hat{\beta}_1$ de la droite des moindres carrés. Avec Minitab, calculez l'équation de la droite des moindres carrés. Retrouvez-vous les mêmes résultats ?

d) Avec l'équation de la droite des moindres carrés, quelle est l'estimation la plus plausible du coût direct de déplacement pour des camions dont la dernière inspection remonte à 6 mois ?

e) D'après les résultats de cette étude, un délai supplémentaire d'un mois pour l'inspection d'un camion occasionnera-t-il une augmentation ou une diminution du coût direct ? Quelle sera vraisemblablement la valeur de cette variation de coût ?

f) Déterminer la variation totale dans le coût direct de déplacement.

g) L'équation de la droite des moindres carrés pour les données de la société est : $\hat{y}_i = 1,54941 + 2,26087\, x_i$. Calculer la variation qui est expliquée par la droite des moindres carrés.

h) Quelle est la variation résiduelle ?

i) Calculer le coefficient R^2 et interpréter le résultat.

Exercice I.3. Une étudiante en sociologie veut analyser, dans le cadre d'un projet de fin de session, s'il existe une relation linéaire entre la densité de population dans les régions métropolitaines et le taux de criminalité correspondant dans ces régions.

Le taux de criminalité (Y) est indiqué en nombre de crimes par 10 000 habitants et la densité de population (X) est mesurée en milliers d'habitants par km carré.

Les calculs préliminaires conduisent à :
$n = 12$; $\sum x_i = 76$; $\sum y_i = 96$; $\sum x_i y_i = 732,6$; $\sum x_i^2 = 576,46$; $\sum y_i^2 = 962$.

Régions no	1	2	3	4	5	6	7	8	9	10	11	12
x_i	7,7	5,8	11,5	2,1	3,7	3,6	7,5	4,2	3,8	10,3	8,6	7,2
y_i	12	9	15	4	4	2	10	3	5	11	10	11

a) Tracer le diagramme de dispersion de ces observations.

b) Calculer le coefficient de corrélation linéaire entre ces 2 variables.

c) Peut-on conclure que la corrélation linéaire entre le taux de criminalité et la densité de population est significative au seuil de $\alpha = 0,05$?

d) Calculer avec les formules du cours les coefficients de la droite des moindres carrés et calculer avec Minitab l'équation de cette droite.

e) A quelle augmentation du taux de criminalité peut-on s'attendre pour une variation unitaire (ici 1 000 habitants par km^2) de la densité de population ?

f) Estimer le taux de criminalité le plus plausible pour une densité de population de 75 000 habitants par km^2.

g) A l'aide des calculs préliminaires, calculer la variation totale du taux de criminalité.

h) Calculer la variation qui est expliquée par la droite des moindres carrés.

i) Quelle proportion de la variation totale est expliquée par la droite des moindres carrés ?

Exercice I.4. Un étudiant en techniques forestières veut utiliser la régression linéaire pour estimer le volume en bois utilisable d'un arbre debout en fonction de l'aire du tronc mesuré à 25 cm du sol. Il a choisi au hasard 10 arbres et a mesuré, à la base, l'aire correspondante (en cm^2). Il a par la suite enregistré, une fois l'arbre coupé, le volume correspondant en m^3. Les calculs préliminaires donnent :

$n = 10; \sum x_i = 5\,313; \sum y_i = 2,66; \sum x_i y_i = 1\,527,63; \sum x_i^2 = 3\,147\,509$; $\sum y_i^2 = 0,7882$.

Vol.	$0,152$	$0,284$	$0,187$	$0,350$	$0,416$	$0,230$	$0,242$	$0,276$	$0,383$	$0,140$
Aire	297	595	372	687	790	520	473	585	762	232

a) Dans l'équation de la droite des moindres carrés :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

le coefficient $\hat{\beta}_1 = 0,00049$. Déterminer la valeur de $\hat{\beta}_0$.

b) Calculer le coefficient de corrélation linéaire entre ces 2 variables.

c) Tester les hypothèses suivantes :

$$H_0 : \rho = 0, H_1 : \rho > 0.$$

Utiliser $\alpha = 0,05$.

d) Son professseur lui mentionne qu'il peut, à l'oeil, évaluer avec une assez bonne précision le volume d'un arbre. L'étudiant un peu perplexe lui lance un défi : “je gage 1 euro que je fais mieux que vous avec le modèle des moindres carrés”. “D'accord.” Ayant justement un arbre tout près, le professeur lui dit, après une expertise de quelques minutes que cet arbre a un volume de $0,22$ m^3. Sans plus tarder ; l'étudaint mesure l'aire de la base de l'arbre et obtient 465 cm^2. Calculer avec la droite des moindres carrés, l'estimation la plus plausible du volume de l'arbre.

e) On s'acharne par la suite à couper l'arbre et le volume correspondant est $0,24$ m^3. Celui qui a le plus faible écart de prévision empoche le pari. Lequel s'est enrichi de 1 euro ?

f) Est ce que le volume moyen des arbres échantillonnés aurait donné une estimation aussi bonne que la droite des moindres carrés pour cet arbre ?

Exercice I.5. L'entreprise INFORMATEX se spécialise dans l'analyse de systèmes et la programmation sur ordinateur de problèmes techniques et de gestion. Elle veut utiliser la régression dans une étude sur le temps requis, par ses analystes-programmeurs, pour programmer des projets complexes.

Cette étude pourrait permettre à la firme d'établir des normes quant au temps requis pour programmer certains projets et d'assurer éventuellement une meilleure planification des ressources humaines. Les données du tableau suivant représentent le temps total en heures requis pour programmer différents projets en fonction du nombre d'instructions dans chaque programme.

Temps total en heures	40	55	62	58	82	94	120
Nombre d'instructions	60	82	100	142	190	220	285
Temps total en heures	134	128	140	152	174	167	218
Nombre d'instructions	354	400	425	440	500	530	640

a) Si on veut expliquer les fluctuations dans le temps requis pour programmer les projets quelle variable doit-on identifier comme variable dépendante ?

Comme variable explicative ?

b) Qu'est qui peut renseigner l'entreprise sur la forme de liaison statistique qui peut exister entre ces 2 variables ?

c) Quelle méthode d'ajustement linéaire doit-on utiliser pour obtenir les estimateurs des coefficients de la droite de régression ?

d) Après quelques calculs préliminaires, on obtient :
$n = 14; \sum x_i = 4\,368; \sum y_i = 1\,624; \sum x_i y_i = 631\,852; \sum x_i^2 = 806\,254$; $\sum y_i^2 = 224\,526$. Calculer à l'aide des formules du cours $\hat{\beta}_0$ et $\hat{\beta}_1$. Retrouvez ces coefficients à l'aide de Minitab.

e) Quelle est l'équation de la droite des moindres carrés ?

f) Calculer le coefficient de la régression linéaire.

g) Peut-on conclure, au seuil $\alpha = 0,05$ en la présence d'une corrélation linéaire positive entre ces 2 variables ?

h) Si on ne tient pas compte du nombre d'instructions, quelle valeur pourrait-on utiliser comme estimation du temps moyen de programmation des projets ?

i) Quelle correction peut-on apporter à l'estimation obtenue en h), en tenant compte du nombre d'instructions par l'entremise de la droite des moindres carrés ?

j) D'après la droite des moindres carrés, à quelle augmentation du temps de programmation peut-on s'attendre lorsque le nombre d'instructions augmente de 50 ?

k) Pour chaque nombre d'instructions suivant, estimer le temps de programmation à l'aide de la droite des moindres carrés

Nombre d'instructions	100	220	440
Estimation du temps de programmation			

l) Selon les résultats observés, quels sont les écarts de prévision de l'équation des moindres carrés pour le nombre d'instruction en k) ?

m) Si on avait utilisé l'estimation obtenue en h) au lieu de celles déduites de l'équation des moindres carrés pour effectuer les prévisions selon le nombre d'instructions spécifié en k), quels auraient été alors, dans chaque cas, les écarts de prévision ?

n) Pour chaque valeur x_i spécifié en k), vérifier la relation $y_i - \overline{y} = (\hat{y}_i - \overline{y}) + (y_i - \hat{y}_i)$.

o) Calculer la variation totale, la variation expliquée par la droite des moindres carrés et la variation résiduelle.

p) Quelle proportion de la variation totale dans le temps de programmation est expliquée par la droite des moindres carrés ? Quelle proportion demeure inexpliquée par la droite ?

q) On avait fixé le R^2 à $0,90$ comme valeur minimale pour considérer la droite des moindres carrés d'utilité pratique. D'après les résultats obtenus, devrait-on utiliser la droite des moindres carrés comme outil de prévision ?