Université Mohamed Khider de Biskra. Faculté des Sciences Exactes et de la Vie
Département: Biologie Module: Biostatistique.
Année Univ: 2021-2022. Master 1 Par Prof Chala Adel.

Exercices de Cours sur la Régression Linéaire Simple

Exercice N°: 01

On a étudié les longueurs respectives des 2 (deux) paires d'ailes d'une espèce de guêpe (Vespa sp) sur un échantillon de 11 individus. Soit X la longueur d'une aile de la première paire et Y celle de l'aile de la deuxième paire mesurée sur le même individu. On a obtenu les résultats suivants:

L'individu	La longueur d'une aile de la première paire	La longueur d'une aile de la deuxième paire
I	294	624
II	271	661
III	314	728
IV	356	782
V	383	819
V I	369	869
V II	402	938
V III	422	1023
IX	475	1136
X	475	1227
XI	486	1317

On veut tester la réalité d'une relation linéaire entre Y et X, soit:

$$Y = \beta + \alpha X + \xi.$$

Les hypothèses classiques de modèle linéaire simple sont supposées réalisées, c'est-à-dire que ξ est une variable aléatoire suivant la loi normale centrée.

1/ Donner les estimations $\widehat{\alpha}$ et $\widehat{\beta}$ de α et β obtenues par la méthode des moindres carrés ordinaires MCO.

2/ Calculer l'estimateur de la variance, l'estimateur de la variance de l'estimateur $\widehat{\beta}$, et le coefficient de détermination de la régression R^2.

3/ On se pose le problème de test:

$$\begin{cases} H_0 : \widehat{\alpha} = 0, \\ H_1 : \widehat{\alpha} \neq 0. \end{cases}$$

A quelle question ce test répond t-il ? Peut-on dire que α est significativement différent de zéro, au risque $0,05$?

Exercice N°: 02

Pour vérifier les relations d'halométrie entre insectes, on a retenu les deux relations x La longueur de l'élytre y La largeur de la tête. Les mesures sur 50 insectes, notées $(x_i; y_i)$ ont fourni les résultats suivants:

$$\sum_1^{50} x_i = 155, \quad \sum_1^{50} y_i = 125, \quad \sum_1^{50} x_i y_i = 391,1.$$

$$\sum_1^{50} x_i^2 = 482,5, \quad \sum_1^{50} y_i^2 = 320,5, \quad \sum_1^{50} x_i^2 y_i^2 = 3468,7.$$

1/ Calculer:

a/ La moyenne et l'écart-type du caractère x sur l'échantillon observée.

b/ La moyenne et l'écart-type du caractère y sur l'échantillon observée.

c/ La covariance empirique et la corrélation empirique des variables x et y.

d/ L'équation de la droite de régression linéaire de y sur x obtenue par estimation sur ces données.

2/ Donner un estimateur de la variance pour l'estimateur du coefficient.

3/ Déterminer l'intervalle de confiance pour le coefficient $\widehat{\alpha}$ au taux de confiance 95%.

Exercice N°: 03

On considère le modèle linéaire:

$$y_t = \alpha x_t + \beta + \varepsilon_t$$

Où xi et yi représentent respectivement le nombre d'heures d'ensoleillement et le montant des ventes de crèmes glacées pour une région donnée et durant le mois de juin de l'année t. On suppose que les hypothèses assurant l'interprétation probabiliste de la méthode des MCO:

T	1	2	3	4	5
X	330	309	279	312	318
Y	32,4	32,1	29,7	31,2	31,1

1/ préciser le modèle.

2/ Donner une estimateur de $V(\alpha)$.

3/ Tester si l'ensoleillement liée significativement les ventes.

Exercice N°: 04

Les œufs d'une espèce particulière d'insecte nécessitants avant leur développement un séjour à basse température(Hivernation). On étudie la corrélation éventuelle qui liée la durée d'hivernation x et le temps de développement de ces œufs(x et y en jours).

1/ Calculer les estimation pour α et β à partir des observations suivantes:

$\overline{x} = 1,8$, $\overline{y} = 1,7$, $s_X^2 = \frac{1}{10}\sum_1^{10}(x_i - \overline{x})^2 = 1,9$, $s_Y^2 = \frac{1}{10}\sum_1^{10}(y_i - \overline{y})^2 = 26,7$, $s_{XY} = \frac{1}{10}\sum_1^{10}(x_i - \overline{x})(y_i - \overline{y}) = -6,2$.

2/ La variable X a-t-elle une influence significative sur Y au niveau $0,01$si $\widehat{\sigma(\widehat{a})} = 0,446$?

Exercice N°: 05

On désire avoir s'il existe une relation entre le poids de naissance d'un enfant et âge de sa mère à l'accouchement. Dans ce but, on prélève 100 dossiers médicaux dans le fichier des naissances d'une maternité. On calcule les quantités suivantes:

	Poids de naissance de l'enfant	L'age de la mère à l'accouchement
Moyenne observée	3100	25
Variance observée S^2	10 000	25

La covariance observée entre le poids de naissance de l'enfant et l'âge de la mère à l'accouchement:

$$s_{XY} = \frac{1}{n}\sum_1^n (x_i - \overline{x})(y_i - \overline{y}) = 450 \text{ gr/ans}$$

1/ Donner la valeur numérique du coefficient de corrélation entre le poids de naissance de l'enfant et l'âge de sa mère à l'accouchement.

2/ Calculer la droite de régression linéaire entre le poids de naissance de l'enfant et l'âge de sa mère à l'accouchement.

3/ La qualité du modèle est jugée par le coefficient de détermination de la régression, exprimer leur valeur.

4/ Au risque de 5% le poids de naissance de l'enfant et l'âge de sa mère à l'accouchement sont-ils significativement liés ?

Exercice N°: 06

D'avril 2008 à juillet 2010 on a relevé les données relatives aux variables suivantes pour le mois t:

y_t : Ecart entre les versements et les retraits dans les Caisses CASNOS.

x_{1t}: Accroissement relatif du revenu des ménages.

x_{2t}: Accroissement relatif de l'indice des prix de détail.

On a obtenu:

$$\sum y_t = 39,82; \qquad \sum x_{1t} = 25,14; \qquad \sum y_t^2 = 63,21;$$

$$\sum_t^2 x_{1t}^2 = 24,36; \qquad \sum x_{1t} y_t = 38,38.$$

1/ Ecrire l'équation de la droite de régression de y_t en x_{1t} et donner un indicateur de la qualité de cette régression.

2/ On effectue maintenant la régression y_t en x_{2t}, le calcul des résidus d'ajustement donne $\sum \varepsilon_i^2 = 1,96$. Comparer avec le modèle précédent.

Exercice N°: 07

Le revenue R_t et l'épargne nette E_t ont été mesurés par trimestre pendant 3 ans pour une catégorie socioprofessionnelle bien déterminée ; Après correction des variations saisonnières, exprimées en million de Dinars, les indicateurs suivants sont disponibles:

$$\overline{R}_t = \frac{1}{12}\sum_1^{12} R_i = 19,7, \qquad \sum_1^{12} R_i^2 = 4827,$$

$$\overline{E}_t = \frac{1}{12}\sum_1^{12} E_i = 6,1, \qquad \sum_1^{12} E_i^2 = 456.$$

$$\sum_1^{12} R_i E_i = 1480$$

On suppose que les variables E_t et R_t sont liée par le modèle:

$$E_t = \beta + \alpha R_t + u_t$$

Les v,a étant de la loi $\mathcal{N}(0;2)$ pour tout t et indépendantes.

1/ Calculer $\widehat{\alpha}$ et $\widehat{\beta}$, estimateurs des MCO de α et β.

2/ Etudier la validité du modèle.

3/ On désire tester l'hypothèse qu'une augmentation absolue de 1% du revenu implique une augmentation absolue de 1% de l'épargne. Ecrire cette hypothèse en fonction des coefficients de la régression et résoudre le problème.