Ecole supérieure d'agro-développement international

istom

Examen de statistiques et analyse de données

Jeudi 5 juin 2025

Promotion 115

Antoine Géré

Document(s) autorisé(s) : ☐ Oui ☒ Non Calculatrice autorisée : ☒ Oui ☐ Non

Remarques :

- Les exercices sont indépendants.
- Il sera tenu compte de la propreté de votre copie, ainsi que de la clarté et de la qualité de la rédaction et du raisonnement.
- **Ne pas écrire avec un crayon papier**, sauf pour dessiner et/ou annoter des croquis, le cas échéant.
- Utiliser les **notations** indiquées dans le texte et **justifier toutes vos réponses**.
- Le sujet est à conserver par l'étudiant-e.

Exercice 1

Un groupe d'istomien étudie, dans le cadre de leur MJE, les gorilles dans une réserve nationale au Rwanda. En particulier ils mesurent les poids et tailles des 10 premiers nouveaux nés d'un mois donné. Ils ont consigné ces données dans le tableau suivant :

$n^\circ i$ du nouveau né	1	2	3	4	5	6	7	8	9	10
Taille X_i en cm	50	52	47	49	50	55	54	51	45	49
Poids Y_i en g	3100	3230	2950	3050	3100	3350	3480	3250	3000	3150

1. Donner la nature des variables X et Y.
2. Calculer les variances des variables X et Y prises séparément, notées respectivement σ_X^2 et σ_Y^2.
3. Construire le nuage de points représentant le poids des nouveaux nés en fonction de leur taille. Commenter ce graphique.
4. Déterminer le point moyen G.
5. Construire le tableau de contingence pour les variables X et Y. On choisira trois intervalles pour X et Y.
6. Calculer la covariance σ_{XY}.
7. Déterminer le coefficient de corrélation. Que pouvons nous en déduire sur le lien entre X et Y
8. Déterminer la droite de régression de Y en fonction X, expliquant le poids d'un nouveau né en fonction de sa taille.

Indication ▼ Correction ▼ [stat-0010]

Exercice 2

Un étudiant istomien a étudié dans le cadre de son stage de fin d'étude l'influence d'un engrais bio sur les récoltes de riz d'une exploitation au Vietnam. Avec l'aide de l'équipe sur place il a effectué 200 mesures sur des zones témoins dans deux rizières, une ayant bénéficié de l'engrais bio, et l'autre non. Le tableau ci-dessous donne la répartition de ces 200 mesures en fonction de la rizière où la mesure a été effectuée. On note

- E la rizière ayant bénéficié de l'apport en engrais,
- $\overline{E}$ la rizière n'ayant pas bénéficié de l'apport en engrais,
- P le poids en kg de riz mesuré sur chacune des zones témoins.

	E	$\overline{E}$
$P \in [0;3[$	26	20
$P \in [3;4[$	61	63
$P \in [4;6[$	8	22

1. Expliquer clairement le tableau ci-dessus, à quoi correspond le nombre 63 ?
2. Déterminer l'effectif marginal de P dans chacune des modalités E et $\overline{E}$.
3. Donner l'effectif total.
4. Calculer la moyenne conditionnelle de P dans chacune des modalités E et $\overline{E}$.
5. Calculer la variance conditionnelle de P dans chacune des modalités E et $\overline{E}$.
6. Rappeler la relation de décomposition de la variance.
7. Calculer la variance.
8. Calculer le rapport de corrélation.
9. Donner une interprétation de votre résultat.

Indication ▼ Correction ▼ [stat-0011]

Indication pour l'exercice 1 ▲

Utiliser les formules de König-Huygens : $V(X) = \overline{x^2} - \bar{x}^2$ et $\mathrm{Cov}(X,Y) = \overline{xy} - \bar{x}\,\bar{y}$. Le coefficient de corrélation est $r = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$ et la droite de régression de Y en X a pour pente $a = \dfrac{\mathrm{Cov}(X,Y)}{V(X)}$ et passe par le point moyen G.

Indication pour l'exercice 2 ▲

Remplacer chaque classe de poids par son centre $(1{,}5\,;\ 3{,}5\,;\ 5)$. La variance totale se décompose en $V(P) = V_{\text{inter}} + V_{\text{intra}}$, où V_{intra} est la moyenne pondérée des variances conditionnelles et V_{inter} la variance des moyennes conditionnelles. Le rapport de corrélation est $\eta^2 = V_{\text{inter}}/V(P)$.

Tournez la page S.V.P.

Correction de l'exercice 1 ▲

1. X (taille) et Y (poids) sont deux variables **quantitatives continues**.
2. $\overline{X} = \dfrac{502}{10} = 50{,}2$ cm, $\overline{Y} = \dfrac{31\,660}{10} = 3\,166$ g.

$$\sigma_X^2 = \frac{25\,282}{10} - 50{,}2^2 = 2\,528{,}2 - 2\,520{,}04 = \mathbf{8{,}16}\ \text{cm}^2, \qquad \sigma_X \approx 2{,}857\ \text{cm},$$

$$\sigma_Y^2 = \frac{100\,475\,800}{10} - 3\,166^2 = 10\,047\,580 - 10\,023\,556 = \mathbf{24\,024}\ \text{g}^2, \qquad \sigma_Y \approx 155{,}0\ \text{g}.$$

3. Le nuage de points (poids en ordonnée, taille en abscisse) est allongé selon une direction croissante : les nouveau-nés les plus grands sont aussi les plus lourds ; une relation linéaire semble adaptée.
4. Point moyen : $G(50{,}2\,;3\,166)$.
5. Exemple avec les classes $X \in [45;49[$, $[49;52[$, $[52;56[$ et $Y \in [2\,900;3\,100[$, $[3\,100;3\,300[$, $[3\,300;3\,500[$:

$X \backslash Y$	[2900;3100[	[3100;3300[	[3300;3500[	Total
[45;49[	2	0	0	2
[49;52[	1	4	0	5
[52;56[	0	1	2	3
Total	3	5	2	10

(D'autres découpages sont possibles ; les effectifs se concentrent sur la diagonale, signe d'une liaison positive.)

6. $\sum x_i y_i = 1\,593\,330$, d'où $\sigma_{XY} = \dfrac{1\,593\,330}{10} - 50{,}2 \times 3\,166 = 159\,333 - 158\,933{,}2 = \mathbf{399{,}8}$.
7. $r = \dfrac{399{,}8}{2{,}857 \times 155{,}0} \approx \mathbf{0{,}90}$: forte corrélation linéaire positive entre la taille et le poids des nouveau-nés.
8. $a = \dfrac{\sigma_{XY}}{\sigma_X^2} = \dfrac{399{,}8}{8{,}16} \approx 49{,}0$ et $b = \overline{Y} - a\overline{X} \approx 3\,166 - 49{,}0 \times 50{,}2 \approx 706{,}4$.
   Droite de régression : $\mathbf{Y \approx 49{,}0\,X + 706{,}4}$: chaque centimètre supplémentaire correspond en moyenne à environ 49 g de plus.

Correction de l'exercice 2 ▲

1. Le tableau croise la rizière (E ou $\overline{E}$) et la classe de poids P récolté sur une zone témoin. Le nombre 63 signifie que, dans la rizière *sans* engrais bio, 63 zones témoins ont donné entre 3 et 4 kg de riz.
2. Effectifs marginaux : $n_E = 26 + 61 + 8 = \mathbf{95}$ zones et $n_{\overline{E}} = 20 + 63 + 22 = \mathbf{105}$ zones.
3. Effectif total : $95 + 105 = \mathbf{200}$.
4. Avec les centres de classes 1,5 ; 3,5 ; 5 :

$$\overline{P}_E = \frac{26 \times 1{,}5 + 61 \times 3{,}5 + 8 \times 5}{95} = \frac{292{,}5}{95} \approx \mathbf{3{,}08}\ \text{kg},$$

$$\overline{P}_{\overline{E}} = \frac{20 \times 1{,}5 + 63 \times 3{,}5 + 22 \times 5}{105} = \frac{360{,}5}{105} \approx \mathbf{3{,}43}\ \text{kg}.$$

5.

$$V_E = \frac{26 \times 1{,}5^2 + 61 \times 3{,}5^2 + 8 \times 5^2}{95} - 3{,}079^2 \approx \mathbf{1{,}107},$$

$$V_{\overline{E}} = \frac{20 \times 1{,}5^2 + 63 \times 3{,}5^2 + 22 \times 5^2}{105} - 3{,}433^2 \approx \mathbf{1{,}229}.$$

6. $V(P) = V_{\text{inter}} + V_{\text{intra}}$, avec

$$V_{\text{intra}} = \frac{n_E V_E + n_{\overline{E}} V_{\overline{E}}}{n} \text{ (moyenne des variances)},$$

$$V_{\text{inter}} = \frac{n_E(\overline{P}_E - \overline{P})^2 + n_{\overline{E}}(\overline{P}_{\overline{E}} - \overline{P})^2}{n} \text{ (variance des moyennes)}.$$

7. Moyenne générale $\overline{P} = \dfrac{292{,}5 + 360{,}5}{200} = 3{,}265$ kg.

$$V_{\text{intra}} \approx \frac{95 \times 1{,}107 + 105 \times 1{,}229}{200} \approx 1{,}171,$$

$$V_{\text{inter}} \approx \frac{95 \times (3{,}079 - 3{,}265)^2 + 105 \times (3{,}433 - 3{,}265)^2}{200} \approx 0{,}031,$$

$$V(P) \approx 1{,}171 + 0{,}031 = \mathbf{1{,}202} \text{ kg}^2 \quad \text{(vérification directe : } \overline{P^2} - \overline{P}^2 \approx 1{,}202).$$

8. $\eta^2 = \dfrac{V_{\text{inter}}}{V(P)} \approx \dfrac{0{,}031}{1{,}202} \approx \mathbf{0{,}026}$.

9. Seulement 2,6 % de la variabilité des récoltes s'explique par la rizière : le lien est très faible. De plus, la rizière avec engrais bio a une récolte moyenne plus *faible*. Rien ne montre ici un effet bénéfique de l'engrais (d'autres facteurs, comme la fertilité initiale des deux rizièress, peuvent intervenir).
