

Examen de statistiques

Lundi 8 juin 2026

Promotion 116

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 Intensification écologique de la culture du riz à Madagascar

Dans les zones de haute altitude de Madagascar, la riziculture irriguée fait face à des contraintes thermiques importantes. Pour optimiser l'efficacité de la fertilisation organique, un centre de recherche agronomique local teste l'effet de doses croissantes de *compost enrichi en biochar* sur le rendement final d'une variété de riz de contre-saison.

L'ingénieur agronome responsable de l'essai a collecté les données suivantes sur $n = 8$ parcelles paysannes élémentaires :

- Variable X : Dose de compost enrichi apportée (en tonnes par hectare, t/ha)
- Variable Y : Rendement en grain obtenu à la récolte (en quintaux par hectare, q/ha)

Numéro de parcelle	Dose de compost X_i (t/ha)	Rendement Y_i (q/ha)
1	1,0	32
2	1,5	35
3	2,0	41
4	2,5	44
5	3,0	48
6	3,5	51
7	4,0	56
8	4,5	59

1. Calculer les moyennes \bar{X} et \bar{Y} .
2. Calculer les variances $\sigma^2(X)$ et $\sigma^2(Y)$. En déduire les écarts-types σ_X et σ_Y .
3. Calculer la covariance $\text{Cov}(X, Y)$. Interpréter le signe de cette covariance d'un point de vue agronomique.
4. Calculer le coefficient de corrélation linéaire. Que peut-on en conclure concernant l'impact du compost sur le rendement du riz ?
5. *Modélisation (Ajustement affine)* :

- Déterminer l'équation de la droite de régression de Y en X sous la forme $Y = aX + b$ (arrondir les coefficients à 2 décimales).
- Interpréter concrètement la valeur du coefficient directeur a (la pente).
- À l'aide de ce modèle, estimer le rendement théorique que l'on pourrait espérer si un paysan appliquait une dose de 5,0 t/ha de compost. Quelles sont les limites d'une telle prédiction ?

Correction ▼

[stat-0055]

Exercice 2 Performance comparée de variétés de sorgho au Burkina Faso

Dans la région du Centre-Nord au Burkina Faso, le sorgho est une culture vivrière de base hautement stratégique pour la sécurité alimentaire. Face au dérèglement climatique, un groupement de producteurs teste le rendement de trois variétés de sorgho sur plusieurs parcelles homogènes :

- *Variété V1* : Variété locale traditionnelle.
- *Variété V2* : Variété améliorée à cycle court (adaptée aux sécheresses précoces).
- *Variété V3* : Variété hybride sélectionnée pour sa tolérance au stress thermique.

L'ingénieur agronome de la coopérative mesure le rendement en grain (exprimé en *quintaux par hectare*, q/ha) sur un échantillon total de $N = 12$ parcelles. Voici les données brutes de l'essai :

Numéro de parcelle	Variété testée X	Rendement obtenu Y (q/ha)
1	V1 (Locale)	12
2	V1 (Locale)	14
3	V1 (Locale)	11
4	V1 (Locale)	15
5	V2 (Cycle court)	18
6	V2 (Cycle court)	16
7	V2 (Cycle court)	15
8	V2 (Cycle court)	19
9	V3 (Hybride)	22
10	V3 (Hybride)	25
11	V3 (Hybride)	21
12	V3 (Hybride)	24

- Déterminer la variable qualitative (ainsi que ses modalités) et la variable quantitative.
- Proposer une représentation graphique adaptée à ce type de données bivariées pour comparer visuellement les performances des trois variétés. .
- (BONUS)** Réaliser la représentation graphique suggérée à la question 2.
- Analyse des profils conditionnels* :
 - Calculer l'effectif de chaque groupe (n_1, n_2, n_3).
 - Calculer le rendement moyen pour chaque variété ($\bar{Y}_1, \bar{Y}_2, \bar{Y}_3$).
 - Calculer le rendement moyen global \bar{Y} de l'ensemble de l'essai.
- Analyse de la dispersion* :
 - Calculer la variance interne à chaque groupe (V_1, V_2, V_3).
 - Quelle variété présente la production la plus homogène (la plus stable) d'une parcelle à l'autre ?
- Vers l'analyse de variance* :
 - Calculer la *Variance Intra-groupes*. Que mesure-t-elle concrètement ?

(b) Calculer la *Variance Inter-groupes*. Que mesure-t-elle concrètement ?

7. *Mesure de l'intensité de la liaison (Rapport de corrélation)* :

(a) À l'aide des résultats de la question 6, calculer le rapport de corrélation.

(b) Interpréter précisément cette valeur. Qu'indique-t-elle quant à l'influence de la variété sur le rendement du sorgho ?

Correction ▼

[stat-0056]

Correction de l'exercice 1 ▲

1. Calcul des moyennes marginales \bar{X} et \bar{Y} .

La moyenne d'une série statistique empirique est obtenue en divisant la somme des valeurs observées par l'effectif total ($n = 8$).

- *Moyenne de la dose de compost (\bar{X})* :

$$\sum_{i=1}^8 X_i = 1,0 + 1,5 + 2,0 + 2,5 + 3,0 + 3,5 + 4,0 + 4,5 = 22$$

$$\bar{X} = \frac{22}{8} = 2,75 \text{ t/ha}$$

- *Moyenne du rendement en grain (\bar{Y})* :

$$\sum_{i=1}^8 Y_i = 32 + 35 + 41 + 44 + 48 + 51 + 56 + 59 = 366$$

$$\bar{Y} = \frac{366}{8} = 45,75 \text{ q/ha}$$

2. Calcul des variances et des écarts-types

En appliquant la formule de Koenig-Huygens ($V = \frac{1}{n} \sum Z_i^2 - \bar{Z}^2$), on calcule les variances.

Note : Afin d'éviter des erreurs d'arrondis en cascade lors des calculs suivants, nous conservons la valeur exacte décimale des variances.

- *Variance et écart-type de X* :

$$\sum_{i=1}^8 X_i^2 = 1,0^2 + 1,5^2 + 2,0^2 + 2,5^2 + 3,0^2 + 3,5^2 + 4,0^2 + 4,5^2 = 71$$

$$V(X) = \frac{71}{8} - (2,75)^2 = 8,875 - 7,5625 = 1,3125$$

$$s_X = \sqrt{1,3125} \approx 1,15 \text{ t/ha}$$

- *Variance et écart-type de Y* :

$$\sum_{i=1}^8 Y_i^2 = 32^2 + 35^2 + 41^2 + 44^2 + 48^2 + 51^2 + 56^2 + 59^2 = 17368$$

$$V(Y) = \frac{17368}{8} - (45,75)^2 = 2171 - 2093,0625 = 77,9375$$

$$s_Y = \sqrt{77,9375} \approx 8,83 \text{ q/ha}$$

3. Calcul et interprétation de la covariance $\text{Cov}(X, Y)$

En appliquant la formule de Koenig-Huygens pour la covariance ($\text{Cov}(X, Y) = \frac{1}{n} \sum X_i Y_i - \bar{X}\bar{Y}$) :

- Somme des produits croisés :

$$\sum_{i=1}^8 X_i Y_i = (1,0 \times 32) + (1,5 \times 35) + \dots + (4,5 \times 59) = 1104,5$$

- Calcul numérique :

$$\text{Cov}(X, Y) = \frac{1104,5}{8} - (2,75 \times 45,75) = 138,0625 - 125,8125 = 12,25$$

- *Interprétation agronomique* : La covariance est strictement positive ($\text{Cov}(X, Y) > 0$). Cela démontre que les deux variables évoluent de concert et dans le même sens : l'augmentation des apports de compost enrichi s'accompagne globalement d'une hausse linéaire du rendement du riz.

4. Coefficient de corrélation linéaire

Le coefficient r évalue l'intensité de la relation affine existant entre X et Y :

$$r = \frac{\text{Cov}(X, Y)}{s_X \times s_Y} = \frac{12,25}{\sqrt{1,3125} \times \sqrt{77,9375}} = \frac{12,25}{10,11357} \approx 0,999$$

La valeur de r est extrêmement proche de 1. Il existe une *corrélation linéaire positive quasi-parfaite* entre la dose de compost et le rendement. Le niveau d'apport en fertilisant est le facteur hautement prédictif et explicatif de la variabilité des récoltes constatée sur ces parcelles.

5. Modélisation par un ajustement affine ($Y = aX + b$)

(a) *Détermination des coefficients (Moindres carrés)* :

- Calcul de la pente a :

$$a = \frac{\text{Cov}(X, Y)}{V(X)} = \frac{12,25}{1,3125} = \frac{14}{1,5} \approx 9,33$$

- Calcul de l'ordonnée à l'origine b :

$$b = \bar{Y} - a\bar{X} = 45,75 - (9,3333 \times 2,75) = 45,75 - 25,6666 = 20,08$$

L'équation de la droite de régression de Y en X s'écrit : $Y = 9,33X + 20,08$

(b) *Interprétation concrète de la pente ($a = 9,33$)* : Le coefficient directeur correspond au taux de réponse agronomique de la culture. Concrètement, chaque apport additionnel d'une tonne par hectare (1 t/ha) de compost enrichi génère une hausse moyenne de rendement de *9,33 quintaux de riz par hectare (q/ha)*.

(c) *Estimation pour $X = 5,0$ t/ha et limites du modèle* :

- *Calcul prédictif* : $Y = (9,3333 \times 5,0) + 20,0833 = 66,75$ q/ha
- *Limites agronomiques (Extrapolation)* : Bien que mathématiquement exacte, cette estimation souffre de limites majeures :
 - Sortie de la plage expérimentale* : L'essai n'a mesuré les effets que jusqu'à 4,5 t/ha. Extrapoler au-delà suppose, sans preuve, que le comportement de la plante reste inchangé.
 - Loi des rendements décroissants* : En biologie végétale, la croissance n'est jamais indéfiniment linéaire. À forte dose, la culture sature (plateau technique) ; de surcroît, un excès de fertilisation organique/biochar peut engendrer des phénomènes de phytotoxicité ou bloquer l'assimilation d'autres nutriments, provoquant une baisse finale du rendement.

Correction de l'exercice 2 ▲

1. Identification des variables

- *La variable qualitative (X)* : C'est la *Variété de sorgho*. Ses trois modalités sont : V_1 (Traditionnelle locale), V_2 (Améliorée à cycle court) et V_3 (Hybride).

- La variable quantitative (Y) : C'est le Rendement obtenu à la récolte, exprimé en q/ha.

2. Représentation graphique adaptée

Le graphique idéal pour croiser ces variables est le diagramme en boxplots parallèles (Boxplots) :

Constat : On note visuellement une hiérarchie ascendante des rendements allant de V_1 à V_3 , avec des dispersions internes qui semblent équivalentes.

3. AJOUTER REPRESENTATIONS GRAPHIQUES! (coming)

4. Analyse des profils conditionnels

(a) *Effectifs* : L'essai est équilibré : $n_1 = n_2 = n_3 = 4$ parcelles.

(b) *Moyennes conditionnelles* :

$$\bar{Y}_1 = \frac{12 + 14 + 11 + 15}{4} = 13 \text{ q/ha} ; \quad \bar{Y}_2 = \frac{18 + 16 + 15 + 19}{4} = 17 \text{ q/ha}$$

$$\bar{Y}_3 = \frac{22 + 25 + 21 + 24}{4} = 23 \text{ q/ha}$$

(c) *Moyenne globale* (\bar{Y}) : L'effectif total est $N = 12$.

$$\bar{Y} = \frac{52 + 68 + 92}{12} = \frac{212}{12} \approx 17,667 \text{ q/ha}$$

5. Analyse de la dispersion

(a) *Variances conditionnelles* :

- Groupe V_1 : $V_1 = \frac{12^2 + 14^2 + 11^2 + 15^2}{4} - 13^2 = \frac{686}{4} - 169 = 171,5 - 169 = 2,5$
- Groupe V_2 : $V_2 = \frac{18^2 + 16^2 + 15^2 + 19^2}{4} - 17^2 = \frac{1166}{4} - 289 = 291,5 - 289 = 2,5$
- Groupe V_3 : $V_3 = \frac{22^2 + 25^2 + 21^2 + 24^2}{4} - 23^2 = \frac{2126}{4} - 529 = 531,5 - 529 = 2,5$

(b) *Homogénéité* : Puisque $V_1 = V_2 = V_3 = 2,5$, toutes les variétés présentent *exactement le même niveau d'homogénéité (et de stabilité)* d'une parcelle à l'autre.

6. Décomposition de la variance

(a) *Variance Intra-groupes* (V_{intra}) : C'est la moyenne pondérée des variances.

$$V_{intra} = \frac{4(2,5) + 4(2,5) + 4(2,5)}{12} = 2,5$$

Sens concret : Elle mesure la dispersion résiduelle à l'intérieur des groupes due aux fluctuations aléatoires et aux micro-variations du sol (le « bruit »).

(b) *Variance Inter-groupes* (V_{inter}) : C'est la variance des moyennes autour de $\bar{Y} = \frac{212}{12}$:

$$V_{inter} = \frac{1}{12} \left[4 \times \left(13 - \frac{212}{12} \right)^2 + 4 \times \left(17 - \frac{212}{12} \right)^2 + 4 \times \left(23 - \frac{212}{12} \right)^2 \right] = \frac{7296}{432} \approx 16,889$$

Sens concret : Elle évalue l'écart des moyennes entre les variétés et quantifie la part de variabilité nette expliquée par le facteur d'étude (l'effet « Variété »).

7. Rapport de corrélation η^2

(a) *Calcul du rapport de corrélation* (η^2) : La variance totale vaut $V(Y) = V_{intra} + V_{inter} = 2,5 + 16,889 = 19,389$.

$$\eta^2 = \frac{V_{inter}}{V(Y)} = \frac{16,889}{19,389} \approx 0,871$$

- (b) *Interprétation agronomique* : La valeur $\eta^2 = 0,871$ prouve que 87,1% de la variabilité globale des rendements du sorgho est déterminée directement par le choix de la variété. La liaison est extrêmement puissante, ce qui fait du matériel génétique un levier agronomique majeur pour sécuriser la production dans cette région.
-