

# Examen de statistiques

Lundi 27 avril 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 Analyse de la performance agricole au Burkina Faso

Une ONG accompagne 10 coopératives agricoles dans la région du Sahel pour améliorer la culture du sorgho. Un agronome a recueilli des données sur deux variables principales pour chaque coopérative :

- *Type de fertilisant utilisé* : Les modalités sont soit Naturel (compost/fumier), soit Synthétique (engrais NPK).
- *Rendement à l'hectare* : Quantité récoltée par hectare, en quintaux.

Les données collectées sont les suivantes :

Coopérative	Type de fertilisant	Rendement (q/ha)
C1	Naturel	12
C2	Synthétique	18
C3	Naturel	14
C4	Naturel	13
C5	Synthétique	20
C6	Naturel	15
C7	Synthétique	17
C8	Naturel	14
C9	Synthétique	16
C10	Naturel	11

1. Donner la nature des variables Type de fertilisant et Rendement.
2. Analyse du type de fertilisant
  - (a) Dressez le tableau des effectifs et des fréquences.
  - (b) Quel type de graphique est le plus adapté pour représenter cette variable ? Justifiez votre choix.

### 3. Analyse du Rendement

- Calculez la moyenne, la médiane, le premier quartile et le troisième quartile de la série des rendements.
- Représentez la série des rendements par un boxplot.
- Calculez la variance et l'écart-type de cette série.
- La médiane est-elle supérieure ou inférieure à la moyenne ? Que pouvez-vous en déduire sur la distribution des rendements ?
- Si l'on souhaite comparer visuellement les rendements, quel graphique utiliseriez-vous ? Pourquoi ?

Correction ▼

[stat-0040]

### Exercice 2 Analyse bivariée de la production agricole au Sahel

Une étude agronomique a été menée dans une zone rurale du Sahel pour comprendre les déterminants de la productivité du mil. Deux axes d'analyse ont été retenus par les chercheurs :

- L'impact de la pluviométrie annuelle sur le rendement.
- Le lien entre le mode d'organisation de l'exploitation et l'adoption de techniques de conservation des eaux et des sols (CES).

#### Pluie vs Rendement

Quatorze parcelles ont été suivies. On note  $X$  la pluviométrie annuelle (en mm) et  $Y$  le rendement en quintaux par hectare (q/ha).

Parcelle	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$X$ (mm)	350	400	420	450	500	520	550	580	600	620	650	680	700	750
$Y$ (q/ha)	6	7	9	8	10	11	10	13	12	15	13	16	15	18

- Calculez la moyenne  $\bar{X}$  et  $\bar{Y}$ .
- Calculez la covariance de  $X$  et  $Y$ .
- En déduire le coefficient de corrélation de  $X$  et  $Y$ . Conclure sur l'intensité de la corrélation linéaire de  $X$  et  $Y$ .
- Calculez l'ajustement linéaire de  $X$  et  $Y$ , c'est à dire l'équation de la droite qui passe au mieux dans le nuage de points décrit par les variables  $X$  et  $Y$ .

#### Mode d'organisation vs Adoption CES

Les chercheurs ont interrogé 120 chefs d'exploitation sur leur mode d'organisation (Coopérative vs Individuel) et sur leur adoption ou non de techniques CES (Oui vs Non).

- 50 exploitations s'organisent en coopérative.
  - 35 exploitations s'organisent en coopérative et ont adopté les techniques CES.
  - 20 exploitations s'organisent individuellement et ont adopté les techniques CES.
- Construire le tableau de contingence
  - Calculez les effectifs théoriques.
  - Donnez la relation du  $\chi^2$  et indiquez ce que tente d'évaluer ce coefficient.
  - Calculer le  $\chi^2$ .
  - Existe-t-il une dépendance apparente entre le mode d'organisation et l'adoption de techniques CES ? Justifiez votre réponse.

Correction ▼

[stat-0041]

## Correction de l'exercice 1 ▲

### 1. Nature des variables :

- Type de fertilisant : C'est une variable **qualitative nominale**. Ses modalités (« Naturel » et « Synthétique ») expriment une caractéristique non mesurable qui ne possède pas d'ordre naturel.
- Rendement : C'est une variable **quantitative continue**. Elle exprime une quantité mesurable (exprimée ici en quintaux par hectare).

### 2. Analyse du type de fertilisant :

#### (a) Tableau des effectifs et des fréquences :

L'effectif total est de  $N = 10$  coopératives.

Type de fertilisant	Effectif ( $n_i$ )	Fréquence ( $f_i$ )
Naturel	6	0,6 (soit 60 %)
Synthétique	4	0,4 (soit 40 %)
<b>Total</b>	<b>10</b>	<b>1,0 (soit 100 %)</b>

TABLE 1 : Distribution de l'échantillon selon le type de fertilisant

#### (b) Graphique le plus adapté :

Le graphique le plus adapté pour représenter cette variable qualitative est un **diagramme en secteurs circulaires** (communément appelé camembert) ou un **diagramme en barres**.

*Le diagramme en secteurs est idéal pour mettre en évidence la répartition relative (les proportions de 60 % et 40 %) des modalités au sein de l'effectif global. Le diagramme en barres permet, quant à lui, de comparer directement et visuellement les effectifs de chaque catégorie.*

### 3. Analyse du Rendement :

Pour faciliter les calculs, ordonnons au préalable la série des  $n = 10$  rendements dans l'ordre croissant :

11 ; 12 ; 13 ; 14 ; 14 ; 15 ; 16 ; 17 ; 18 ; 20

#### (a) Calcul des paramètres de position :

- **Moyenne ( $\bar{x}$ ) :**

$$\bar{x} = \frac{11 + 12 + 13 + 14 + 14 + 15 + 16 + 17 + 18 + 20}{10} = \frac{150}{10} = 15 \text{ q/ha}$$

- **Médiane ( $Me$ ) :** L'effectif de la série est pair ( $n = 10$ ). La médiane se situe donc à la transition entre la 5<sup>e</sup> et la 6<sup>e</sup> valeur de la série ordonnée :

$$Me = \frac{x_{(5)} + x_{(6)}}{2} = \frac{14 + 15}{2} = 14,5 \text{ q/ha}$$

- **Premier quartile ( $Q_1$ ) :** On cherche la valeur correspondant à l'indice  $\frac{n}{4} = \frac{10}{4} = 2,5$ . On arrondit à l'entier supérieur, ce qui donne la 3<sup>e</sup> valeur de la série ordonnée :

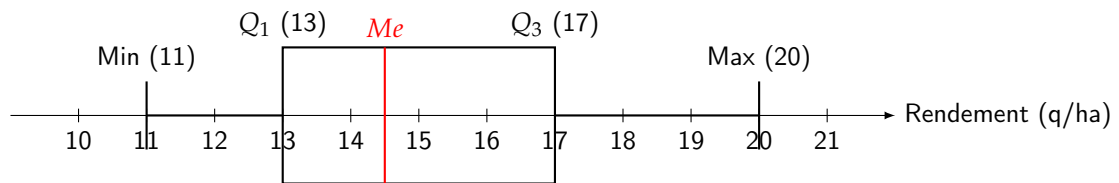
$$Q_1 = x_{(3)} = 13 \text{ q/ha}$$

- **Troisième quartile ( $Q_3$ ) :** On cherche la valeur correspondant à l'indice  $\frac{3n}{4} = \frac{30}{4} = 7,5$ . On arrondit à l'entier supérieur, ce qui donne la 8<sup>e</sup> valeur de la série ordonnée :

$$Q_3 = x_{(8)} = 17 \text{ q/ha}$$

#### (b) Représentation par un boxplot (boîte à moustaches) :

La boîte est délimitée par  $Q_1 = 13$  et  $Q_3 = 17$ , la ligne rouge centrale représente la médiane  $Me = 14,5$ , tandis que les moustaches s'étendent du minimum ( $Min = 11$ ) au maximum ( $Max = 20$ ).



(c) **Calcul de la variance et de l'écart-type :**

- **Variance ( $s^2$ ) :** En appliquant la formule des écarts à la moyenne  $\bar{x} = 15$  :

$$\begin{aligned}
 s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \frac{(11 - 15)^2 + (12 - 15)^2 + (13 - 15)^2 + 2(14 - 15)^2 + (15 - 15)^2 + (16 - 15)^2 + \dots}{10} \\
 &\quad \dots + (17 - 15)^2 + (18 - 15)^2 + (20 - 15)^2 \\
 &= \frac{(-4)^2 + (-3)^2 + (-2)^2 + 2(-1)^2 + 0^2 + 1^2 + 2^2 + 3^2 + 5^2}{10} \\
 &= \frac{16 + 9 + 4 + 2 + 0 + 1 + 4 + 9 + 25}{10} \\
 &= \frac{70}{10} \\
 &= 7 \text{ q/ha}^2
 \end{aligned}$$

- **Écart-type ( $s$ ) :**

$$s = \sqrt{s^2} = \sqrt{7} \approx 2,65 \text{ q/ha}$$

(d) **Comparaison et interprétation de la forme de la distribution :**

On remarque que la médiane (14,5 q/ha) est **inférieure** à la moyenne (15 q/ha).

On peut en déduire que la distribution des rendements présente une *asymétrie positive* (ou étalée vers la droite). Cette dissymétrie s'explique par la présence de quelques coopératives ayant de très bons rendements (notamment celle à 20 q/ha) qui tirent mathématiquement la moyenne vers le haut sans affecter la valeur de la médiane.

(e) **Graphique recommandé pour une comparaison visuelle :**

Pour comparer de façon optimale les rendements selon le type de fertilisant, il faudrait utiliser des **boîtes à moustaches comparatives** (représentées côte à côte ou l'une au-dessus de l'autre sur le même repère).

*Ce type de graphique permet de visualiser et d'évaluer instantanément les différences majeures entre les deux groupes de fertilisation : l'écart entre les tendances centrales (médianes), l'étendue globale des rendements, ainsi que la dispersion des données à travers l'écart interquartile de chaque pratique.*

## Correction de l'exercice 2 ▲

### Pluie vs Rendement

L'échantillon comporte  $n = 14$  parcelles (remarque : bien que l'énoncé mentionne « Dix parcelles », le tableau fournit bien les données pour 14 parcelles).

1. **Calcul des moyennes  $\bar{X}$  et  $\bar{Y}$  :**

- Pour la pluviométrie  $X$  :

$$\sum_{i=1}^{14} X_i = 350 + 400 + 420 + 450 + 500 + 520 + 550 + 580 + 600 + 620 + 650 + 680 + 700 + 750 = 7770$$

$$\bar{X} = \frac{7770}{14} = 555 \text{ mm}$$

- Pour le rendement  $Y$  :

$$\sum_{i=1}^{14} Y_i = 6 + 7 + 9 + 8 + 10 + 11 + 10 + 13 + 12 + 15 + 13 + 16 + 15 + 18 = 163$$

$$\bar{Y} = \frac{163}{14} \approx 11,6 \text{ q/ha}$$

2. **Calcul de la covariance de  $X$  et  $Y$**  : On utilise la formule de König-Huygens pour la covariance :

$$\sigma(X, Y) = \left( \frac{1}{n} \sum_{i=1}^n X_i Y_i \right) - \bar{X} \bar{Y}$$

Calculons d'abord la somme des produits  $\sum X_i Y_i$  :

$$\begin{aligned} \sum_{i=1}^{14} X_i Y_i &= (350 \times 6) + (400 \times 7) + (420 \times 9) + (450 \times 8) + (500 \times 10) + (520 \times 11) \\ &\quad + (550 \times 10) + (580 \times 13) + (600 \times 12) + (620 \times 15) + (650 \times 13) \\ &\quad + (680 \times 16) + (700 \times 15) + (750 \times 18) \\ &= 2100 + 2800 + 3780 + 3600 + 5000 + 5720 + 5500 + 7540 \\ &\quad + 7200 + 9300 + 8450 + 10880 + 10500 + 13500 \\ &= 95870 \end{aligned}$$

On en déduit la covariance :

$$\sigma(X, Y) = \frac{95870}{14} - (555 \times 11,643) \approx 386$$

3. **Calcul du coefficient de corrélation linéaire et conclusion** : Le coefficient de corrélation linéaire est défini par :

$$\rho = \frac{\sigma(X, Y)}{\sigma_X \sigma_Y}$$

Calculons les écarts-types  $\sigma_X$  et  $\sigma_Y$  :

- Variance de  $X$  :

$$\sum_{i=1}^{14} X_i^2 = 350^2 + 400^2 + \dots + 750^2 = 4502500$$

$$\sigma_X^2 = \frac{4502500}{14} - 555^2 = 13582 \implies \sigma_X \approx 117$$

- Variance de  $Y$  :

$$\sum_{i=1}^{14} Y_i^2 = 6^2 + 7^2 + \dots + 18^2 = 2063$$

$$\sigma_Y^2 = \frac{2063}{14} - 11,6^2 \approx 12,8 \implies \sigma_Y \approx 3,58$$

On calcule alors  $r$  :

$$\rho = \frac{386}{117 \times 3,58} \approx 0,92$$

Le coefficient de corrélation  $\rho \approx 0,92$  est très proche de 1. Il existe une **très forte corrélation linéaire positive** entre la pluviométrie annuelle et le rendement du mil. Plus la pluviométrie est élevée, plus le rendement est fort.

4. **Calcul de l'ajustement linéaire (droite de régression de  $Y$  en  $X$ )** : L'équation de la droite d'ajustement par les moindres carrés est sous la forme  $Y = aX + b$ .

- Calcul de la pente  $a$  :

$$a = \frac{\sigma(X, Y)}{\sigma_X^2} \approx \frac{386}{117} \approx 0,0284$$

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

$$b = \bar{Y} - a\bar{X} \approx 11,6 - 0,0284 \times 555 \approx 11,643 - 15,762 = -4,119$$

L'équation de la droite de régression est donc :

$$Y = 0,0284 X - 4,162$$

### Mode d'organisation vs Adoption CES

5. **Construction du tableau de contingence** : On dispose des données suivantes sur  $N = 120$  exploitations :

- Total Coopérative = 50. Donc, Total Individuel =  $120 - 50 = 70$ .
- Coopérative et CES (Oui) = 35.
- Individuel et CES (Oui) = 20.

On en déduit par soustraction les autres cases du tableau :

	CES (Oui)	CES (Non)	Total
Coopérative	35	15	<b>50</b>
Individuel	20	50	<b>70</b>
Total	<b>55</b>	<b>65</b>	<b>120</b>

TABLE 2 : Tableau de contingence observé

6. **Calcul des effectifs théoriques** : Sous l'hypothèse d'indépendance entre les deux critères, l'effectif théorique d'une case se calcule par :

$$n_{ij}^* = \frac{\text{Total ligne } i \times \text{Total colonne } j}{\text{Total général}} = \frac{n_{i\bullet} \times n_{\bullet j}}{N}$$

On trouve

$$n_{11}^* = 22,92 \quad n_{12}^* = 27,08 \quad n_{21}^* = 32,08 \quad n_{22}^* = 37,92$$

7. **Relation du  $\chi^2$  et objectif du coefficient** :

- La relation permettant de calculer le  $\chi^2$  est :

$$\chi^2 = \sum \frac{(n_{ij}^* - n_{ij})^2}{n_{ij}^*}$$

où  $n_{ij}$  représente l'effectif observé et  $n_{ij}^*$  l'effectif théorique.

- Ce coefficient tente d'évaluer s'il existe une liaison (dépendance statistique) significative entre deux variables qualitatives, en comparant la répartition observée à celle que l'on obtiendrait sous une hypothèse d'indépendance parfaite.

8. **Calcul du  $\chi^2$**  :

$$\begin{aligned} \chi^2 &= \frac{(35 - 22,92)^2}{22,92} + \frac{(15 - 27,08)^2}{27,08} + \frac{(20 - 32,08)^2}{32,08} + \frac{(50 - 37,92)^2}{37,92} \\ &= \frac{(12,08)^2}{22,92} + \frac{(-12,08)^2}{27,08} + \frac{(-12,08)^2}{32,08} + \frac{(12,08)^2}{37,92} \\ &= \frac{145,926}{22,92} + \frac{145,926}{27,08} + \frac{145,926}{32,08} + \frac{145,926}{37,92} \\ &\approx 6,367 + 5,389 + 4,549 + 3,848 \\ &\approx \mathbf{20,15} \end{aligned}$$

La valeur calculée du  $\chi^2$  est d'environ **20,15**.

#### 9. Analyse de la dépendance et intensité de la relation (Coefficient de Cramer) :

Pour déterminer s'il existe une dépendance et mesurer son intensité, nous procédons en deux étapes : le test de décision du  $\chi^2$ , suivi du calcul du coefficient de Cramer.

##### Calcul du coefficient de Cramer (V) :

Pour mesurer la force de cette dépendance, on utilise le coefficient de Cramer défini par la formule :

$$V = \sqrt{\frac{\chi^2}{N \times \min(r-1, c-1)}} = \sqrt{\frac{20,15}{120 \times 1}} = \sqrt{\frac{20,15}{120}} \approx \sqrt{0,1679} \approx 0,41$$

**Conclusion et interprétation agronomique :** Le coefficient de Cramer obtenu est de  $V \approx 0,41$ . Un coefficient compris entre 0,3 et 0,5 indique une liaison d'intensité moyenne à forte entre les deux variables.

**Justification concrète :** Le mode d'organisation de l'exploitation joue un rôle non négligeable dans la transition agroécologique au Sahel. Faire partie d'une coopérative favorise nettement l'adoption des techniques de conservation des eaux et des sols (CES) : 70 % des producteurs en coopérative les appliquent, contre seulement 28,6 % des producteurs individuels. Les structures collectives facilitent probablement l'accès aux formations, à l'entraide de main-d'œuvre et aux financements nécessaires à ces aménagements.

---