

1. Statistiques et données agronomiques

Statistique Descriptive

Anicet Ebou

Institut National Polytechnique Félix Houphouët-Boigny

2026-2027

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

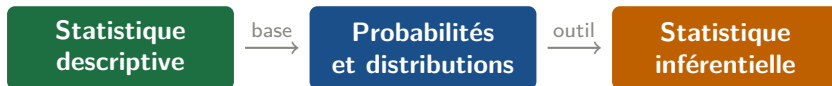
Quelques questions agronomiques typiques :

- Rendement moyen maïs : irrigation vs sans irrigation ?
- Variabilité de la teneur en azote dans un sol ?
- Lien entre pluviométrie et biomasse du soja ?
- Proportion de plants malades dans une parcelle ?

Ce cours

Statistique **descriptive** : résumer et visualiser des données avant toute décision ou inférence.

Les trois grandes familles de statistiques



Ce cours couvre **le premier bloc** qui est le fondement indispensable des deux suivants.

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

Trois grandes catégories d'études

Type	Objectif	Exemple
Enquête / recensement	Décrire une population existante	Rendements des exploitations d'une région
Essai expérimental	Comparer des traitements	Comparaison de variétés de riz en station
Suivi / monitoring	Suivre une évolution dans le temps	Teneur en MO du sol, mois par mois

Attention : Causalité vs association

Essai expérimental → causalité possible.

Enquête / monitoring → association seulement.

Définition : Unité d'observation

Élément sur lequel une mesure est effectuée.

Ex. : un plant, une parcelle, un échantillon de sol.

Définition : Population

Ensemble **total** des unités d'intérêt.

Ex. : tous les champs de cacao de Côte d'Ivoire.

Définition : Échantillon

Sous-ensemble **observé** de la population.

La qualité des conclusions dépend du mode de prélèvement.

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

Variable quantitative : valeur numérique, opérations arithmétiques valides.

Sous-type	Définition	Exemples agronomiques
Continue	Toute valeur dans un intervalle	Rendement (t/ha), pluviométrie (mm), pH
Discrète	Valeurs entières seulement	Nombre de grains/épi, nombre de plants malades

Continue

Histogramme, moyenne, écart-type.

Discrète

Diagramme en bâtons, médiane.

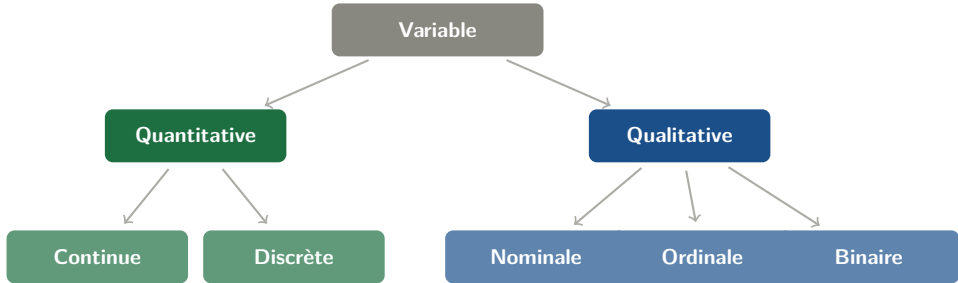
Variable qualitative : catégories, pas d'arithmétique.

Sous-type	Définition	Exemples agronomiques
Nominale	Catégories sans ordre	Type de sol, variété cultivée
Ordinale	Catégories avec ordre	Stade phénologique, classe de fertilité
Binaire	Deux modalités seulement	Irrigué / non irrigué, présence / absence

Attention : Piège courant

Un code numérique (1 = argileux, 2 = limoneux. . .) reste une variable **nominale**. Sa moyenne n'a aucun sens.

Arbre de classification des variables



1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

Les quatre échelles de Stevens

Stevens (1946) a classifié les données en quatre niveaux de complexité, définis par les opérations mathématiques permises :

Échelle	Propriété	Exemple	Stats autorisées
Nominale	Classification en catégories distinctes	Type de sol, variété	Effectif, mode
Ordinale	+ Ordre	Stade phénologique	+ Médiane, percentiles
Intervalle	+ Distances égales entre les points (zéro arbitraire)	Température (°C)	Moyenne, écart-type
Ratio	+ Zéro absolu, permettant de comparer des rapports	Rendement (t/ha), hauteur (cm)	Moyenne, CV, rapport

Température °C vs rendement

20 °C n'est *pas* deux fois plus chaud que 10 °C (zéro arbitraire).

En revanche, 4 t/ha est bien le double de 2 t/ha (zéro absolu → échelle **ratio**).

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

Sources d'erreur dans les données agronomiques

Source	Exemple	Prévention
Erreur de mesure	Balance mal étalonnée	Étalonnage, double mesure
Erreur de saisie	1 240 au lieu de 124	Double saisie, validation de plage
Valeur aberrante	Rendement négatif	Vérification graphique (box-plot)
Donnée manquante	Mesure oubliée	Protocole de substitution planifié
Biais de sélection	Enquêter seulement les accessibles	Échantillonnage probabiliste

Détecter des erreurs de saisie avec R

```
# Rendements (t/ha), une valeur suspecte
rendement <- c(
  3.1, 2.8, 3.5, 3.2, 3.4, 2.9,
  3.8, 3.0, 32.0, 3.3, 3.6, 2.7
)

summary(rendement)
boxplot(rendement, ylab = "Rendement (t/ha)")
```

Attention : Principe fondamental

Garbage in, garbage out.

Un contrôle graphique *systématique* précède toute analyse.
La valeur 32.0 est probablement 3.2 (erreur de frappe).

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

- **Libre et gratuit** : accès pérenne
- **Reproductible** : le script est la documentation
- **Standard** dans la littérature agronomique internationale
- **Packages spécialisés** : agricolae, lme4, vegan...

Interface recommandée : RStudio ou VS Code.

Ce cours

Chaque module contient des exemples R reproductibles.
Le code est disponible sur le site du cours.

```
# Vecteur de rendements en t/ha  
rdts <- c(2.8, 3.1, 3.5, 2.4, 3.8, 2.9, 4.1, 3.3)  
  
length(rdts) # nombre d'observations -> 8  
mean(rdts)    # moyenne -> 3.2375  
median(rdts)  # médiane -> 3.2  
sd(rdts)      # écart-type -> 0.554  
range(rdts)   # [min, max] -> 2.4 4.1
```

Créer un tableau de données (data.frame)

```
parcelles <- data.frame(  
  id = paste0("P", 1:5),  
  variete = c("TZPB", "TZPB", "CMS8704", "CMS8704", "OPV"),  
  rendement = c(3.1, 2.9, 4.2, 4.0, 3.5), # t/ha  
  azote = c(60, 60, 90, 90, 75) # kg N/ha  
)  
parcelles
```

Convention de nommage

Noms de colonnes en minuscules avec _ comme séparateur : rdt_tha, sol_ph.
Éviter les accents et les espaces dans les noms d'objets.

1. Pourquoi les statistiques en agronomie ?
2. Types d'études agronomiques
3. Types de variables
4. Échelles de mesure
5. Qualité des données
6. Prise en main de R
7. Résumé et exercices

Objectifs atteints :

1. Rôle des statistiques en agronomie
2. Types d'études : enquête / essai / monitoring
3. Types de variables : quantitative vs qualitative
4. Échelles de Stevens → statistiques autorisées
5. Sources d'erreur et stratégies de prévention
6. Premiers pas dans R : vecteur, `data.frame`, statistiques de base

Prochain module

Collecte et organisation des données : échantillonnage, tableaux de fréquences, données *tidy*.

Exercice 1.1 : Identifier les types de variables

Pour chaque variable, identifier : type, sous-type, échelle de Stevens.

Variable	Description
ph_sol	pH du sol mesuré au pH-mètre (0-14)
stade	Stade phénologique du maïs (végétatif, floraison...)
rdt_kg	Rendement en grain (kg/ha)
traitement	Dose d'engrais : T0, T1, T2, T3
nb_epis	Nombre d'épis récoltés par parcelle
couleur_grain	Couleur du grain : blanc, jaune, rouge

Exercice 1.2 : Contrôle qualité avec R

```
hauteurs <- c(
  112, 108, 115, 110, 113, 109, 1150, 114, 111, 107,
  116, 112, 109, 113, 108, 115, 114, 110, 112, 111
)
summary(hauteurs)
```

Questions :

1. Quelle valeur est aberrante ? Quelle erreur de saisie est probable ?
2. Corrigez et recalculez la moyenne et l'écart-type.
3. Commentez l'impact de ces nouveaux calculs sur les statistiques.

Merci de votre attention

Statistiques Descriptives · Module 1

Equipe Bioinformatique et Biostatistique

Laboratoire de Microbiologie, Biotechnologie et Bioinformatique

Institut National Polytechnique Félix Houphouët-Boigny

