

Travaux Dirigés n°1

Types de variables et protocoles d'échantillonnage

Modules M1 - M2 | Durée : 1h30 | Documents autorisés : cours

Dr. Anicet EBOU, ediman.ebou@inphb.ci

Objectifs du TD

- Identifier et classer correctement les types de variables statistiques
- Choisir l'échelle de mesure adaptée à chaque variable
- Analyser et critiquer un protocole d'échantillonnage
- Proposer un plan d'échantillonnage adapté à une problématique agronomique

Exercice 1

(30 points - 25 min)

Classification de variables dans une étude agronomique

Contexte. Dans le cadre d'une étude sur la production cacaoyère en Côte d'Ivoire, un chercheur de l'INP-HB collecte les informations suivantes sur un échantillon de 120 planteurs de la région du Lôh-Djiboua :

N°	Variable collectée	Exemples de valeurs
1	Nom du village	Gagnoa, Lakota, Divo
2	Superficie de la plantation	2,5 ha ; 4,1 ha ; 0,8 ha
3	Rendement en fèves sèches	380 kg/ha ; 520 kg/ha
4	Variété principale cultivée	Amelonado, Trinitario, Forastero
5	Nombre d'arbres productifs	210 ; 350 ; 480
6	Niveau d'instruction du planteur	Aucun, Primaire, Secondaire, Supérieur
7	Utilisation d'engrais	Oui / Non
8	Âge de la plantation	8 ans ; 15 ans ; 23 ans
9	pH du sol (couche 0-20 cm)	5,2 ; 6,1 ; 4,8
10	Classe de rendement	Faible (<300), Moyen (300-500), Élevé (>500)

Questions :

1. Pour chaque variable du tableau, indiquez :

- a) Son **type** (qualitative ou quantitative) et sa **sous-catégorie** (nominale, ordinale, discrète, continue).

- b) L'échelle de mesure correspondante (nominale, ordinale, intervalle, rapport).
2. Parmi les variables quantitatives listées, lesquelles ne peuvent **jamais** être négatives ? Justifiez brièvement.
3. La variable « classe de rendement » (n° 10) est dérivée de la variable « rendement » (n° 3).
- a) Quel type d'information perd-on lors de ce passage ?
- b) Dans quel cas ce regroupement en classes peut-il néanmoins être utile ?

Exercice 2

(40 points - 35 min)

Critique et amélioration d'un protocole d'échantillonnage

Contexte. Un étudiant en Master à l'INP-HB souhaite estimer le rendement moyen en arachide dans la région de Bouaké. Il dispose d'une liste des 8 villages de sa zone d'étude et d'un budget lui permettant de visiter **40 parcelles** au total. Voici le protocole qu'il a proposé :

“Je vais me rendre dans le village le plus proche de Bouaké, puis interroger les premiers planteurs que je rencontre jusqu'à obtenir 40 réponses.”

Questions :

1. Identifiez **au moins trois défauts** du protocole proposé. Pour chaque défaut, expliquez quel biais il peut introduire dans les résultats.
2. L'étudiant dispose des informations suivantes sur les 8 villages :

Village	Nombre de planteurs	Proportion
Broukro	120	0,24
Koffikro	80	0,16
Assikro	60	0,12
Djébonoua	50	0,10
N'Zuénoua	45	0,09
Languibonou	40	0,08
Kotokro	35	0,07
Yocoboué	70	0,14
Total	500	1,00

Proposez un **échantillonnage stratifié proportionnel** en utilisant les villages comme strates. Calculez le nombre de planteurs à enquêter dans chaque village.

3. Comparez l'échantillonnage stratifié que vous venez de calculer avec un **échantillonnage aléatoire simple** et un **échantillonnage systématique**. Remplissez le tableau comparatif suivant :

Critère	Aléatoire simple	Stratifié	Systématique
Représentativité			
Facilité de mise en oeuvre			
Coût / logistique			
Risque de biais			
Adapté ici ?			

4. Dans ce contexte agronomique, quel protocole recommanderiez-vous finalement, et pourquoi ?

Exercice 3

(30 points - 20 min)

Concevoir un protocole de collecte de données

Contexte. Vous êtes chargé(e) de conduire une étude sur la qualité nutritionnelle des sols maraîchers de la périphérie de Yamoussoukro. La zone d'étude compte environ 200 exploitations maraîchères réparties sur 5 quartiers périphériques. Vous disposez d'un budget pour analyser **30 échantillons de sol** au laboratoire.

Questions :

- Définissez clairement :
 - La **population statistique** de cette étude.
 - L'**unité statistique**.
 - La **variable d'intérêt principale** et son type.
- Proposez **cinq variables** à mesurer sur chaque exploitation, en indiquant pour chacune son type et son échelle de mesure. Au moins deux variables doivent être quantitatives et deux doivent être qualitatives.
- Décrivez en 5 à 8 lignes un **protocole d'échantillonnage complet** pour sélectionner les 30 exploitations, en justifiant votre choix de méthode.
- Citez **deux sources d'erreur non liées à l'échantillonnage** qui pourraient affecter la qualité de vos données, et proposez une mesure préventive pour chacune.