ANALYSE DES DONNÉES APPLIQUÉES

DE LA PRÉPARATION DES DONNÉES À LA PRODUCTION DES RÉSULTATS



Découvrir, analyser, anticiper et décider. Parcours d'initiation hybride de 10 h sous Stata, sur les vraies données de l'EDS Madagascar 2021



CONTENU

01 INTRODUCTION

02 DESCRIPTION

03 COMPÉTENCES ACQUISES

04 INTERÊT DU COURS

05 POUR QUI EST LE COURS ?

06 CONTENU DU COURS

07 MODALITÉS



INTRODUCTION

Comment passer d'un fichier d'enquête brut à des résultats fiables ? Comment nettoyer une base sans biaiser les résultats, choisir le bon test et l'interpréter correctement ? En 10 h sur 3 semaines, faites-le sous Stata, sur les vraies données de l'EDS Madagascar 2021, avec une seule matinée bloquée.

DESCRIPTION

Disposer de données ne suffit pas : il faut les nettoyer, les résumer, puis les comparer et les tester sans leur faire dire plus qu'elles ne permettent. Ce cours associe théorie et pratique : une unité en ligne sur les fondamentaux, une matinée en présentiel pour installer Stata et écrire son premier do-file, puis des unités à son rythme sur trois jeux de données : **une base de suivi de 480 ménages au format Kobo** (avec de vraies erreurs de saisie), **le registre électoral 2023 des 19 340 fokontany** et **l'EDS Madagascar 2021** (20 510 ménages). 17 capsules vidéo, 2 devoirs corrigés et 2 webinaires courts de correction ; scripts R équivalents fournis.

À LA FIN DU COURS, VOUS SAUREZ



Nettoyer et préparer une base :
import Excel ou Kobo, étiquettes, doublons, codes de non-réponse, valeurs impossibles.



Identifier les types de variables et l'unité d'analyse, et choisir le résumé approprié (pourcentage, moyenne, médiane, écart-type).



Utiliser Stata pour des analyses reproductibles :
un do-file commenté, avec son journal.



Créer des graphiques clairs pour des profils non techniques et les exporter vers Excel avec vos tableaux.



Tirer un échantillon, calculer un intervalle de confiance, choisir et lire un khi-deux, un test de proportions ou un test de Student.



Interpréter et présenter les résultats (data storytelling) en 3 phrases, sans confondre différence et effet.

INTERÊT DU COURS

AUTONOMIE

refaites seul.e toute la chaîne, du fichier brut au tableau interprété, sans dépendre d'un statisticien,

RIGUEUR

évitez les erreurs qui invalident une étude ; dans l'EDS, un code 996 oublié fait passer le temps d'accès à l'eau de 25 à 228 minutes,

IMPACT

chaque analyse se termine par 3 phrases prêtes pour un rapport ou un comité de pilotage,

ADAPTABILITÉ

Découvrez et maîtrisez des méthodes efficaces, adaptées aux réalités du terrain où les données sont imparfaites dans 90% des enquêtes sur terrain

POUR QUI EST LE COURS ?

Des étudiants de Master et doctorants (mémoire, thèse) et des chercheurs,

Des chargés d'études, assistants de recherche et chargés de suivi-évaluation

Des consultants et chefs de projets qui produisent des études et des rapports.

Aucune maîtrise des statistiques ou de la programmation n'est exigée : il suffit d'être à l'aise avec Excel. Stata est fourni et installé par iLab ; chaque participant télécharge gratuitement l'EDS avec son propre compte DHS.

CONTENU DU COURS

1. **Fondements de l'analyse de données (1 h) :** types de variables, unité d'analyse, fiche d'indicateur.

2. **Préparation et gestion des bases de données (4 h) :** do-file, import, étiquettes, nettoyage, création de variables ; devoir 1 sur une nouvelle base de 360 ménages.

3. **Statistiques descriptives :** effectifs, pourcentages, moyenne, médiane, mode, dispersion.

4. **Échantillon et analyses bivariées (2h 15) :** intervalle de confiance, p-valeur, khi-deux, test de proportions, test de Student. Corrélations, ANOVA et tests non paramétriques : formation de 40 h.

5. **Analyse et visualisation (1 h 15) :** tableaux croisés, graphiques, export Excel, facteur de confusion ; devoir 2 sur l'EDS.

6. **Application pratique sur STATA et R :** Stata 18 de bout en bout, do-files « faites avec moi » et corrigés ; scripts R équivalents.

MODALITÉS

APPROCHE PÉDAGOGIQUE

1 matinée en présentiel (3 h), 5 unités à votre rythme (5 h 30) avec 17 capsules vidéo, 2 webinaires de 45 min (1 h 30) ; démonstrations, exercices, 2 devoirs corrigés, correction collective.

LIVRABLES PÉDAGOGIQUES

supports de cours, aide-mémoire Stata, bases d'exercice, do-files et corrigés, scripts R, enregistrements des webinaires ; attestation de formation (10 h) avec une note d'au moins 70 % (quiz, 2 devoirs, participation).

EN PRATIQUE

hybride, 10 h sur 3 semaines, plateforme en ligne du site iLab · 15 à 20 participants · prérequis : Excel, portable avec droits administrateur, compte DHS gratuit · recommandé : Éthique de la recherche et de l'évaluation · [prix et dates à confirmer], prix déduit de la formation de 40 h · inclus dans Premium.



DÉROULEMENT : Le parcours en 3 semaines

Semaine	Séances	Format	Durée	Vous produisez
1	Les données du S&E	En ligne	1h	Fiche d'indicateur
1	Prise en main de Stata, résumer, nettoyer	Présentiel	3h	Base propre, do-file
1	Devoir 1 : nettoyer une nouvelle base	En ligne	1h	**Devoir 1**
2	De l'échantillon à la population	En ligne	1h 15	Échantillon et IC
2	Tester une différence sur l'EDS	En ligne	1h	Do-file EDS
2	Webinaire 1 : mettre en commun et tester	En direct	45 min	Un test ensemble
2-3	Devoir 2 : le livrable EDS	En ligne	1h 15	Devoir 2
3	Webinaire 2 : corriger, interpréter, conclure	En direct	45 min	Attestation

DES DONNÉES AUX DÉCISIONS :

donnez du sens à vos résultats avec iLab !