



HAL
open science

Projets d'exploitation de données

Pierre Levasseur

► **To cite this version:**

| Pierre Levasseur. Projets d'exploitation de données. Master. France. 2018. hal-02942843

HAL Id: hal-02942843

<https://hal.inrae.fr/hal-02942843v1>

Submitted on 18 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PROJET D'EXPLOITATION DE DONNÉES

Enseignant référent : Pierre Levasseur, pierre.levasseur@u-bordeaux.fr (D323)

Titre du projet : Les nouveaux enjeux nutritionnels des pauvres dans les pays émergents : le cas du Mexique.

Mots clés : Mexique ; Pauvreté ; Connaissances et habitudes nutritionnelles ; Obésité ; Programmes sociaux.

Compétences techniques à exploiter au cours du projet :

- « Décrypter » une enquête ménage dans une langue étrangère (espagnol)
- Fusion de bases de données
- Construction d'indicateurs de richesse et de santé (uni- et multi-dimensionnels)
- Analyse descriptive et régressions bivariées et multivariées (linéaires et binomiales)
- Lecture socioéconomique des problématiques nutritionnelles
- Mettre en œuvre une stratégie d'identification capable d'établir un lien de causalité

Base de données à exploiter : Encuesta Nacional de Salud y Nutricion – ENSANUT, 2016
http://ensanut.insp.mx/ensanut2016/descarga_bases.php#.WniVjSFmpAE

Contexte d'analyse

À partir d'une enquête récente portant sur les conditions de vie et de santé des ménages mexicains (**Encuesta Nacional de Salud y Nutricion – ENSANUT, 2016**), le groupe de travail devra étudier le lien qui existe entre le niveau de richesse du ménage et les connaissances et habitudes nutritionnelles des individus, puis ensuite (éventuellement) évaluer le rôle des programmes sociaux comme vecteur d'amélioration des conditions de santé des populations pauvres.

Après un état des lieux synthétique du panorama nutritionnel dans lequel se trouve aujourd'hui le Mexique, en s'appuyant notamment sur des statistiques descriptives pertinentes, le rendu final devra permettre de répondre aux questions suivantes :

- *Existe-t-il un lien entre le niveau de richesse du ménage et ses habitudes nutritionnelles ?*
- *Si oui, quelle est la forme fonctionnelle de ce lien (linéaire ou non-linéaire)?*
- *Quelle sont les problématiques nutritionnelles auxquelles les pauvres sont aujourd'hui confrontés ?*
- *Est-ce que ces problématiques varient en fonction de facteurs observables, tels que le genre, le niveau d'éducation et la zone d'habitat ?*
- *Quel est le rôle des programmes sociaux face à l'émergence de ces nouveaux enjeux nutritionnels ?*
- *Les programmes de transferts de fonds conditionnels représentent-ils une solution ?*

Protocole d'investigation

Ce protocole d'investigation en quatre étapes est simplement indicatif. Il doit servir de guide, mais ne constitue pas une finalité ; d'autres voies peuvent être empruntées et développées (tant qu'elles restent pertinentes).

Pour réaliser ce projet, il sera indispensable de fusionner entre eux les modules d'enquête jugés pertinents pour l'analyse. Le but est de rassembler dans une base « propre » les variables d'intérêt sur lesquelles on souhaite potentiellement travailler. Il faut également penser à intégrer des variables de contrôle (âge, genre, éducation, emploi, facteurs sociodémographiques et communautaire, etc.)¹. L'enjeu sera de faire attention aux différentes échelles d'analyse lors de la fusion des différentes bases (variables d'identifiant individuel, ménage, communautaire).

Le choix des techniques d'analyse quantitative est libre, mais devra à chaque fois être justifié. Il est vivement conseillé au groupe de travail de présenter leurs projets et avancés à l'enseignant référent afin de convenir de la stratégie empirique la plus adéquate.

1. Générer et comparer des indicateurs de richesse/pauvreté

Vous trouverez l'ensemble des variables nécessaires à l'élaboration d'indicateurs de richesse dans le module « Hogar » de la base ENSANUT 2016, en particulier le sous-module « *hogar_socioeconomico_NSE* ».

L'objectif de cette partie sera de comparer les différentes approches permettant de mesurer le niveau de richesse du ménage : indicateurs brutes (revenu, patrimoine, etc.) versus indicateurs multidimensionnels (indicateurs composites, analyse factorielle des correspondances, etc.). Il est également attendu que les étudiants comparent leurs indicateurs avec la variable de statut socioéconomique directement calculée et fournie par l'enquête.

Au préalable, les étudiants devront se questionner sur les aspects qui apparaissent aujourd'hui (ou du moins en 2016) comme des marqueurs de richesse au Mexique, un pays à revenu intermédiaire élevé. Autrement dit, quelles sont les variables les plus discriminantes pour distinguer les ménages « riches » des ménages « moins riches » ? Notons qu'une étude comparée en fonction de la zone d'habitat peut se révéler intéressante : la pauvreté se mesure-t-elle différemment en ville qu'à la campagne ?

2. Sélectionner des indicateurs de santé et de nutrition (variables de résultat)

La base de données ENSANUT 2016 offre une information très riche sur la santé et la nutrition des Mexicains dans les modules prévus à cet effet (« Salud » et « Nutricion »).

L'objectif est de rassembler dans une même base de données un jeu de variables permettant d'apprécier les modes de vie et de nutrition du ménage ou de l'individu (échelle d'analyse à déterminer).

Même si le choix des indicateurs de santé et de nutrition reste libre, nous suggérons aux étudiants de s'intéresser particulièrement aux connaissances et habitudes nutritionnelles des ménages ou individus². Notons que le calcul des taux de non-réponse pourra se révéler utile pour sélectionner les variables permettant de maximiser le nombre d'observations.

¹ Voir notamment le sous-module « *hogar_integrantes_procesada* » dans le module « Hogar ».

² Notamment aux sous-modules suivants : « *percepcion* », « *actividad fisica* », « *enfermedad cronica* ».

3. Analyse de la relation entre les indicateurs de richesse et les indicateurs de santé et nutrition

À l'aide d'outils statistiques exploratoires (régression bivariées et multivariées), cette partie devra étudier comment les indicateurs de nutrition et de santé sélectionnés dans la partie précédente varient en fonction du niveau de richesse du ménage. L'objectif est de déterminer quelles problématiques nutritionnelles touchent davantage les pauvres aujourd'hui au Mexique.

On attend dans cette partie que les étudiants se questionnent sur la présence d'hétérogénéité observable (liée par exemple au genre, au niveau d'éducation, à la zone d'habitat) et non-observable (préférences individuelles) capable de biaiser les estimations. De même, l'éventualité d'une causalité inverse devra aussi faire l'objet d'une discussion.

Les étudiants devront enfin proposer une stratégie d'identification permettant d'établir un lien de causalité en ce qui concerne la relation richesse/nutrition.

(BONUS) L'impact des programmes sociaux sur les connaissances et habitudes nutritionnelles des pauvres

Des points bonus seront délivrés aux étudiants qui prendront l'initiative d'évaluer l'impact du programme de transferts de fonds conditionnels mexicain (Progres-Oportunidades-Prospera) sur les indicateurs de santé et de nutrition sélectionnés. On peut en effet se demander si la participation au programme, qui prévoit une incitation monétaire en contrepartie de contrôles de santé réguliers et de réunions de prévention sanitaire et nutritionnelle, a un effet sur le niveau de connaissance nutritionnelle des membres du ménage, de même que sur leurs habitudes et leur santé.

La variable de participation au programme est disponible dans le sous-module « *programas* », du module « *Nutricion* ».

Pour évaluer l'impact du programme, les étudiants devront mettre en place une stratégie d'identification capable de corriger le possible biais de sélection « caché » qui existe entre les ménages participants et les ménages non-participants. Il est en effet possible que des facteurs non-observables expliquent simultanément le choix de participer et les indicateurs de santé et de nutrition, comme l'intérêt du ménage aux questions de santé et d'éducation.

Nous conseillons aux étudiants de s'inspirer d'Andalón (2011)³ en mettant en place une méthode de régression de discontinuité de type floue (*Fuzzy Regression Discontinuity Design*) à partir de l'un des indicateurs de richesse multidimensionnels créés dans la partie 1. L'idée sous-jacente à cette approche est de déterminer si le passage du statut de « non-traité » au statut de « traité » engendre un écart significatif de la variable de résultat pour les individus ayant un niveau de richesse similaire.

L'originalité de l'étude pourra être par exemple d'étudier l'impact du programme sur les connaissances et habitudes nutritionnelles des ménages bénéficiaires, thématique encore peu explorée par la littérature.

³ ANDALÓN Mabel, 2011, « Oportunidades to reduce overweight and obesity in Mexico? », *Health economics*, 20 Suppl 1, p. 1-18.