Les données administratives : quelques références internationales à partir du site du projet Big_Stat

Giulia Ferrari

Institut National d'études démographiques (INED)

giulia.ferrari@ined.fr

Rencontres de statistique appliquée en collaboration avec le Pôle Perspectives internationales

INED, 10 décembre 2018

Le projet Big_Stat

- Projet financé par l'ANR (mars 2017 - mars 2022)
- Responsable du projet: Laurent Toulemon
- Responsables du site-web : Arnaud Bringé et moi

Objectifs:

- Recherche
  - → Population et situations familiales
  - → Jeunes adultes
  - → Enfants de parents séparés
  - → Autres projets : orphelins, immigrés
- Documentation des fichiers et retours vers les producteurs
- Formation des utilisateurs
- Collaborations entre utilisateurs et avec les producteurs
  - → En utilisant les données nous comprenons les enjeux
  - → Et nous fournissons des variables créées et validées par les chercheurs

Le site web Big_Stat : structure

- www://big-stat.site.ined.fr
- Description du projet
- Sources de données et modalités d'accès
- Recherches en cours et évènements
- Liens aux
  - → définitions & théories
  - → open data
  - → données contextuelles
  - → utilisation des données administratives en France et à l'étranger
- Publications

Le site web Big_Stat : sources de données

- Des sources exploitées par les membres du projet (en cours)

Éléments présents :

- Description de la source
- Accès aux données
- Documentation
- Contribution à son amélioration

Le site web Big_Stat : définitions et théories

- **Que sont les données administratives?**
  - → Des données dérivées des opérations des systèmes administratifs, typiquement du secteur public
  - → Elles sont souvent liées à la prestation d'un service
  - → Exemples de systèmes administratifs qui offrent leurs données:
    - Éducation
    - Santé
    - Impôts
    - Registres électoraux
    - Registres et recensement

  - → Très importantes puisqu'elles peuvent aider à repondre à des questions fondamentales en sciences sociales et sans coût de collecte

Le site web Big_Stat : définitions et théories

- Liste d'articles et documents de travail (en français et en anglais) sur :
  - → le besoin croissant de mobilisation de données existantes (Heran & Sebag, 2000)
  - → les opportunités offertes par les données administratives en sciences sociales et leurs défis (Connelly et al. 2016)
  - → la comparaison entre données de sources administratives et données d'enquête
  - → les enjeux du couplage des données issues des sources différentes

Le site web Big_Stat : données contextuelles

- Les données contextuelles sont très utiles dans les recherches en sciences sociales pour mieux comprendre les comportements individuels.
- GGP contextual database
  niveaux national et infra-national
  (Caporali et al. 2013, 2016)

- The LawsAndFamilies database
  (Digoix et al. 2016)

Le site web Big_Stat : open data

- Ce sont des données dont l'accès et l'usage sont libres de droit pour les usagers, sans restriction technique, juridique ou financière.
- Elles peuvent être utilisées dans des analyses sur fichiers individuels, en tant que données contextuelles, ou pour des cadrages de données d'enquêtes.
- www.data.gouv.fr recense la très grande majorité des données ouvertes disponibles.
- La DREES recense des exemples dans le domaine de la santé

Le site web Big_Stat: exemples d'utilisation des données administratives

- Utilisation croissante de ce type de données
- Impossible de donner une liste exhaustive
  - →Quelques exemple d'utilisation des données françaises
    - EDP : publications, CASD
  - →Panorama sur les principaux pays européens utilisant ce type de données.
    - Poulain & Herm 2013
    - Pays Nordiques

Conclusions

- Opportunités offertes par les données administratives en sciences sociales :
  - → Échantillons plus larges par rapport aux enquêtes sociales traditionnelles : cela permet l'étude des évènements rares et des sous-groupes de population spécifiques
  - → Possibilité de « panelisation » de données (beaucoup d'attrition dans les enquêtes)
  - → Évaluation des politiques publiques et support quantitatif aux politiques publiques ciblées

Conclusions

- Défis qui peuvent se poser avec les données administratives :
  - →Accès aux données et respect de la vie privée
  - →Échantillons des bénéficiaires d'un certain service
  - →Enjeux liés aux collectes de données contextuelles en rétrospectif (Schoumaker et al. 2006)
  - →Gros effort de compréhension et d'harmonisation des procédures qui ont conduit à la production des données

Conclusions

- Défis qui peuvent se poser avec les données administratives :
  - →Construction de données statistiques : appariement, nettoyage, harmonisation
  - →Documentation des fichiers et variables construites, collaboration entre producteurs et utilisateurs