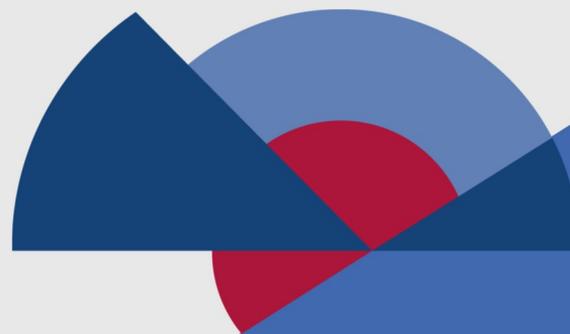


D'Octopusse à Nautille

ou changement de source pour la base de sondage des enquêtes ménages de l'Insee

29 novembre 2018



Séminaire de Méthodologie Statistique - Fidéli-Nautille : une nouvelle source d'informations et base de sondage sur les individus et les ménages



01

Utiliser Fidéli comme base de sondage

02

Le nouvel échantillon-maître dans Fidéli

01

Utiliser Fidéli comme base de sondage



1.1. Les qualités d'une bonne base de sondage

1.2. La base actuelle : le recensement de la population

1.3. Fidéli, une source envisageable ?

- Unicité
 - Toutes les unités du champ, une et une seule fois
- Fraîcheur
 - Des informations à jour, correspondant aux unités enquêtées *in fine*
- Exhaustivité
 - Ensemble complet du champ
- Informations auxiliaires intéressantes
 - Possibilité d'avoir des plans de sondage de qualité
 - Permettre un repérage précis

- Depuis les années 60
 - Utilisation des recensements de la population pour le tirage des échantillons des enquêtes ménages

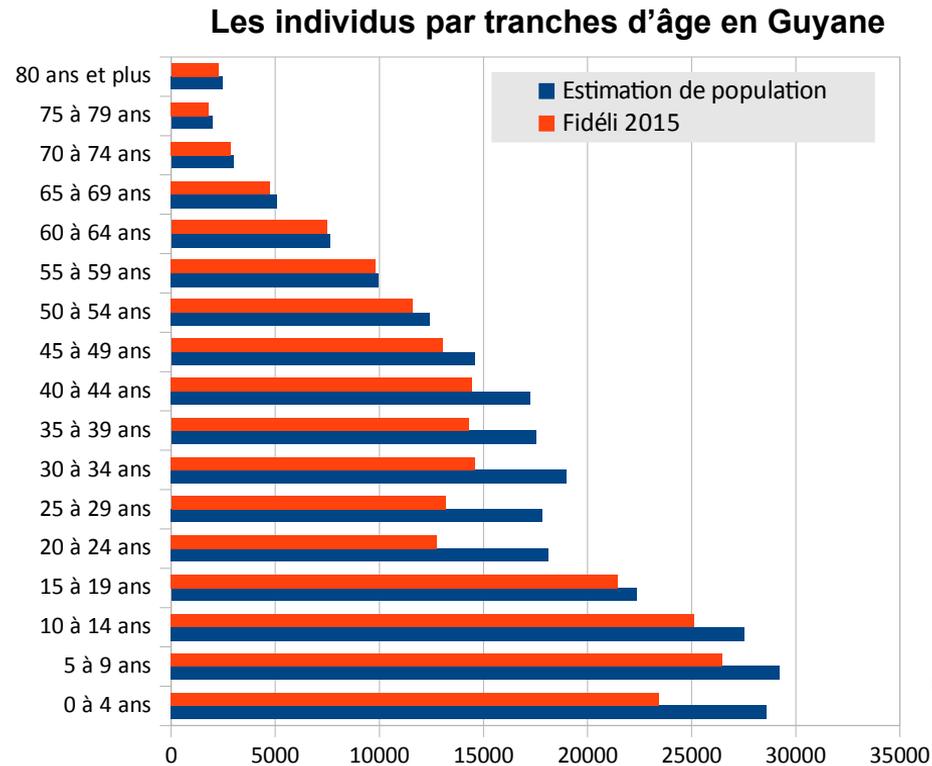
- Unicité
 - OUI : on ne recense qu'une fois
- Fraîcheur
 - NON : un recensement tous les 8 ans environ
- Exhaustivité
 - OUI
- Informations auxiliaires intéressantes
 - Possibilité d'avoir des plans de sondage de qualité : OUI
 - Permettre un repérage précis : OUI

- Depuis le recensement rénové de 2004
 - Le recensement devient rotatif :
 - Communes de moins de 10 000 habitants réparties en 5 groupes de rotation
 - Enquête exhaustive d'un groupe chaque année
 - Adresses des communes de 10 000 habitants et plus réparties en 5 groupes de rotation
 - Enquêtes de 40 % des adresses d'un groupe chaque année
 - Tirage des échantillons des enquêtes ménages dans la dernière EAR

- Unicité
 - OUI : on ne recense qu'une fois
- Fraîcheur
 - mOUI : information de l'année précédente, mais uniquement sur une partie de la population
- Exhaustivité
 - NON : un seul groupe de rotation chaque année
 - NON : une enquête dans les grandes communes
- Informations auxiliaires intéressantes
 - Possibilité d'avoir des plans de sondage de qualité : OUI
 - Permettre un repérage précis : OUI

- Unicité
 - OUI : Traitement des doublons par Fidéli
- Fraîcheur
 - OUI : information de l'année précédente sur toute la base
- Exhaustivité
 - OUI : France entière couverte, y compris Mayotte
- Informations auxiliaires intéressantes
 - Possibilité d'avoir des plans de sondage de qualité : OUI
 - Permettre un repérage précis : OUI
- Et en plus : une seule base de sondage pour toutes les enquêtes !

- Même si localement, on peut trouver des différences
 - Notamment dans les DOM



- D'autres variables pour la stratification et les post-traitements
 - Quelques variables ne sont plus disponibles
 - Diplôme
 - Catégorie socio-professionnelle
 - Mais de nouvelles variables importantes
 - Montant des revenus individuels (et par logements)
 - Types de revenus (salaires, pensions...)

- D'autres variables pour le repérage :
 - Disparition des informations « recensement »
 - Repérage « normalisé » pour le recensement
 - Mais déjà, des défauts avec le recensement par internet
 - De nouvelles opportunités d'interrogation et de repérage à étudier
 - Une information géolocalisée et géocodée
 - Présence du mail et de numéros de téléphones des déclarants
 - Informations complémentaires sur l'adresse
 - Dépendance, présence de commerces, communautés...
 - Plusieurs adresses pour un individu

- Valider la cohérence des concepts
 - Le logement ordinaire
 - Des logements connus de la taxe d'habitation
 - D'autres logements aux contours plus flous
 - Présence de communautés
 - Foyers fiscaux non rattachés à des structures identifiées
 - Le concept de résidence principale
 - Des intérêts divergents entre bases administratives et enquêtes de recensement

- Les risques potentiels d'utiliser cette nouvelle base
 - Mauvaise couverture de la base
 - Hors champ et champ non couvert
 - Doubles comptes
 - D'autres variables
 - Mauvais repérage
 - Stratification moins pertinente
 - Un concept différent
 - Impact sur les estimations produites
 - Réflexion sur la définition du logement ordinaire, vis-à-vis du recensement

02

Le nouvel échantillon- maître dans Fidéli



2.1. Le principe de l'échantillon-maître

2.2. L'échantillon-maître actuel

2.3. Pourquoi changer ?

2.4. Le nouvel échantillon-maître

- Les enquêtes ménages de l'Insee :
 - Des thèmes variés
 - Enquêtes sur les revenus, la santé, les conditions de vie, la mobilité...
 - Souvent en face-à-face
 - Devant représenter tout le territoire
- L'Insee possède un réseau d'enquêteurs sous contrat
 - Mais un nombre limité d'enquêteurs...
 - ...ne pouvant pas arpenter l'ensemble du territoire.

→ Recours à la méthode de l'échantillon-maître (EM)

- D'un point de vue pratique



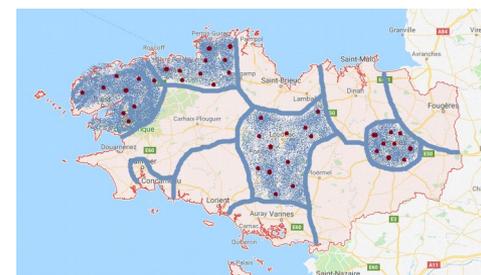
Créer une partition du territoire



Sélectionner *intelligemment* les zones



Affecter un enquêteur à chaque zone



Tirer des logements dans chaque zone, pour chaque enquête, à partir d'une base de sondage.

- D'un point de vue théorique : tirage à deux degrés

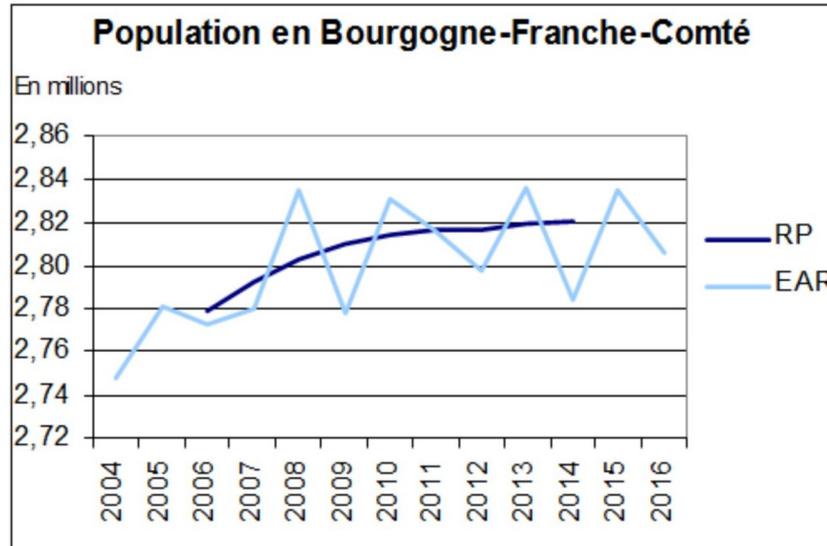
- 1^{er} degré : Tirage d'unités primaires (UP), les zones constituant l'échantillon-maître
 - Un échantillon de zones représentant le territoire
 - Des zones qu'un enquêteur peut couvrir
- 2^e degré : Tirage d'unités secondaires au sein de chaque unité primaire, les logements

- Avec le recensement rénové de 2004
 - Partition du territoire en unités primaires (les ZAE)
 - Comme ensemble de communes contiguës
 - Sous contraintes :
 - Chaque UP possède suffisamment de logements pour que l'enquêteur affecté ait du travail chaque année.
 - Des logements de chaque groupe de rotation dans chaque UP
 - Chaque UP possède suffisamment de logements pour éviter des réinterrogations
 - au moins 300 résidences principales de chaque groupe de rotation dans chaque UP

- Avec le recensement rénové de 2004
 - Tirage équilibré de l'échantillon-maître en 2009
 - équilibré sur les 5 groupes de rotation
 - Au deuxième degré, sélection des logements
 - au sein des unités primaires de l'échantillon-maître
 - parmi les logements recensés dans l'année (EAR).

- Un besoin régulier de renouveler l'échantillon-maître
 - Ne pas toujours réinterroger les mêmes logements
 - Épuisement des zones de collecte
 - Des zones qui ne représentent plus le territoire
 - Perte de qualité de l'échantillon-maître

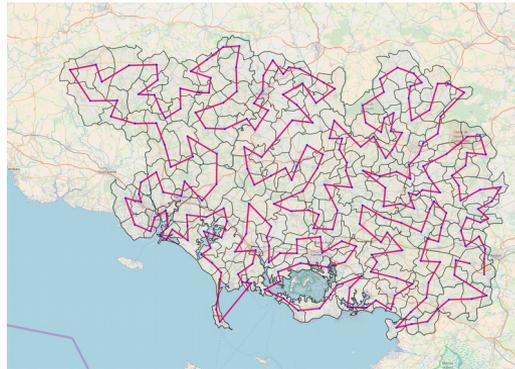
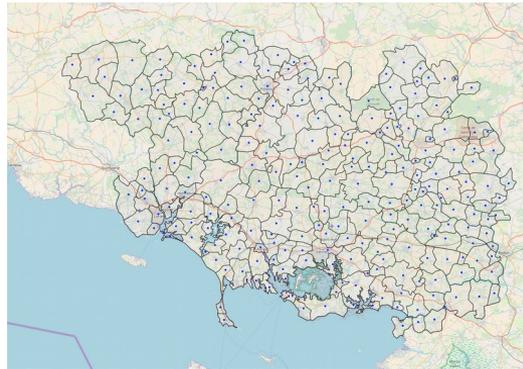
- Une méthode à revoir
 - Déséquilibre des groupes de rotation et donc de tout échantillon-maître issu de la même méthode



Évolution des estimations de la population de la Bourgogne-Franche-Comté entre 2004 et 2016, à partir du RP d'une part et de l'EAR d'autre part.

- Un système trop complexe
 - Répartition en groupes de rotation
 - Sélection d'adresses
 - Partition liée au groupe
 - Tirage équilibré par groupe
 - ...

- Partition du territoire en 5000 zones
 - Ensembles de communes
 - De 2500 résidences principales minimum
 - Les moins étendues possible
 - Grâce à une réponse au problème du voyageur de commerce



- L'abandon du recensement de la population permet de s'affranchir de la contrainte sur les 5 groupes de rotation
 - Des unités primaires avec plus de résidences principales
 - donc pouvant durer plus longtemps, sans réinterrogation
 - Des unités primaires moins étendues
 - inutile d'aller chercher une commune de chaque groupe dans chaque unité primaire

- Sélection de l'échantillon-maître : les objectifs
 - Obtenir un échantillon-maître représentatif du territoire
 - pour un maximum d'enquête, sur des thèmes différents
 - pendant 10 ans
 - Respecter les contraintes de précision imposées par Eurostat
 - Sélectionner un nombre de zones que le réseau des enquêteurs peut couvrir

- Sélection de l'échantillon-maître : la méthode
 - Tirage spatialement équilibré (Grafström et Tillé, 2008)
 - Équilibrage par région
 - Sur des variables majeures
 - Sur les premiers axes d'une ACP sur les UP, portant sur un grand nombre de variables socio-démographiques
 - Spatialement équilibré
 - Sélectionner des unités réparties sur le territoire
 - Pour bénéficier de l'auto corrélation spatiale
 - Conséquences :
 - L'échantillon-maître n'est pas directement équilibré sur toutes les variables socio-démographiques souhaitées
 - Mais est plus représentatif pour un plus grand nombre de variables et d'enquêtes

- L'utilisation de Fidéli permet
 - d'utiliser des variables de revenus
 - Très explicatives pour beaucoup d'enquêtes
 - de s'affranchir de contraintes de représentativité sur les 5 groupes de rotation
 - Moins de contraintes, donc plus de précision
 - Ou moins de zones tirées dans l'échantillon-maître, et donc moins de territoire à couvrir

Conclusion



- Une simplification de la méthode par l'abandon du recensement de la population au profit de Fidéli
 - Gain en précision en s'affranchissant des groupes de rotation
 - Utilisation d'une base exhaustive
- Mais un changement de source à anticiper
 - Modification des variables disponibles pour la stratification et les post-traitements
 - Changement d'informations pour le repérage
 - Modification de l'appréhension du champ des enquêtes ménages du fait de sources divergentes
- Une méthodologie innovante pour la sélection du nouvel échantillon-maître
 - Tirage spatialement équilibré sur variables et axes d'ACP
 - Coordonné avec l'échantillon de l'enquête emploi en continu

Retrouvez-nous sur :

insee.fr



Ludovic VINCENT

Pierre-Arnaud PENDOLI

Thomas MERLY-ALPA

DMCSI