

# Méthodologie d'estimation dans l'enquête Sans-Domicile 2012

---

Alexandre Lebrère



# Plan

---

Introduction: Le problème posé

1. L'apurement des questionnaires

2. Le calage des prestations

3. L'imputation des semainiers et l'échantillonnage indirect

# Introduction : position du problème

---

On dispose en sortie de collecte de trois ensembles de questionnaires :

- Questionnaires longs (francophones) : QIL
- Questionnaires courts francophones : QAAF
- Questionnaires courts non francophones : QAANF

On dispose d'autre part de différents jeux de poids :

- Poids agglomération
- Poids service x jour
- Poids contact
- Poids prestation

# Introduction : position du problème

---

L'information collectée concerne :

- Les questionnaires individuels,
- Les services,
- Les service x jours (= contacts).

On veut obtenir des poids individuels

- Sur la population des usagers de services d'aide,
- Sur la population des usagers francophones
- Sur la population des usagers non francophones

Deux questions sont posées : représenter au mieux l'univers des prestations avec l'information disponible, et passer d'un poids prestation à un poids individuel.

# Introduction : Que faire ?

---

La démarche suivie pour obtenir des poids individuels est la suivante :

## 1. Obtention des poids services x jours

*remarque : le tirage des réserves impose un traitement complexe.*

## 2. Les poids prestations sont calés sur l'information disponible au niveau service x jours et QILQAAF

*remarque : l'estimation du nombre de prestations est disponible à ce stade.*

## 3. Un partage des poids est effectué pour obtenir des poids individuels.

*On sépare les populations francophone et non francophone.*

# 1. Apurement des données

---

Apurement des données indispensables :

- concernant les services
- nécessaires au calage
- concernant le semainier

# 1. Apurement des données : infos sur le service

---

- Fusions, éclatements, doublons : redressement sur les poids de tirage
- Fréquentation : nombre de personnes ayant fréquenté le service le jour de la visite
- Données service exploitables :
  - fréquentation renseignée
  - nombre de contacts
  - nombre de femmes francophones
  - nombre d'hommes francophones
  - nombre de femmes non francophones
  - nombre d'hommes non francophones

# 1. Apurement des données : QIL (1)

## BON REMPLISSAGE

**B2** Où avez-vous mangé les 7 derniers jours ?

code restauration p. 10- 11 et **carte 4**

Pour les distributions de repas à l'extérieur, coder le service et indiquer le nom de l'organisme et l'adresse la plus précise possible du lieu de distribution (ex : Restos du Cœur derrière la gare de Perrache, LYON, 69)

Jour semaine	Repas du matin	Repas du midi
<b>J-1</b> (hier)	Type de repas   0   1   RMA1T Si type 01 → Code service   3   4   1   4   E   P   ↓ nom : ..Petits déjeuners du Cœur..... .....RMA1S	Type de repas   0   1   RM1T Si type 01 → Code service   7   8   6   5   E   M   ↓ nom : Restos du Cœur..... .....RM1S
Lun	.....RMA1N	.....RM1N
Mar	adresse : ..13 rue du Palais..... .....RMA1A	adresse : 13 rue du Palais..... .....RM1A
Mer	.....RMA1C	.....RM1C
Jeu	commune : ..Bordeaux..... .....RMA1D	commune : ..Bordeaux..... .....RM1D
Ven	dép   3   3	dép   3   3
Sam		
Dim		
<b>J-2</b>	Type de repas   0   1   RM2T Si type 01 → Code service   3   4   1   4   E   P   ↓ nom : .....IDEM..... .....RMA2S	Type de repas   0   1   RM2T Si type 01 → Code service   7   8   6   5   E   M   ↓ nom : .....IDEM..... .....RM2S
Lun	.....RMA2N	.....RM2N
Mar	.....RMA2A	.....RM2A
Mer	adresse : ..... .....RMA2C	adresse : ..... .....RM2C
Jeu	.....RMA2D	.....RM2D
Ven	commune : ..... .....RMA2D	commune : ..... .....RM2D
Sam	dép	dép
Dim		
<b>J-3</b>	Type de repas   0   1   RM3T Si type 01 → Code service   3   4   1   4   E   P   ↓ nom : .....IDEM..... .....RMA3S	Type de repas   0   1   RM3T Si type 01 → Code service   7   8   6   5   E   M   ↓ nom : .....IDEM..... .....RM3S
Lun	.....RMA3N	.....RM3N
Mar	.....RMA3N	.....RM3N
Mer	.....RMA3N	.....RM3N
Jeu	adresse : ..... .....RMA3N	adresse : ..... .....RM3N
Ven	.....RMA3N	.....RM3N
Sam	.....RMA3N	.....RM3N
Dim	.....RMA3N	.....RM3N

# 1. Apurement des données : QIL (2)

---

- Semainier très important pour la pondération
  - La vérification des codes services était réalisée lors de la saisie
  - Traitement des prestations dans le champ avec une adresse renseignée et un code service à blanc ou noté inconnu
  - Pour un code service : vérification des adresses du semainier et de la base
  - Vérification des extensions des codes services : doivent correspondre à la bonne prestation
  - Vérification des prestations codées hors champ avec un code service
  - Déterminer si le semainier est exploitable ou non
- Pour les exploitables :
  - Imputations
  - Calcul du nombre de liens

# 1. Apurement des données : QIL (3)

---

QIL exploitable si :

- Semainier exploitable
- On peut déterminer la situation d'hébergement de la personne la veille de l'enquête (ce qui définit sans-domicile ou non) :
  - Sans-domicile : hébergé (logement / hôtel / hébergement regroupé) ou sans-abri
  - Locataire ou propriétaire
  - Sans logement personnel (personnes hébergées par un tiers, squatt)

# 1. Apurement des données : QAA non francophones

---

- Apurement des variables nécessaires au calage :
  - âge, sexe, pays de naissance et diplôme si la personne est née en France
- Semainier agrégé renseigné
- Déterminer la situation d'hébergement de la personne la veille de l'enquête

## 2. Le calage de l'enquête : Introduction

---

**But du calage** : améliorer la précision des estimateurs et assurer la cohérence entre sources différentes de données (et éventuellement corriger de la non réponse) en utilisant de l'information auxiliaire disponible sur la population complète.

*Idée : estimateur par la régression généralisée*

### **Problèmes :**

- Pas de « vraie » information auxiliaire sur la population cible.
- Qualité de la base de sondage pour le calage.

## 2. Le calage de l'enquête : prestations

---

Les poids service x jours et la fréquentation observée donnent l'estimation du nombre de prestations.

Les données obtenues au niveau service x jour (*i.e.* la fiche contacts) sont considérés pour estimer deux variables :

- Sexe
  - Francophone
- 
- On effectue ensuite une post-stratification des poids contacts sur ces deux variables au niveau agglomération x type de service.

## 2. Le calage de l'enquête : auto-calage des questionnaires

---

A ce stade, nous avons un poids prestation utilisable pour le calcul des poids individuels. Cet ensemble de poids représente séparément les populations francophones et non francophones.

L'échantillon des utilisateurs francophones est l'ensemble des QIL. Nous pouvons tirer parti de l'information disponible dans les QAAF.

On fait un calage des QIL sur l'information disponible dans l'ensemble QIL + QAAF.

### 3. Poids individuels : le partage des poids

---

**Problématique du sondage indirect** : la population d'inférence est celle des sans domicile, mais la population (statistique) observable est celle des prestations offertes dans les services d'aide.

**La solution est la méthode du partage des poids** : pour chaque questionnaire de l'échantillon, compter le nombre de liens avec la population.

- *Intuitivement : Il faut diviser le poids prestation par le nombre de fois où l'individu était atteignable par le plan d'échantillonnage.*
- Le nombre de liens est observé dans le semainier.

### 3. Poids individuels : semainiers et liens

---

Les liens sont observés grâce au « semainier. » On note sur les sept derniers jours :

- Dans les QIL : questions détaillées sur les sept derniers jours. 4 prestations enregistrées par jour : petit-déjeuner, déjeuner, dîner, nuit.
- Dans les QAA : quatre questions résument la fréquentation des services au cours de la semaine précédente.

La qualité des semainiers observés est cruciale pour le calcul des poids. Les rapports de poids extrêmes possibles varient de 1 à 22.

Les problèmes de mémoire ont un effet direct sur la variance des estimations.

### 3. Poids individuels : imputation des semainiers

---

Le nombre de liens est indispensable pour avoir un poids individuel. → Il faut absolument utiliser l'information des semainiers

Dans le cas de non réponse partielle, on doit imputer les semainiers. Pour les QIL, on utilise une méthode d'imputation par donneur, en tirant les liens d'un individu ayant les mêmes fréquentations que celles observées dans les données du receveur.

Pour les QAA, un fort effet questionnaire est observable.

# Conclusion

---

- Les difficultés de l'estimation proviennent de plusieurs facteurs :
  - L'échantillonnage complexe, prenant en compte des contraintes diverses (base de sondage, collecte, files)
  - L'absence de données de cadrage
  - Une population difficile à atteindre
  - Une population difficile à interroger

## Les résultats sont de deux ordres

- L'estimation de la taille de la population est la plus fragile, mais semble raisonnable (études de sensibilité)
- Les études de comportement dépendent avant tout de la qualité de l'apurement.