

# Algorithme CURIOS et méthode de priorisation pour les enquêtes en face-à-face : application à l'enquête Patrimoine 2014

Antoine Rebecq, Thomas Merly-Alpa

INSEE, DMCSI, Division Sondages

Journées de Méthodologie Statistique, 02/04/2015

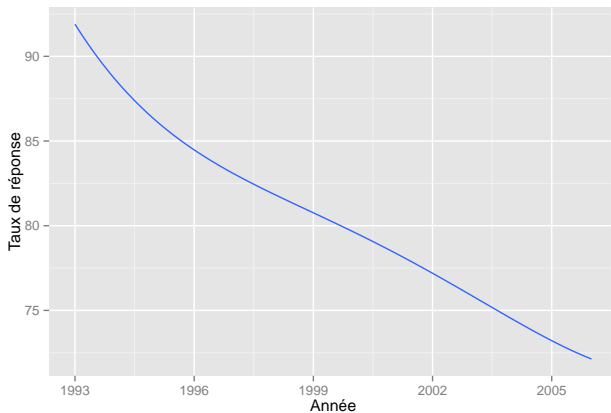


# Sommaire

- 1 La priorisation
  - Problématique
  - La méthode
- 2 Description de CURIOS
  - Collecte en deux vagues
  - Objectifs de l'algorithme
- 3 Application à l'enquête Patrimoine 2014
  - Contexte
  - Résultats

# Introduction

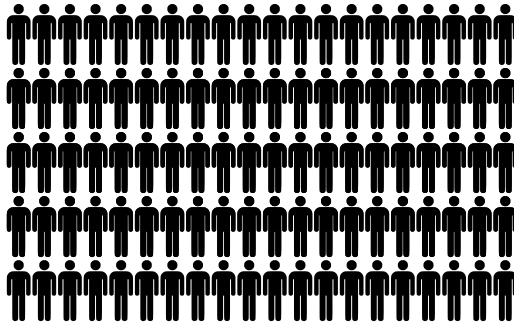
Contexte de baisse des taux de réponse (Särndal, JMS 2012).



(Source : Enquête EDTR sur la Dynamique du Travail et du Revenu - Canada)

# L'échantillonnage usuel

Notre population-cible est



On donne usuellement



# La priorisation

Parmi les individus échantillonnés, certains sont non-répondants (entourés de bleu).



# La priorisation

Les individus rouges ont un comportement de réponse particulier : leur sélection s'effectue selon des indicateurs de représentativité.



# Première vague

Au lieu de donner l'échantillon classique



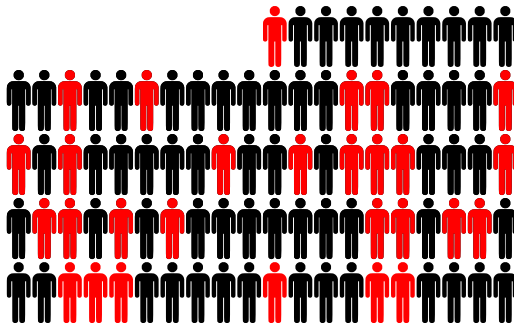
On sépare la collecte en deux vagues



La première vague permet d'identifier les individus à prioriser.

# Retour à la population

On identifie les individus concernés dans la population



et on échantillonne





# Une collecte en deux vagues

Au lieu de donner l'échantillon classique



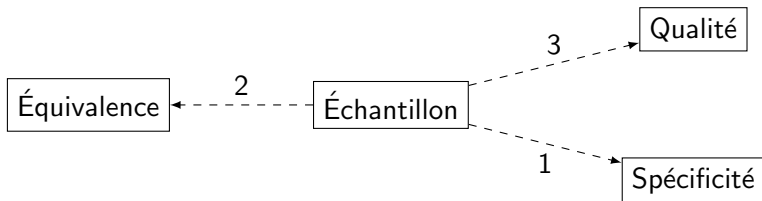
On sépare la collecte en deux vagues



En pratique, on recommande d'attendre au moins 6 semaines et au moins 75% de la collecte.

# Principe de l'algorithme

L'échantillon de deuxième phase est choisi selon des contraintes opposées :



Dans la pratique, on choisit des termes de **Qualité**, **Spécificité** et **Équivalence** qu'on intègre à un programme de minimisation.

# Respecter le plan de sondage initial

L'objectif de **Spécificité** consiste à respecter le plan de sondage spécifiquement créé pour l'enquête.

Par exemple, l'allocation de Neyman peut être utilisée pour optimiser la précision.

En pratique, on veut minimiser l'écart de l'échantillon de seconde vague au plan de sondage initial.

# Minimiser la dispersion des poids

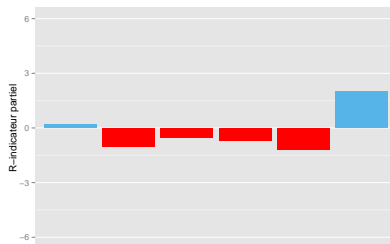
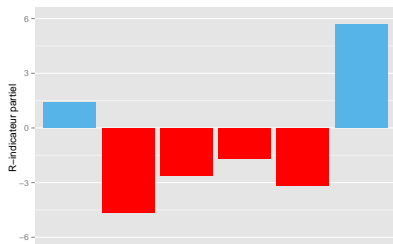
L'objectif d'**Équivalence** consiste à minimiser la dispersion des poids corrigés de la non-réponse.

Une faible dispersion des poids permet une meilleure robustesse, et améliore le processus de calage.

En pratique, pour calculer les poids corrigés de la non-réponse, on utilise des méthodes de GRH sur des critères pertinents.

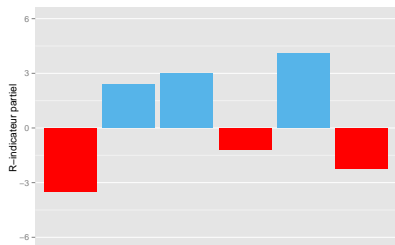
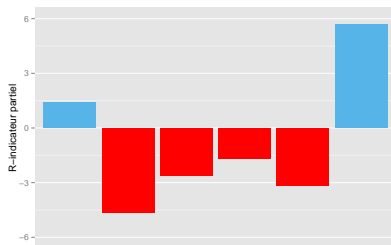
# Les R-Indicateurs

Pour l'objectif de **Qualité**, on utilise des indicateurs de représentativité, tels que les R-indicateurs (Schouten, 2009) qui sont des mesures du manque d'association entre réponse et variables auxiliaires.



# Les R-Indicateurs

La priorisation doit être limitée pour éviter de recréer un déséquilibre :



## Programme d'optimisation

On détermine l'échantillon de phase 2 en minimisant le programme suivant faisant intervenir les objectifs d'équivalence et de qualité :

$$\arg \min_{S_2} \mathbb{E} [\Sigma(w_{CNR}) + \lambda \cdot \Gamma(S)]$$

avec :

$\Sigma(X)$  = dispersion (variance empirique) du vecteur  $X$

$S = S_1 \cup S_2$  = échantillon total

$w_{CNR}$  = poids corrigé de la non-réponse des unités de  $S$

$\Gamma(S)$  = mesure de l'équilibre de l'échantillon de répondants

$\lambda \in [0, +\infty[$

## Anticipation de la seconde vague

L'espérance se calcule sur l'ensemble  $S_1 \cup S_2$  composé des deux échantillons.

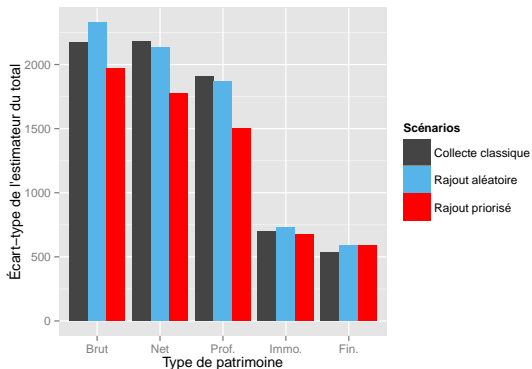
Cela suppose en particulier que l'on connaisse le comportement général de la population enquêtée en seconde vague.

La probabilité de réponse est estimée par un modèle prédictif qu'on suppose semblable à celui en première vague. On utilise des méthodes de Monte-Carlo pour simuler cette non-réponse.



# Étude préliminaire sur Patrimoine 2010

La priorisation réduit les écarts-types pour l'estimation des patrimoines totaux :



# L'enquête Patrimoine 2014

Contexte : Taux de réponses faibles dans certains départements d'Île-de-France.

On applique donc la priorisation à 25% de l'échantillon, à partir de janvier 2015.

Cela concerne 820 Fiches-Adresses.

## Analyse à la fin de la première vague

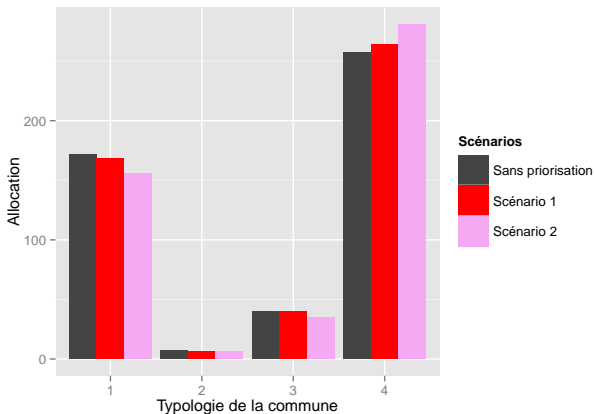
On réalise une étude de la représentativité des répondants à la fin de la première vague, en novembre 2014.

On remarque que les modalités **Locataires**, **Appartements** ainsi que les **Petites Surfaces** sont sous-représentées parmi les répondants.

Cependant, afin de permettre le suivi des recommandations européennes, les groupes d'optimisation choisis sont basés sur la **Strate** et la **Typologie de la commune**.

## Deux scénarios

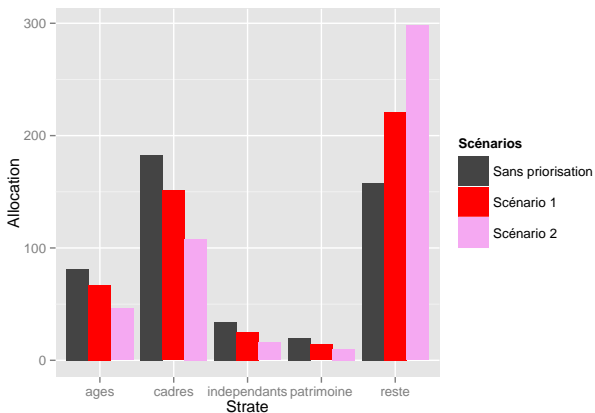
Allocations par groupes d'optimisation (**typo** × strates).



(1 : riches urbaines, 2 : riches rurales, 3 : rurales revenus modérés, 4 : urbaines revenus modérés.)

## Deux scénarios

Allocations par groupes d'optimisation (typo × **strates**).



## Résultats finaux

Variations entre le scénario proposé et le scénario sans priorisation :

Typo. commune × strate	Locataires	Appartements	Petites surfaces
1 × ages	-0.17	-0.06	-0.08
1 × independants	-0.18	0.01	-0.14
1 × revenus patrimoine	-0.24	-0.02	-0.13
1 × cadres	-0.14	0.01	-0.07
1 × reste	<b>0.28</b>	<b>0.09</b>	<b>0.14</b>
(2,3) × ages	-0.21	-0.12	0.05
(2,3) × independants	-0.12	0.02	-0.09
(2,3) × revenus patrimoine	0	0	-0.24
(2,3) × cadres	-0.04	-0.04	-0.08
(2,3) × reste	<b>0.14</b>	<b>-0.27</b>	<b>0.24</b>
4 × ages	-0.10	-0.04	-0.06
4 × independants	-0.23	-0.20	-0.14
4 × revenus patrimoine	-0.18	-0.10	-0.24
4 × cadres	-0.11	-0.01	-0.04
4 × reste	<b>0.28</b>	<b>0.09</b>	<b>0.14</b>

# Conclusion

La méthode de priorisation permet d'adapter les échantillons tirés dans le cadre d'une seconde vague pour une enquête ménage en face-à-face.

Les objectifs de l'algorithme CURIOS sont paramétrables et permettent ainsi une adaptation à toutes les enquêtes.

L'application de la méthode à l'enquête Patrimoine 2014 a permis de tester une première mise en œuvre.

# Conclusion

**Merci pour votre attention !**



<http://nc233.com/2015/03/curios/>



@nc233