

7ème colloque francophone sur les sondages

# Les calculs de précision dans le recensement rénové

---

Gwennaëlle BRILHAULT  
INSEE – DSDS  
Département de la démographie  
Division « Méthodes et traitements des recensements »



# Introduction

---

- **Nombreux facteurs influant sur la qualité des résultats de tout recensement traditionnel :**
  - non-réponse totale, redressements des variables collectées pour non-réponse ou incohérence, erreurs de codification, etc.
- **Introduction du sondage dans le recensement rénové en France : « choc culturel »**
  - nombreux travaux sur la précision liée à l'introduction du sondage
    - notamment présentation au séminaire SFdS consacré au recensement rénové du 25 juin 2002, puis du 19 mai 2011
- **Objectif de la présentation :**
  - point sur les différents travaux portant sur la précision du RP rénové et données de précision du recensement disponibles aujourd'hui

# Plan de la présentation

---

- I. Rappel succinct du plan de sondage du RP rénové
- II. Evolution des méthodes de calculs de précision du RP
- III. Résultats obtenus et diffusion de ces résultats
- IV. Cas particulier des DOM

# I. Rappel succinct sur le plan de sondage du RP

- On se restreint ici à la métropole et au champ des ménages
- Exhaustif dans les communes de moins de 10 000 habitants (PC) :
  - réparties en 5 groupes équilibrés
  - enquête exhaustive de chaque groupe tous les 5 ans
- Échantillonnage dans les communes de 10 000 habitants et plus (GC) :
  - répartition des adresses en 5 groupes équilibrés
  - grandes et nouvelles adresses enquêtées exhaustivement
  - chaque année, enquête par sondage des logements des « autres adresses » d'un groupe
    - ➔ au total, 40% des logements des grandes communes sont enquêtés en 5 ans

# Méthode d'estimation dans ce contexte

- Des enquêtes de recensement annuelles depuis 2004
- Chaque année depuis 2008 (pour le RP2006) : utilisation des 5 dernières années d'enquêtes consécutives pour :
  - connaître la population légale de toutes les communes pour l'année de référence = l'année médiane
  - diffuser des résultats détaillés sur les variables collectées
- Comment ?
  - en grandes communes : nombre moyen de personnes par logement sur 5 ans \* nombre de logements au 1<sup>er</sup> janvier de l'année médiane (« calage » au niveau iris)
  - en petites communes : interpolation ou extrapolation pour ramener toutes les petites communes à une même date de référence

# Conséquence en termes de précision

- **Le nouveau facteur en termes de précision provient** du sondage concernant la population des ménages vivant dans les « autres adresses » (ni grandes, ni nouvelles) dans les communes de plus de 10 000 habitants
- **Le reste de la population continue à être enquêté exhaustivement**
  - les petites communes enquêtées exhaustivement sur 5 ans
  - les communautés enquêtées exhaustivement sur 5 ans
  - les grandes et les nouvelles adresses des grandes communes enquêtées exhaustivement sur 5 ans
  - les autres populations (habitations mobiles et sans abris, marinières) collectées exhaustivement une fois tous les 5 ans

## II. Les estimations de précision du RP avant 2009

- **Reposaient sur les données du RP1999**  
dans le cadre des réflexions sur le RP rénové  
et en l'absence de données d'un cycle du RP rénové complet
- **Etaient basées sur des simulations**  
Plusieurs expériences en ce sens, variant notamment selon :
  - la méthodologie retenue (taille des grandes adresses, niveau du calage, etc.) pour mettre au point la définition de la strate des grandes adresses
  - les variables d'intérêt (population seule ou jeu de plusieurs variables collectées au RP) pour anticiper la précision de la diffusion
  - le nombre d'échantillons simulés
  - le niveau géographique retenu (plusieurs communes, quelques iris) pour anticiper la précision de la diffusion

## II. Les estimations de précision du RP avant 2009

- Parmi les résultats issus de ces simulations :  
les simulations de référence = rapport CNIS décembre 2005 :

La précision des résultats pour les communes de 10 000 habitants ou plus

Tranches d'effectif	Précision (CV)
50 000 ou plus	< 1,0 %
20 000 - 49 999	1,5%
10 000 - 19 999	2,0%
6 000 - 9 999	2,5%
3 000 - 5 999	3,0%
2 000 - 2 999	3,5%
1 000 - 1 999	4,5%
500 - 999	6,0%
250 - 499	8,0%
Moins de 250	> 8,0 %

Source : Rapport du CNIS « Utilisation des données produites par le recensement rénové de la population et leur diffusion » - Décembre 2005



# Un contexte différent à partir de 2009

## En 2009, le contexte évolue :

- données du premier cycle complet du RP rénové disponibles
- analyse des données à l'infra-communal dans la perspective de leur diffusion courant 2009 (données du RP2006)
- calcul de précision au niveau iris pour la métropole : possible grâce à un nouvel outil :
  - basé sur les formules statistiques sur les échantillons équilibrés issues des travaux de J.C. Deville et Y. Tillé (résidus)
  - tenant compte de l'équilibrage du tirage et du calage à l'iris
  - sans prise en compte de l'erreur de mesure temporelle (étalement sur 5 ans des collectes utilisées pour un millésime RP)

# Calculs avec ce nouvel outil depuis 2009

- **Ce nouvel outil permet :**
  - des calculs plus systématiques : à différents niveaux (commune, iris), sur toutes les variables
  - une confirmation / un affinement des résultats de précision donnés précédemment par les simulations
- **Il a été utilisé au département de la Démographie :**
  - fin 2009 : pour toutes les variables du RP2006 au niveau iris
  - début 2010 : pour la variable population du RP2006 au niveau région, département, communes
- **Outil complété en 2010 pour le projet DIAF (diffusion du RP sur des zones à façon)**
  - inclus un calage supplémentaire au niveau de la zone

### III. Les résultats obtenus sur les données du RP2006

- **Calculs (début 2010) des CV de la variable population (des ménages) aux niveaux national, régional et départemental :**

la population française est connue à + ou – 15 800 personnes  
( $15800 \times 2 / \text{pop française} = 0,05\%$ )

	Distribution du CV de la variable population au niveau région	Distribution du CV de la variable population au niveau département
75% Q3	0,20	0,53
50% Médiane	0,16	0,35
25% Q1	0,12	0,25

### III. Les résultats obtenus sur les données du RP2006

- **Calcul au niveau des communes de métropole :**

- pour 98% des GC au RP2006, la précision obtenue est meilleure que celle estimée pour le rapport CNIS de décembre 2005
- distribution :

	<b>Communes de 10000 à 19999 habitants</b>	<b>Communes de 20000 à 49999 habitants</b>	<b>Communes de 50000 à 100000 habitants</b>	<b>Communes de 100000 habitants et plus</b>
	(50% des communes)	(36% des communes)	(9% des communes)	(5% des communes)
<b>75% Q3</b>	1,16	0,87	0,71	0,43
<b>50% Médiane</b>	1,02	0,78	0,56	0,39
<b>25% Q1</b>	0,90	0,68	0,50	0,34

# Diffusion actuelle de ces données de précision

- **CV de la variable population du RP2006 de niveaux régional, départemental, communal :**
  - interne à l'Insee pour l'instant en 2010
  - information communiquée à la CNERP en 2011
  - note succincte bientôt en ligne sur insee.fr
- **CV de niveau iris calculés pour toutes les variables du RP2006 :**
  - CV de la variable population utilisés pour donner un label aux iris en diffusion, pour qualifier l'utilisation possible des données de niveau iris
  - CV non diffusés tels quels mais sous forme de CV « résumé » pour une meilleure robustesse des résultatssur insee.fr : <http://www.recensement.insee.fr/basesInfracommunales.action>

# Diffusion actuelle de ces données de précision

## Exemple des données de niveau iris diffusées :

Données infra-communales - Diplômes - Formation							
France - IRIS							
Découpage géographique au 01/01/2009							
© Insee		Source : Insee, Recensement de la population 2007 exploitation principale.					
IRIS	Région	Libellé commune ou ARM	Libellé de l'IRIS	Type d'IRIS	Label de l'IRIS en 2007	Pop scolarisée 2-5 ans en 2007 (princ)	Pop scolarisée 6-14 ans en 2007 (princ)
IRIS	REG	LIBCOM	LIBIRIS	TYP_IRIS	LAB_IRIS	P07_SCOL0205	P07_SCOL0614
010040101	82	Ambérieu-en-Bugey	Les Perouses-Triangle d'Activite	H	1	56	110
010040102	82	Ambérieu-en-Bugey	Longeray-Gare	H	1	117	393
010040201	82	Ambérieu-en-Bugey	Centre-St Germain-Vareilles	H	1	114	345
010040202	82	Ambérieu-en-Bugey	Tiret-Les Allymes	H	1	152	559

Données infra-communales - Diplômes - Formation							
Précision des variables							
Pour les IRIS appartenant à une commune de France métropolitaine de 10 000 habitants ou plus							
© Insee		Source : Insee, Recensement de la population 2007 exploitation principale.					
Variable	Libellé	Précision (coefficient de variation en %) selon la tranche d'effectif de la variable					
VAR_ID	VAR_LIB	VAR00_49	VAR50_99	VAR100_199	VAR200_299	VAR300_499	VAR500_699
P07_SCOL0205	Pop scolarisée 2-5 ans	> 18	18	13	8	< 8	so
P07_SCOL0614	Pop scolarisée 6-14 ans	> 18	18	15	11	9	7

## IV. Cas particulier des DOM

---

- **Plan de sondage différent :**
  - absence d'une base de sondage d'adresses → enquête niv. îlots
  - pas de traitement particulier des grandes et nouvelles adresses
  - pas de « calage » sur un nombre de logements au niveau iris
  - DOM hors champ des travaux du GT sur la validation des données infra-communales
  - Outil de calcul de CV DOM disponible après l'outil de métropole (mi-2010)
- **Premiers travaux de calcul de précision dans le cadre d'un mémoire FCDA (Sophie Planson)**
  - Calcul pour la variable population à l'été 2011
    - au niveau commune et iris → moins bons qu'en métropole au niveau iris
    - Calcul d'un label des iris des DOM proche de celui des iris de métropole, pour la base iris du RP2008 (octobre 2011)
    - Le calcul des CV de niveau iris pour toutes les variables reste à faire

# Conclusions à ce stade

---

- **Utilisation de ces résultats des calculs de précision du RP :**
  - en cas de contestation d'un maire, recours aux intervalles de confiance pour juger de la significativité des fluctuations de la population légale d'une commune d'une année sur l'autre
  - utilisation prudente en direction des maires
- **Ces résultats sur la précision du nouveau recensement sont à compléter dans les années à venir :**
  - les résultats de précision obtenus pour un millésime de populations légales ne sont pas forcément valables pour le suivant :
    - calculs à refaire (en théorie chaque année)
  - à termes, d'autres calculs de précision sont à entreprendre en ce qui concerne les évolutions issues des données du recensement
  - si possible, rédiger un document complet pour mettre à plat les hypothèses de calcul que font nécessairement de tels outils de calcul de variance liée au sondage



# Merci de votre attention !



Contact :  
Gwennaëlle Brilhaut  
Tél. : 01 41 17 53 97  
Courriel : gwennaelle.brilhaut@insee.fr

## Insee

18 bd Adolphe-Pinard  
75675 Paris Cedex 14

[www.insee.fr](http://www.insee.fr)

Informations statistiques :  
[www.insee.fr](http://www.insee.fr) / Contacter l'Insee  
09 72 72 4000  
(coût d'un appel local)  
du lundi au vendredi de 9h00 à 17h00