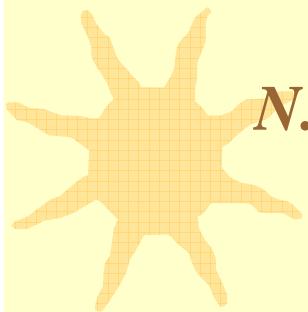
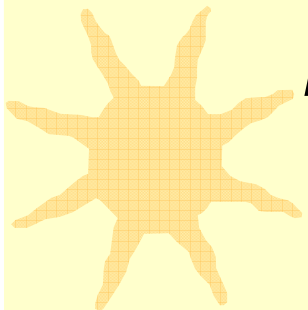


# *Redressement et calcul de précision de l'enquête Santé*



---

*N. Caron, Insee UMS puis Ministère de l'Education Nationale  
et S. Rousseau, Insee UMS puis Insee DSE*



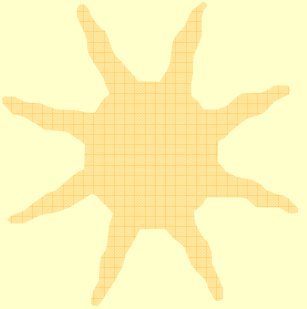
*Présentation aux Journées de Méthodologie Statistique*

*Mardi 24 mars 2009*



# *Sommaire*

---

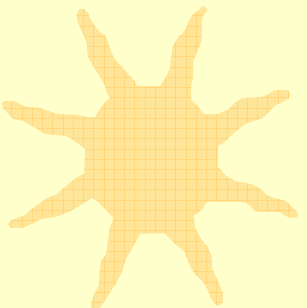
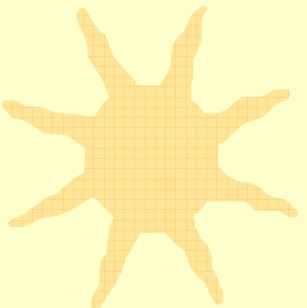


★ I. Caractéristiques de l'enquête Santé

★ II. Méthodologie adoptée

★ III. Quelques résultats

★ IV. Calcul de précision





# *I. Caractéristiques de l'enquête*

---



## ★ Objectif

- Informer sur l'état de santé des français, leur consommation de soins



## ★ Collecte

- Réalisée tous les 10 ans, dernière édition en 2002/2003
- Etalée sur un an (5 vagues de taille identique)
- 5 questionnaires posés en 3 visites successives :
  - 1 au niveau ménage (recueilli lors de la 1<sup>ère</sup> visite)
  - 3 au niveau individuel propres à chaque visite
  - 1 auto-questionnaire (adressé aux individus de 11 ans et plus ; distribué à la 1<sup>ère</sup> visite et repris ensuite)
- Seuls les répondants à une visite donnée sont réinterrogés à la suivante
- Au lieu de résidence principale à la date de l'enquête



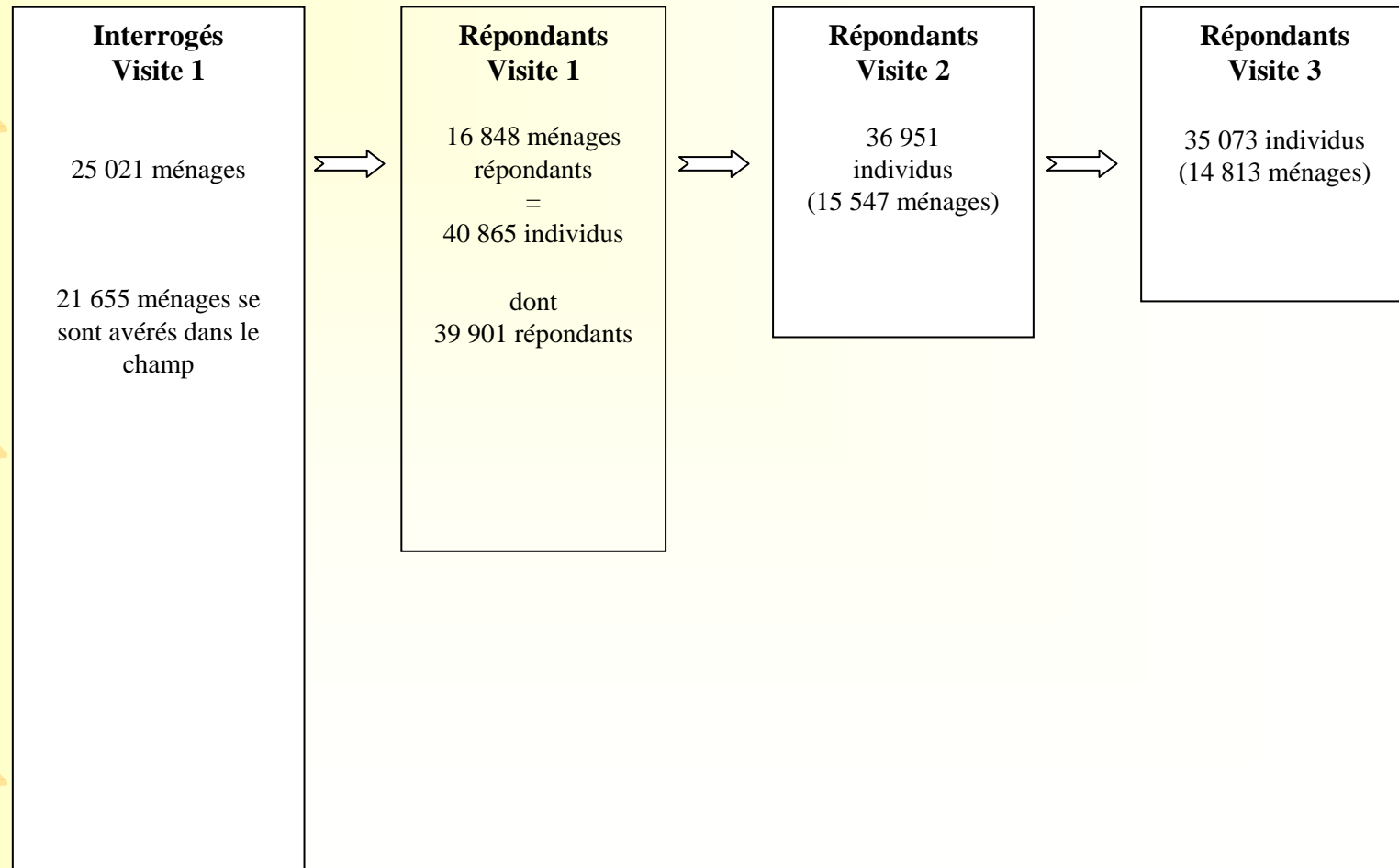
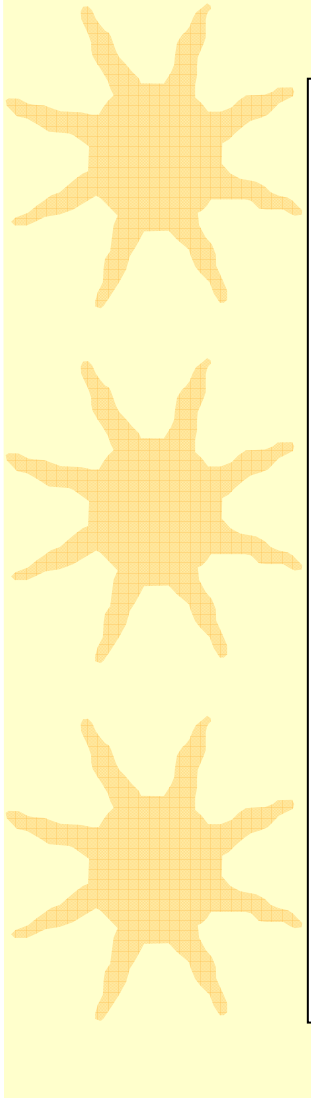
# I. Caractéristiques de l'enquête (suite)

## ★ Plan de sondage

- Sélection de 25 021 logements
- Incluant 5 extensions régionales en Ile-de-France, Picardie, Nord-Pas-de-Calais, Champagne-Ardenne, PACA
- Bases de sondage de logements :
  - Echantillons-Maîtres tirés du recensement de 1999 (EM+EMEX)
  - Base de sondage de logements neufs (BSLN)
- Tirage à plusieurs degrés, stratifié par catégorie de communes,
  - A probabilités égales pour les logements principaux (hors extensions)
  - Pondération des logements sélectionnés entre 352 et 3100
    - sous-représentation des logements non principaux en 1999
    - sur-représentation des logements des régions à extension



# I. Caractéristiques de l'enquête (suite)



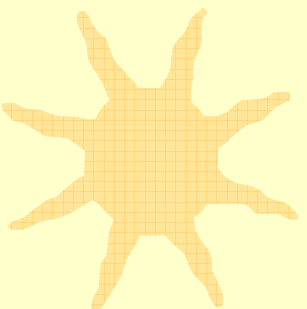
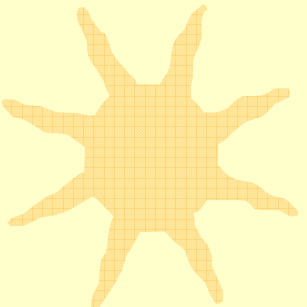
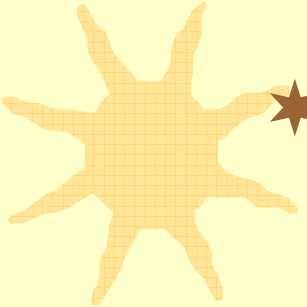


## *II. Méthodologie adoptée*

---

### ★ Objectifs d'exploitation

- Dresser un panorama de l'état de santé des individus
- Exploiter l'enquête au niveau régional dans les régions à extension et comparer les régions entre elles
- Comparer les résultats de l'auto-questionnaire avec les réponses aux 3 visites





## *II. Méthodologie adoptée*

---

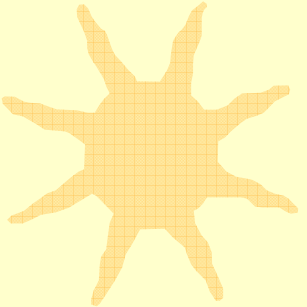
### ★ Choix retenus

- Optique individu  $\Rightarrow$  redressement niveau individu
- Prise en compte des extensions régionales dans le redressement
- Prise en compte de la saisonnalité de la collecte
- Choix de la simplicité d'exploitation : limiter à 3 le nombre de jeux de pondération (parmi 5 possibles)
  - la visite 1 à partir des individus répondants à la 1<sup>ère</sup> visite
  - l'ensemble des 3 visites à partir des individus ayant répondu aux 3 visites
  - l'auto-questionnaire



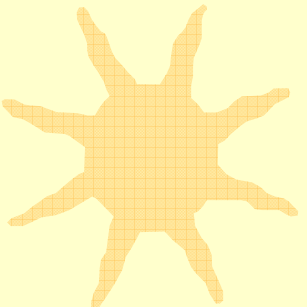
## *II. Méthodologie adoptée (suite)*

---



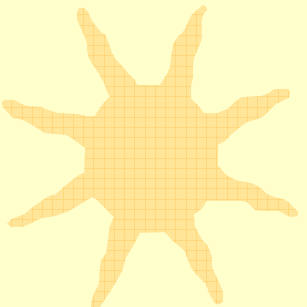
### ★ Cheminement

- 1) Correction de la non-réponse totale
- 2) Calage



### ★ Mise en œuvre

- 1) Au calcul des pondérations servant à exploiter les données de la 1<sup>ère</sup> visite
- 2) Reconduite pour la 3<sup>ème</sup> visite







## II. Méthodologie adoptée (suite)

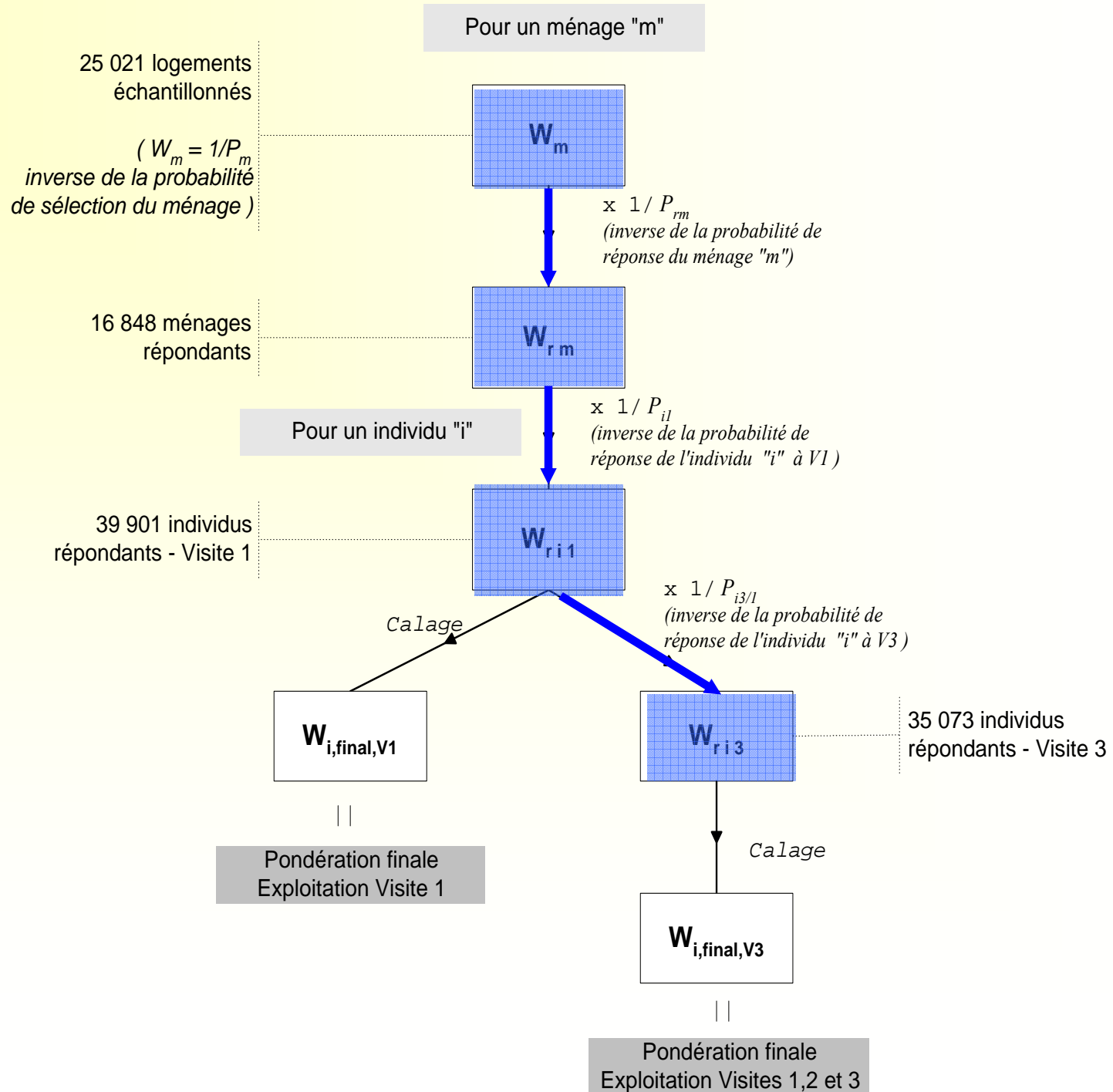
---

### ★ 1<sup>ère</sup> étape : Traiter la non-réponse totale

– Principe : correction par groupe de réponse homogène, déterminants du comportement de réponse obtenus par régression logistique

– Appliqué à 3 stades successifs :

- 1) Aux ménages enquêtés du champ (résidences principales à la date de l'enquête)
  - Recours aux variables « logement » de la base de sondage
- 2) Aux individus interrogés à la visite 1
  - Recours aux variables « individus » de la base de sondage
- 3) Aux individus interrogés à la visite 3
  - Utilisation des variables de la 1<sup>ère</sup> visite (notamment, l'état de santé des enquêtés)





## II. Méthodologie adoptée (suite)

★ 2<sup>ème</sup> étape : Caler au niveau individu sur des marges exogènes

– L'outil : la macro SAS CALMAR (Insee)

– La source : Enquête Emploi 2003

- 54 000 logements interrogés par trimestre avec un taux de rotation d'1/6, 80 000 ménages distincts par an
- L'enquête Emploi est spécifiquement calée sur la pyramide des âges de la population vivant en ménage ordinaire seulement, à l'exclusion des communautés
- Marges retenues : moyenne simple des résultats des 4 trimestres de 2003

– non théoriquement optimal au sens de la précision mais simple et permettant de conserver une cohérence de diffusion



## II. Méthodologie adoptée (suite)

### – Les variables de calage

- Les mêmes pour V1 et pour V3
- Mesurées aux niveaux ménage ou individu (calage simultané)

- Niveau individu : répartition par sexe\*âge, par groupe de région, par catégorie de communes, par niveau de diplôme
- Niveau ménage : répartition par groupe de région, par vague, par taille de ménage, par type de ménage.

- Pour une caractéristique mesurée au niveau ménage selon  $p$  modalités , on utilise pour le calage les variables individuelles :

$$\frac{I_{m \in X_1}}{T_m}, \dots, \frac{I_{m \in X_p}}{T_m}$$

- où  $T_m$  = taille réelle du ménage  $m$  et  $I(m \in X_l) = 1$  si le ménage  $m$  présente la modalité  $X_l$  et 0 sinon



### *III. Quelques résultats*

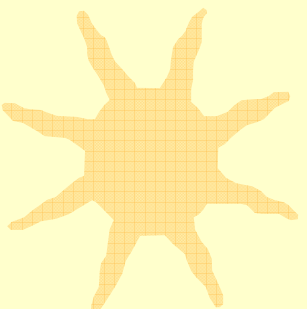
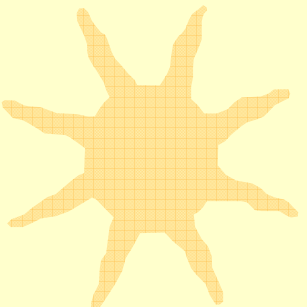
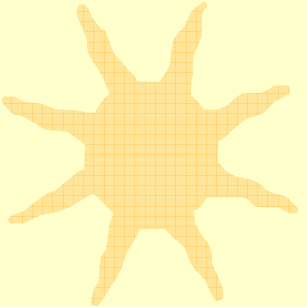
---

#### ★ Correction de la non-réponse à V1

- Ménage : 8 modélisations différentes
- Individus : 964 non répondants parmi la population de 40 865 individus des ménages répondants.

#### ★ Correction de la non-réponse à V3

- Individus : 6 modélisations différentes



21 655  
ménages  
du champ  
dont  
16 848  
répondants

	19 846 résidences principales au RP 99 et en 2003
7	1 241 logements secondaires, occasionnels ou vacants au RP99 devenus principaux en 2003
8	568 logements neufs

1	Nord Pas-de-Calais 1696 RP (210 NR)
2	Picardie 1287 RP (277 NR)
3	Champagne-Ardennes 1229 RP (263 NR)
4	Ile-de-France 4858 RP (1405NR)
5	PACA 2245 RP (617 NR)
6	Reste de la France 8531 RP (1553 NR)

vague  
strate de gestion  
habitat individuel ou collectif  
nombre de pièces  
taille du ménage  
âge de la personne de référence  
sexe de la personne de réf.  
département  
zeat

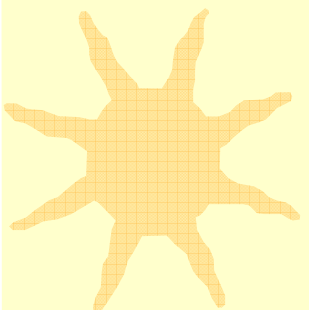
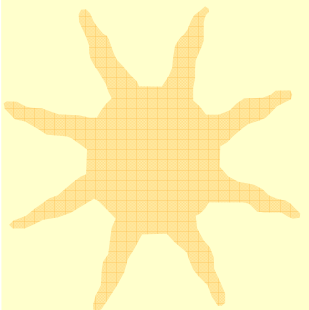
	362 NR
	120NR

vague  
strate de gestion  
zeat  
habitat individuel ou collectif

vague  
strate de gestion  
zeat



### III. Quelques résultats (suite)



39 901 individus  
dont  
35 073  
répondants



Vague de collecte  
Densité d'habitat (5 tranches)  
Logement collectif ou individuel  
Nombre de pièces du logement  
Taille du ménage  
Age de l'individu  
Sexe de l'individu  
Diplôme de l'individu  
Catégorie socio-professionnelle de l'individu  
Individu bénéficiant d'une couverture sociale ad-hoc pour raison de santé  
Etat de santé de l'individu

Nord-Pas-de-Calais	Picardie	Champagne-Ardennes	Ile-de-France	PACA	Reste de la France
(3 922 ; 3 492)	(2 680 ; 2 407)	(2 488 ; 2 274)	(9 160 ; 7 844)	(4 025 ; 3 262)	(17 626 ; 15 794)
1	2	3	4	5	6



### *III. Quelques résultats (suite)*

#### ★ Distributions des rapports des poids

	Calage V1 /poids initial	Calage V3 / poids initial
Maximum	2.79	3.64
Médiane	1.29	1.45
Minimum	0.82	0.87

#### ★ Distributions des poids finaux

	Pondération initiale	Pondération après calage V1	Pondération après calage V3
Maximum	3100.58	6116.12	8933.34
Médiane	857.94	1374.57	1584.81
Minimum	352.10	338.24	360.23

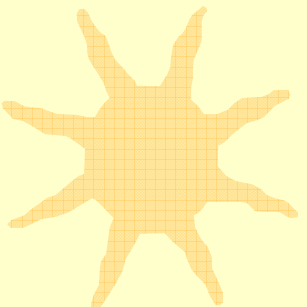
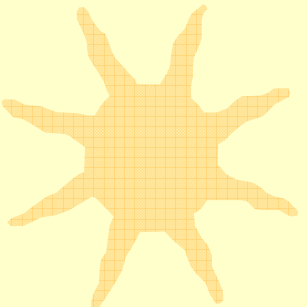
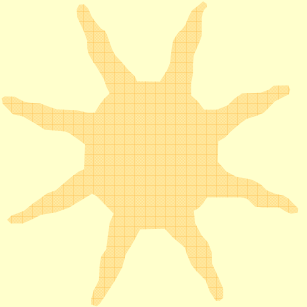




## *IV. Estimation de la précision*

---

- ★ But : fournir des intervalles de confiance pour des estimateurs calculés à partir de l'enquête
- ★ La précision mesurée
  - Appréhende
    - l'aléa d'échantillonnage (modélisation du plan de sondage stratifié à plusieurs degrés)
    - la correction de la non-réponse totale
    - le calage
  - Exclut les effets
    - des erreurs de mesure
    - de la non-réponse partielle
- ★ L'outil : la macro SAS POULPE (Insee)





## *IV. Estimation de la précision (suite)*

### ★ Quelques résultats

#### – Apport des extensions régionales

*Exemple : proportion de personnes se sentant en bonne santé*

<i>Région</i>	<i>Nombre de répondants 1<sup>ère</sup> visite</i>	<i>Proportion estimée</i>	<i>Intervalle de confiance à 95% 1<sup>ère</sup> visite</i>	<i>Coefficient de variation (%)</i>	<i>Effet de Effet de sondage en sondage l'absence d'extension</i>	
France	39 901	0,685	[0,68 - 0,69]	0,4		
Ile-de-France	9 160	0,725	[0,71 - 0,74]	0,8	1,03	1,21
Provence-Alpes-Côte-d'Azur	4 025	0,650	[0,63 - 0,67]	1,4	0,91	1,23
Nord-Pas-de-Calais	3 922	0,660	[0,64 - 0,68]	1,5	0,94	1,16
Rhône-Alpes	2 536	0,700	[0,68 - 0,72]	1,6		
Champagne-Ardennes	2 488	0,695	[0,67 - 0,72]	1,7	0,51	0,77
Picardie	2 680	0,690	[0,66 - 0,72]	2,3	1,14	1,55
Auvergne	543	0,630	[0,58 - 0,69]	4,5		



## *IV. Estimation de la précision (suite)*

### – Apport du calage

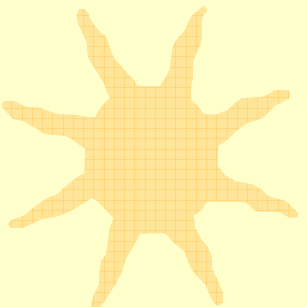
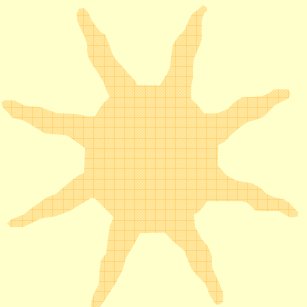
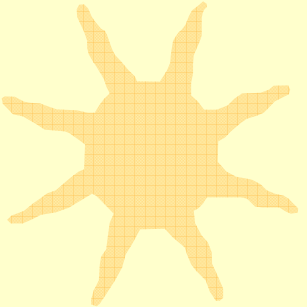
*Exemple : coefficient de variation de l'estimateur du nombre de personnes se sentant en bonne santé (%)*

<i>Région</i>	<i>Avec toutes les variables de calage</i>	<i>Sans les variables régionales de calage</i>	<i>Sans calage</i>
France	0,4	0,4	0,8
Ile-de-France	0,9	1,0	1,4



## *En conclusion*

---



- ★ Méthodologie générale : “Bonnes pratiques” recommandées et applicables pour toutes sortes d’enquêtes
  - d’abord, connaître les objectifs d’exploitation
  - prendre en compte les spécificités de l’enquête, comme ici :
    - utiliser les réponses à l’enquête pour expliquer la non-réponse en cas d’interrogations successives d’un même enquêté
    - caler sur des dimensions régionales
    - Utilisation de la variable « Vague » comme variable de calage
  
- ★ Intérêt des extensions régionales et d’un processus harmonisé
  - Répondre à des besoins locaux
  - Enrichir les analyses nationales
  - D’autres usages de l’EMEX depuis Santé : Logement 07, IVQ 05



# Bibliographie

- ★ Caron N., Rousseau S. (2005), "Redressement et calage de l'enquête Santé 2003", *Document de travail n°M0501 de la série "Méthodologie Statistique"*, INSEE (sur le site [www.insee.fr](http://www.insee.fr))
- ★ Caron N. (2005), "La correction de la non-réponse par repondération et par imputation", *Document de travail n°M0502 de la série « Méthodologie Statistique »*, INSEE (sur le site [www.insee.fr](http://www.insee.fr))
- ★ Deville J.-C., Särndal C.-E., Sautory O. (1993), "Generalized raking procedures in survey sampling", *JASA*, vol.88, n°423
- ★ Deville J.-C., Särndal C.-E. (1992), "Calibration estimators in survey sampling", *JASA*, Vol. 87, pp. 376-382
- ★ Rousseau S. (2006), "Accuracy of regional estimators", *European Conference on Quality in Survey Statistics*

