

Calcul de précision pour une région ayant procédé à une extension régionale et locale de l'Enquête Logement 2006

Guillaume Chauvet

Direction Régionale de RENNES - ENSAI

Journées de Méthodologie Statistique

Paris

25/03/2009

Principaux éléments

L'Enquête Logement 2006 a donné lieu à une extension régionale et à plusieurs extensions locales au niveau de la région Bretagne. Un complément d'échantillon a également été sélectionné dans des bases externes.

Principaux éléments

L'Enquête Logement 2006 a donné lieu à une extension régionale et à plusieurs extensions locales au niveau de la région Bretagne. Un complément d'échantillon a également été sélectionné dans des bases externes.

Un plan de sondage et une technique d'estimation complexes ont été nécessaires pour la mise en commun et l'exploitation des différents sous-échantillons.

Principaux éléments

L'Enquête Logement 2006 a donné lieu à une extension régionale et à plusieurs extensions locales au niveau de la région Bretagne. Un complément d'échantillon a également été sélectionné dans des bases externes.

Un plan de sondage et une technique d'estimation complexes ont été nécessaires pour la mise en commun et l'exploitation des différents sous-échantillons.

Un outil SAS de calcul de précision basé sur des formules de variance estimée a été mis au point pour les partenaires de l'Enquête et pour les chargés d'étude de la DR.

- 1 Présentation de l'Enquête
- 2 Estimation de précision
- 3 Un exemple d'utilisation de la macro

Présentation de l'Enquête

Présentation

L'Enquête Logement est une des plus grosses enquêtes réalisées par l'Insee auprès des ménages. Elle a lieu environ tous les quatre ans (dernières éditions en 2002 et 2006).

Présentation

L'Enquête Logement est une des plus grosses enquêtes réalisées par l'Insee auprès des ménages. Elle a lieu environ tous les quatre ans (dernières éditions en 2002 et 2006).

L'objectif de l'enquête est double:

- Connaître le parc de logements (ancienneté de la construction, nombre de maisons individuelles/appartements, nombre de propriétaires/locataires,...)
- Décrire les conditions de vie des ménages (mobilités et causes de mobilité, confort du logement, emprunts,...)

Sélection de l'échantillon

Dans l'Enquête Logement, le champ de l'enquête est celui des logements résidences principales en 2006, que l'on enquête à l'aide du RP99 et de la BSLN.

Sélection de l'échantillon

Dans l'Enquête Logement, le champ de l'enquête est celui des logements résidences principales en 2006, que l'on enquête à l'aide du RP99 et de la BSLN.

Pour une région avec extension régionale et locale telle que la Bretagne, la sélection de l'échantillon se fait en trois temps (Le Guennec, 2009) :

Sélection de l'échantillon

Dans l'Enquête Logement, le champ de l'enquête est celui des logements résidences principales en 2006, que l'on enquête à l'aide du RP99 et de la BSLN.

Pour une région avec extension régionale et locale telle que la Bretagne, la sélection de l'échantillon se fait en trois temps (Le Guennec, 2009) :

- Sélection de l'échantillon national dans l'Echantillon Maître de 99 (RP99, BSLN) et de l'extension régionale dans l'EMEX,

Sélection de l'échantillon

Dans l'Enquête Logement, le champ de l'enquête est celui des logements résidences principales en 2006, que l'on enquête à l'aide du RP99 et de la BSLN.

Pour une région avec extension régionale et locale telle que la Bretagne, la sélection de l'échantillon se fait en trois temps (Le Guennec, 2009) :

- Sélection de l'échantillon national dans l'Echantillon Maître de 99 (RP99, BSLN) et de l'extension régionale dans l'EMEX,
- Sélection d'extensions d'échantillon au niveau des 6 principales aires urbaines,

Sélection de l'échantillon

Dans l'Enquête Logement, le champ de l'enquête est celui des logements résidences principales en 2006, que l'on enquête à l'aide du RP99 et de la BSLN.

Pour une région avec extension régionale et locale telle que la Bretagne, la sélection de l'échantillon se fait en trois temps (Le Guennec, 2009) :

- Sélection de l'échantillon national dans l'Echantillon Maître de 99 (RP99, BSLN) et de l'extension régionale dans l'EMEX,
- Sélection d'extensions d'échantillon au niveau des 6 principales aires urbaines,
- Sélection d'échantillons complémentaires dans des bases externes (base ZUS, fichiers d'allocataires).

Schéma du tirage

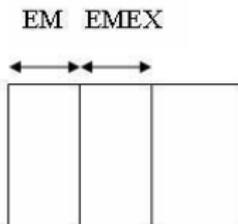


Schéma du tirage

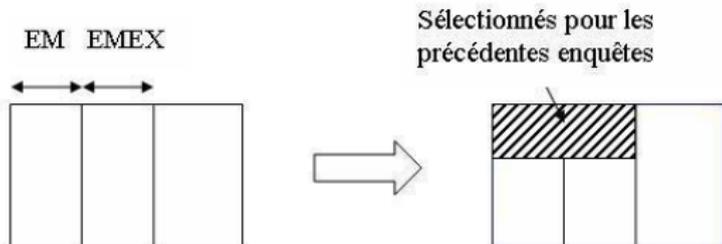


Schéma du tirage

EM EMEX

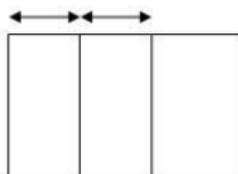
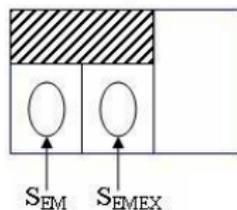
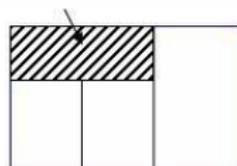
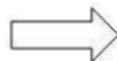
Sélectionnés pour les
précédentes enquêtes

Schéma du tirage

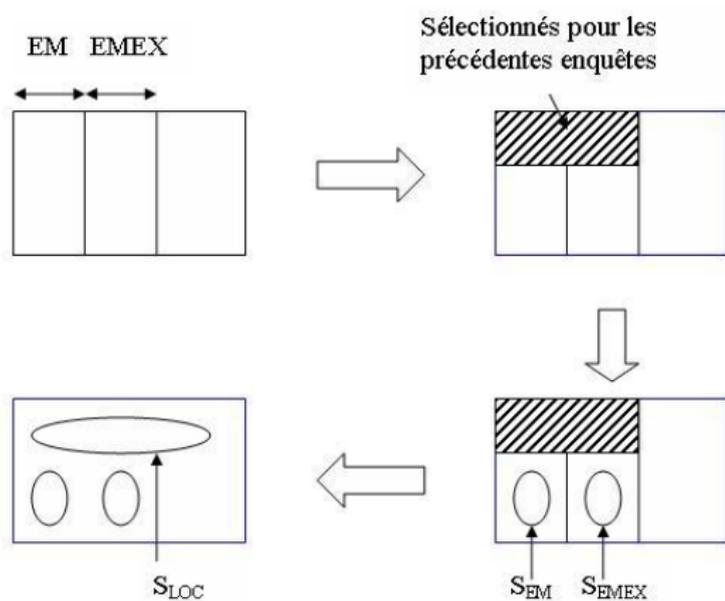


Schéma du tirage

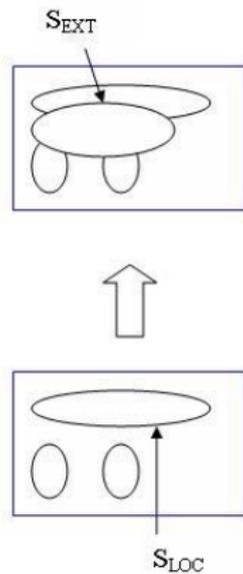
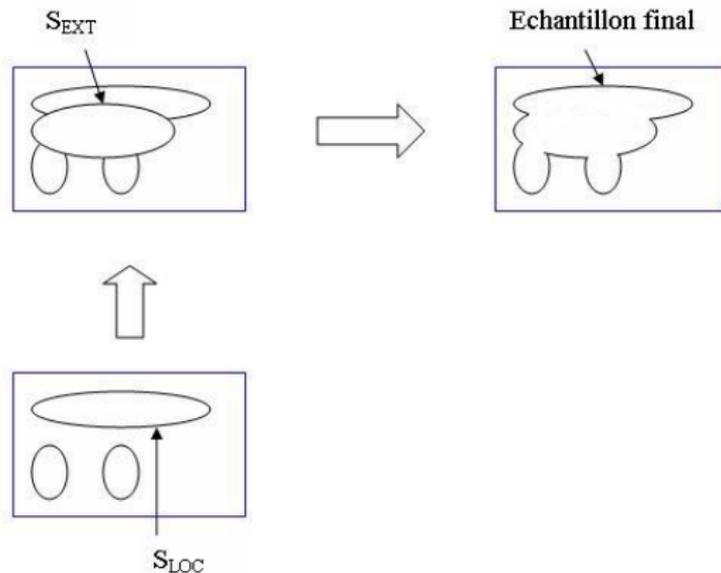


Schéma du tirage



Estimation de précision

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

- 1 Sélection des sous-échantillons,

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

- 1 Sélection des sous-échantillons,
- 2 Mise en commun des sous-échantillons,

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

- 1 Sélection des sous-échantillons,
- 2 Mise en commun des sous-échantillons,
- 3 Redressement de la non-réponse partielle,

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

- 1 Sélection des sous-échantillons,
- 2 Mise en commun des sous-échantillons,
- 3 Redressement de la non-réponse partielle,
- 4 Redressement de la non-réponse totale (méthode des groupes homogènes de réponse),

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

- 1 Sélection des sous-échantillons,
- 2 Mise en commun des sous-échantillons,
- 3 Redressement de la non-réponse partielle,
- 4 Redressement de la non-réponse totale (méthode des groupes homogènes de réponse),
- 5 Calage sur une information externe.

Les étapes de l'enquête

Les différentes étapes de traitement de l'enquête (Le Guennec, 2008) sont :

- 1 Sélection des sous-échantillons,
- 2 Mise en commun des sous-échantillons,
- 3 Redressement de la non-réponse partielle,
- 4 Redressement de la non-réponse totale (méthode des groupes homogènes de réponse),
- 5 Calage sur une information externe.

Toutes ces étapes doivent normalement être prises en compte dans le calcul de précision.

Les étapes de l'enquête

En particulier, la non-réponse partielle a des conséquences en termes de biais et de variance (Caron, 2005), mais cette variance supplémentaire n'est pas prise en compte dans le calcul :

Les étapes de l'enquête

En particulier, la non-réponse partielle a des conséquences en termes de biais et de variance (Caron, 2005), mais cette variance supplémentaire n'est pas prise en compte dans le calcul :

- nécessite d'identifier dans l'échantillon les valeurs imputées,

Les étapes de l'enquête

En particulier, la non-réponse partielle a des conséquences en termes de biais et de variance (Caron, 2005), mais cette variance supplémentaire n'est pas prise en compte dans le calcul :

- nécessite d'identifier dans l'échantillon les valeurs imputées,
- nécessite autant d'estimateurs de variance que de techniques d'imputation utilisées,

Les étapes de l'enquête

En particulier, la non-réponse partielle a des conséquences en termes de biais et de variance (Caron, 2005), mais cette variance supplémentaire n'est pas prise en compte dans le calcul :

- nécessite d'identifier dans l'échantillon les valeurs imputées,
- nécessite autant d'estimateurs de variance que de techniques d'imputation utilisées,
- pose problème pour des statistiques complexes (même des ratios).

Les étapes de l'enquête

En particulier, la non-réponse partielle a des conséquences en termes de biais et de variance (Caron, 2005), mais cette variance supplémentaire n'est pas prise en compte dans le calcul :

- nécessite d'identifier dans l'échantillon les valeurs imputées,
- nécessite autant d'estimateurs de variance que de techniques d'imputation utilisées,
- pose problème pour des statistiques complexes (même des ratios).

On notera respectivement d_k et w_k le poids de sondage et le poids après calage pour une unité k , \hat{t}_{yd} et \hat{t}_{yw} les estimateurs d'un total t_y associés.

Calcul de précision

$$V(\hat{t}_{yw})$$

Prise en compte du calage

$$V(\hat{t}_{yw}) \simeq V(\hat{t}_{ed})$$

Utilisation de la variable de résidus de régression de la variable d'intérêt sur les variables de calage $e_k = y_k - \mathbf{x}'_k \beta$.

Prise en compte de la non-réponse totale

$$\begin{aligned} V(\hat{t}_{yw}) &\simeq V(\hat{t}_{ed}) \\ &= V_p(\hat{t}_{ed}) + V_{NR}(\hat{t}_{ed}) \end{aligned}$$

Variance associée à l'échantillonnage

$$V_p(\hat{t}_{ed})$$

Mise en commun des sous-échantillons

$$\begin{aligned} & V_p(\hat{t}_{ed}) \\ = & V_p(\hat{t}_{ed}^N + \hat{t}_{ed}^L + \hat{t}_{ed}^Z) \end{aligned}$$

Utilisation de la variable synthétique $\tilde{e}_k = f(e_k)$, associée à la technique d'estimation composite permettant de mettre en commun les sous-échantillons.

Prise en compte du plan de sondage

$$\begin{aligned} & V_p(\hat{t}_{ed}) \\ &= V_p(\hat{t}_{ed}^N + \hat{t}_{ed}^L + \hat{t}_{ed}^Z) \\ &= V_p(\hat{t}_{ed}^N) + V_p(\hat{t}_{ed}^L) + V_p(\hat{t}_{ed}^Z) \end{aligned}$$

Utilisation de

- l'indépendance entre échantillon externe et le reste,
- l'indépendance (supposée) entre les extensions locales et l'échantillon national + régional.

Un exemple d'utilisation de la macro

On souhaite estimer, sur l'ensemble des résidences principales de l'aire urbaine de Rennes :

- La structure des logements selon le nombre de chambres,
- Le nombre total de chambres,
- La surface moyenne par habitant.

On est dans le cas d'une estimation sur un **domaine** D , c'est à dire sur une sous-population de U . Cette estimation ne pose pas de problème particulier, en remarquant que

$$t_{yD} = \sum_{k \in D} y_k = \sum_{k \in U} y_k 1_{k \in D}$$

où $1_{k \in D}$ vaut 1 si le logement k est dans le domaine, et 0 sinon.

On va donc estimer :

- des effectifs (nombre de logements ne comptant aucune chambre, comptant 1 chambre, ...)
- un total (nombre de chambres),
- un ratio (surface totale rapportée au nombre d'habitants).

Dans les deux premiers cas, on estime un total (variable indicatrice pour un effectif, variable quantitative pour un total). Dans le second cas, on estime un ratio de totaux : l'estimation de variance se fait par linéarisation (Deville, 1999).

Résultats obtenus

Paramètre	Estim.	Var.	CV (%)	BI (95%)	BS (95%)	DEFF	DCAL	NR (%)
Nb cha.	578 349	$5,6 \cdot 10^7$	1,23	$5,64 \cdot 10^5$	$5,92 \cdot 10^5$	0,11	0,24	22,07
Surf. moy.	38,27	0,23	1,25	37,34	39,21	0,48	0,42	21,79
% Log.								
0 cha.	0,08	$4,3 \cdot 10^{-5}$	7,75	0,07	0,10	0,46	0,99	17,34
1 cha.	0,18	$1,1 \cdot 10^{-4}$	5,69	0,16	0,20	0,59	0,90	21,80
2 cha.	0,26	$2,1 \cdot 10^{-4}$	5,68	0,23	0,29	0,91	0,96	19,15
3 cha.	0,26	$3,0 \cdot 10^{-4}$	6,64	0,23	0,29	1,26	0,98	18,06
4 cha.	0,18	$1,7 \cdot 10^{-4}$	7,43	0,15	0,20	0,97	0,80	18,82
5 cha.	0,04	$8,5 \cdot 10^{-5}$	23,6	0,02	0,06	1,86	0,97	19,96
6 cha.	$3 \cdot 10^{-3}$	$1,7 \cdot 10^{-6}$	48,5	10^{-4}	$5,2 \cdot 10^{-3}$	0,52	1,01	17,36
+ 6 cha.	$4 \cdot 10^{-4}$	$3,3 \cdot 10^{-7}$	152	-10^{-3}	$1,5 \cdot 10^{-3}$	0,72	1,01	17,21

Résultats obtenus

Si on se fixe (par exemple) un seuil à 10% pour le coefficient de variation, la précision est insuffisante pour les logements comportant un grand nombre de chambres (peu de logements correspondant \Rightarrow imprécision importante).

On va donc réaliser un regroupement pour définir une variable à 5 modalités :

- Logements ne comportant pas de chambre,
- Logements comportant 1 chambre,
- Logements comportant 2 chambres,
- Logements comportant 3 chambres,
- Logements comportant 4 chambres et plus.

Résultats obtenus (suite)

Paramètre	Estim.	Var.	CV (%)	BI (95%)	BS (95%)	DEFF	DCAL	NR (%)
Nb cha.	578 349	$5,6 \cdot 10^7$	1,23	$5,64 \cdot 10^5$	$5,92 \cdot 10^5$	0,11	0,24	22,07
Surf. moy.	38,27	0,23	1,25	37,34	39,21	0,48	0,42	21,79
% Log.								
0 cha.	0,08	$4,3 \cdot 10^{-5}$	7,75	0,07	0,10	0,46	0,99	17,34
1 cha.	0,18	$1,1 \cdot 10^{-4}$	5,69	0,16	0,20	0,59	0,90	21,80
2 cha.	0,26	$2,1 \cdot 10^{-4}$	5,68	0,23	0,29	0,91	0,96	19,15
3 cha.	0,26	$3,0 \cdot 10^{-4}$	6,64	0,23	0,29	1,26	0,98	18,06
+ 4 cha.	0,22	$2,0 \cdot 10^{-4}$	6,49	0,19	0,25	0,97	0,81	19,94

Bibliographie

Caron, N. (2005), *La correction de la non-réponse par repondération et par imputation*, document de travail UMS n° M0502, Insee, Paris.

Chauvet, G. (2008), *Calcul de précision pour une région ayant procédé à des extensions régionale et locale de l'Enquête logement 2006*, document de travail, SED, Insee, Rennes.

Deville, J-C. (1999), *Variance estimation for complex statistics and estimators : linearization and residual techniques*, Survey Methodology, 25, P. 193-204.

Le Guennec, J. (2008), *Note de bilan de l'Enquête Logement*, document de travail, Pôle ISM, Insee, Rennes.

Le Guennec, J. (2009), *Enquête Logement 2006, des extensions d'échantillon pour les estimations locales : échantillonnage et redressement*, Journées de Méthodologie Statistique, Paris.