

Sources et méthodes

Le dispositif d'enquêtes permanentes des Conditions de vie (EPCV)

Le dispositif d'enquêtes permanentes des Conditions de vie (EPCV) a plusieurs objectifs. Il permet de produire tous les ans des indicateurs sociaux non monétaires et qu'il est possible de relier entre eux. De part son mode de collecte (trois enquêtes par an : une en janvier, une en mai, la dernière en octobre), instauré depuis 1996, elle répond à des besoins en termes d'information statistique sur les conditions de vie et approfondie, selon l'enquête, les thèmes suivants : qualité de l'habitat et de l'environnement, santé, logement et situation financière des ménages et enfin la participation (à la vie associative, électorale, etc.) et les contacts sociaux.

Présentation du dispositif EPCV

Le dispositif d'enquêtes permanentes des Conditions de vie (EPCV) a donné lieu à trois enquêtes par an (une en janvier, une en mai, la dernière en octobre) depuis 1996 et ce jusqu'en 2004. Il comprend trois parties (tableau de composition du ménage, indicateurs sociaux, partie variable).

Son objectif général est :

- d'une part d'obtenir des indicateurs sociaux (non monétaires et que l'on puisse relier entre eux) avec une périodicité annuelle ;

- d'autre part d'approfondir certains thèmes, pour répondre en premier lieu à des besoins d'information statistique, dans le champ des conditions de vie. Les parties variables peuvent être ponctuelles, visant à explorer un domaine nouveau (par exemple, services de proximité, « bonheur et travail », vie de quartier, comportements vis à vis de la santé, vie associative) ; ou répétées, et reprendre à un rythme régulier (généralement 4 ou 5 ans) des thématiques dont on souhaite suivre l'évolution (vacances, environnement, nouvelles technologies...). La plupart des enquêtes sont réalisées en partenariat avec d'autres organismes du système statistique public.

Dans sa composante « indicateurs sociaux », chaque enquête (janvier, mai, octobre) poursuit des objectifs qui lui sont propres.

Enquête indicateurs sociaux de janvier (qualité de l'habitat et de l'environnement)

L'enquête « qualité de l'habitat et de l'environnement » a pour but d'analyser l'environnement du [logement](#) des [ménages](#), du point de vue des équipements disponibles, de la qualité du cadre de vie et de l'insécurité.

Un premier objectif est de mesurer la sensibilité des ménages aux nuisances liées aux bruits et celles liées à la pollution. Le questionnaire permet d'évaluer le nombre de ménages concernés par ces nuisances ainsi que l'intensité ressentie et les causes de ces nuisances.

Un autre objectif est la mesure de la qualité de l'environnement social du logement. Il est tout d'abord appréhendé par le biais des services et équipements disponibles à proximité de l'habitat. Le bureau de poste, l'école, le cinéma, le restaurant, le supermarché, la boulangerie et le café sont considérés comme des services et équipements représentatifs de la qualité de cet environnement. Le questionnaire permet aussi d'évaluer la qualité des transports en commun et leur utilisation. Au total, il devrait être possible de repérer les manques et imperfections des équipements de proximité.

L'enquête porte également sur les problèmes de sécurité personnelle. Globalement, cette partie du questionnaire a pour objectif de mesurer un large éventail d'agressions ou de violences subies par les membres du ménage et d'analyser les démarches entreprises suite à ces événements. Des éléments sont également collectés sur les conséquences pour les victimes, les changements de comportements et les précautions particulières prises par la suite.

Enquête indicateurs sociaux de mai (Santé, logement et situation financière des ménages)

L'enquête de mai intitulée « Santé, logement et situation financière des ménages » a pour objectif de collecter des indicateurs sur le confort du logement, les équipements en biens durables, les difficultés financières rencontrées pour les ménages, et les problèmes de santé des individus.

Les questions sur le confort du logement et les biens d'équipement se présentent sous forme de listes d'une vingtaine d'éléments plus ou moins répandus dans les ménages, et de quelques questions subjectives sur les critiques qu'on peut apporter au logement.

Des questions sur la variation du revenu sur les douze derniers mois ([revenu imposable](#)), la satisfaction du ménage quant à son [niveau de vie](#) apportent une approche plus subjective sur la situation financière des ménages. Le questionnaire donne aussi une indication sur le niveau de l'épargne des ménages en interrogeant sur la variation des placements financiers au cours des douze derniers mois et le montant global de l'épargne, relativement au revenu.

Les difficultés financières sont abordées en premier lieu sous l'angle de l'endettement. Le questionnaire permet de connaître les raisons du recours aux crédits ou à l'emprunt, donne une évaluation du poids des remboursements sur le revenu du ménage c'est-à-dire le taux d'effort réalisé par les ménages, et interroge sur les découverts bancaires et les difficultés pour payer des traites ou certaines factures. Dans le registre « pauvreté en condition de vie », des indicateurs de « déprivation » ont été introduits en 1997 dans le questionnaire : on cherche à repérer les ménages qui ne peuvent pas, pour des raisons financières, financer certaines dépenses généralement jugées aujourd'hui nécessaires, comme chauffer correctement leur logement, acheter des vêtements neufs, recevoir des parents ou des amis, partir en vacances ..., ce qui permet de mesurer les conséquences de la situation financière sur les conditions de vie.

La santé est appréhendée tout d'abord à l'aide d'une auto-évaluation par les individus sur leur état de santé, puis de critères plus « objectifs » : le nombre et la longueur des congés maladie, le recours à des consultations médicales. En complément on interroge sur l'état dentaire, la vision, ainsi que la consommation de médicaments et celle de cigarettes. On connaît également le nombre des recours à l'aide d'une personne et la proportion des ménages bénéficiant d'une aide financière liée à la santé.

Enquête indicateurs sociaux d'octobre (Participation et contacts sociaux)

L'enquête « Participation et contacts sociaux » a pour objectif d'appréhender différents aspects des conditions de travail, de l'effort de formation et de la participation (associative, électorale) des individus ainsi que les différentes dimensions de leur mobilité sociale. Comme pour les enquêtes de janvier et de mai, l'objectif est d'être le plus complémentaire possible aux enquêtes existant déjà, spécialisées dans les thèmes abordés, en l'occurrence l'[enquête Emploi](#), les enquêtes Conditions de travail, ainsi que des enquêtes sur l'éducation, les contacts sociaux, les pratiques culturelles...

L'enquête permet tout d'abord de faire un point régulier sur les contraintes matérielles (horaires, pénibilité du travail, montant et flexibilité des rémunérations) et organisationnelles (dépendance hiérarchique, usage de nouvelles technologies) dans le travail, et pose aux actifs des questions sur la façon dont ils sont prêts à échanger une modification de leur durée de travail contre des modifications (plus ou moins) compensatrices de leurs conditions de rémunération.

S'agissant de l'effort de formation, on essaie d'en prendre la mesure pour ceux qui ont un emploi (formation continue) et pour les [chômeurs](#). Le questionnaire cherche à repérer les critères régissant l'octroi des promotions et de formation dans les entreprises (critères dont la connaissance est essentielle pour juger de l'effet de la formation), ainsi que le rapport entre le niveau de formation des salariés et la fonction qu'ils occupent.

Le questionnaire cherche à appréhender l'intensité des relations hors travail avec les collègues de travail. L'enquête, par ailleurs, reconstitue l'entourage familial (encore vivant) des individus et pose des questions sur la fréquence des contacts avec cet entourage. L'ensemble de ces questions permet de cerner l'isolement social. Le questionnaire interroge sur des problèmes familiaux (décès, séparation) rencontrés pendant la scolarité des individus dans l'objectif d'analyser l'effet de ces éléments de biographie sur la carrière sociale.

L'enquête fournit également des indicateurs sur les principaux aspects de la participation sociale : participation aux principales élections, participation à la vie associative, fréquentation des différentes activités culturelles (cinéma, spectacle sportif et musées ...), pratiques amateurs.

La seconde partie, dite « partie variable », est une enquête consacrée à un aspect particulier des conditions de vie des ménages. Le thème de cette partie est variable d'une enquête à l'autre.

Les deux parties (variable et fixe) s'articulent autour d'une partie commune, par ailleurs commune à toutes les enquêtes réalisées par l'Insee auprès des ménages : le tableau de composition des ménages (TCM) décrivant les caractéristiques socio-démographique des individus du ménage (sexe, âge, niveau de diplôme, activité, profession, nationalité, type de commune de résidence), et un court questionnaire sur les revenus du ménage.

Le champ de l'enquête

Le champ de ces enquêtes concerne les ménages vivant en [France métropolitaine](#). Au sein de ces ménages, on interroge les personnes de 15 ans ou plus.

En termes de représentativité géographique, les enquêtes PCV ont vocation à donner des résultats au niveau national. En données annuelles, elles ne peuvent donc fournir des résultats fiables au niveau régional, sauf lorsqu'une extension régionale est réalisée. Toutefois, quelques résultats relativement agrégés peuvent être fournis au niveau [ZEAT](#) (zones d'études pour l'aménagement du territoire, regroupement de régions).

Cependant, l'empilement des données sur plusieurs années permet de descendre à un niveau géographique plus fin.

Le mode de collecte

Principes de recueil des données

Les enquêtes EPCV sont réalisées du 1er au 31 du mois de l'enquête (janvier, mai, octobre) au moyen d'un questionnaire (Indicateurs sociaux d'octobre) administré par un enquêteur en face à face sous un mode de collecte assistée par informatique (CAPI).

Exhaustivité ou échantillonnage

L'échantillon des enquêtes EPCV est constitué de 8 800 ménages (5 600 répondants en moyenne), et de un (octobre) à trois (en mai et janvier) individu(s) tiré(s) au sort par ménage répondant au questionnaire, soit, en janvier, environ 11 000 individus répondants.

Redressements

Comme pour toutes les enquêtes, des redressements sont nécessaires pour compenser les biais introduits par les non-réponses et réduire les fluctuations d'échantillonnage.

Tout d'abord, des pondérations sont calculées ex post afin d'assurer une représentativité de l'échantillon par rapport à la population présente sur le territoire. Elles permettent à la fois de limiter les fluctuations d'échantillonnage liées au fait que l'enquête est réalisée par sondage et de redresser la non-réponse à l'enquête. La population de référence est celle de l'[enquête Emploi](#) de l'année précédente (moyenne des quatre enquêtes trimestrielles de l'année précédente).

La non-réponse totale est ensuite corrigée par une méthode usuelle de « calage sur marge » consistant à modifier les poids des ménages répondants de telle manière que le total de certaines variables (en l'occurrence les critères cités ci-dessous) estimé à partir des répondants de l'échantillon soit égal au « vrai » total (ou « marges ») connu par ailleurs. Les marges sont fournies par l'enquête Emploi (ensemble des enquêtes emploi de l'année précédente). Plusieurs critères permettent de calculer les poids qui seront ensuite attribués à chaque ménage et à chaque individu enquêté. Ce calage tient compte des critères suivants :

- au niveau du ménage : le nombre total de personnes composant le ménage, le statut d'activité de la [personne de référence](#), la catégorie de commune de résidence (taille de l'[unité urbaine](#)) ;
- au niveau des personnes composant le ménage : le nombre de personnes dans chaque sous-catégorie de sexe et âge décennal ;
- au niveau des personnes éligibles (de 15 ans ou plus) du ménage : le sexe, l'âge et le statut d'activité.

Qualité des résultats : calcul des intervalles de confiance

Le tirage des échantillons est effectué dans l'[échantillon-maître](#) issu du recensement de la population et dans la [base de sondage des logements neufs](#) achevés entre deux recensements (BSLN), alimentée par le système de suivi des permis de construire du Ministère en charge de l'Équipement.

Il résulte une marge d'erreur, liée à l'échantillonnage, et qu'il convient de mesurer pour apprécier la qualité et la fiabilité des résultats publiés.

Les formules présentées ici ont été établies en supposant que les logements ont été sélectionnés par un plan de sondage aléatoire simple sans remise (SAS) où le taux de sondage est négligeable. Or, dans la réalité ce n'est pas exactement le cas. En effet, les logements sont sélectionnés selon un plan de sondage stratifié (par catégories de communes) à plusieurs degrés, le nombre de degrés étant

variable selon les strates. Cependant, en première approximation, pour la grande majorité des indicateurs, on peut considérer que ce plan de sondage complexe est proche d'un sondage aléatoire simple, c'est-à-dire considérer comme négligeable « l'effet de grappe » dû au tirage des logements dans les zones géographiques fixes, dites « unités primaires ». On suppose de plus que les non-répondants se comportent comme les répondants.

Cas d'un pourcentage ou d'une évolution en point dans la population totale des ménages

Calcul pour un pourcentage sur une année

Soient n le nombre de ménages répondants à l'enquête et \bar{p} l'estimateur pondéré de la proportion p des ménages possédant la caractéristique dans la population.

Par exemple : en 2002, les $n = 652$ répondants dans le Sud-Ouest donnent $\bar{p} = 48\%$ de personnes étant allées au moins une fois dans l'année au cinéma.

L'[intervalle de confiance](#) à 95 % s'obtient par la formule :

$$\bar{p} \pm 2\sqrt{\frac{S^2}{n}}$$

où S^2 est la variance statistique modifiée (estimateurs sans biais de la variance) au sein de l'échantillon.

Puisque nous faisons l'hypothèse d'un sondage aléatoire simple sans remise (SAS), nous nous intéressons à l'estimation d'une proportion, nous obtenons :

$$\bar{p} \pm 2\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Calcul pour l'évolution d'une proportion entre deux années consécutives

Soient k le taux de renouvellement de l'échantillon entre les deux années consécutives, ρ le coefficient de corrélation entre les deux estimations sur la partie commune de l'échantillon, S^2 la variance statistique « calculée » sur la proportion et n la taille de l'échantillon sur lequel cette variance est calculée. On suppose S^2 et n constants entre les deux années t_1 et t_2 (sachant que l'on se place sur deux années consécutives, cette hypothèse est réaliste).

Pour un phénomène estimé par une moyenne empirique \bar{p}_t , l'estimateur naturel de l'évolution étant noté $\bar{p}_2 - \bar{p}_1$, on démontre que la variance de cette évolution est ¹ :

¹ « Estimation dans les enquêtes répétées », Caron N., Ravalet P., document de travail 00005, Unité Méthodologie Statistique, Insee, 2000.

$$V(\bar{p}_2 - \bar{p}_1) = 2(1 - \rho(1 - k)) \frac{S^2}{n}$$

Si l'échantillon est renouvelé par moitié, comme c'est le cas en général dans les EPCV entre 1996 et 2004, $k = \frac{1}{2}$; la taille de l'échantillon est supposée constante, et, du fait de l'hypothèse d'un SAS, l'intervalle de confiance à 95 % autour de $\bar{p}_2 - \bar{p}_1$ devient :

$$(\bar{p}_2 - \bar{p}_1) \pm 2\sqrt{(2 - \rho) \frac{[\bar{p}_2(1 - \bar{p}_2)]}{n}}$$

Il arrive qu'il y ait des extensions d'échantillon certaines années (par exemple en octobre 2003). Dans ce cas k est différent de $\frac{1}{2}$. De plus n ne peut plus alors être considéré comme constant. Si n_c est l'échantillon commun aux deux années, on obtient alors la formule suivante :

$$(\bar{p}_2 - \bar{p}_1) \pm 2\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} - \frac{\rho}{n_c}\right) * [\bar{p}_2(1 - \bar{p}_2)]}$$

Calcul pour une évolution d'une proportion entre deux années non consécutives

Dans ce cas, on peut supposer que les deux échantillons sont indépendants. On obtient donc :

$$V(\bar{p}_2 - \bar{p}_1) = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$$

En supposant de plus que le plan de sondage est un plan de sondage aléatoire simple, on obtient comme intervalle de confiance à 95 % autour de $\bar{p}_2 - \bar{p}_1$:

$$(\bar{p}_2 - \bar{p}_1) \pm 2\sqrt{\frac{[\bar{p}_1(1 - \bar{p}_1)]}{n_1} + \frac{[\bar{p}_2(1 - \bar{p}_2)]}{n_2}}$$

Cas d'un pourcentage ou d'une évolution dans une sous population dans les ménages

Dans ce cas, il faut remplacer n par n_D dans toutes les formules précédentes, n_D étant l'effectif dans l'échantillon de la sous population considérée.

Cas d'un pourcentage ou d'une évolution en point parmi les individus

Les individus Kish² qui répondent à l'enquête sont tirés aléatoirement au sein des ménages une fois leur composition connue et non dans la population totale (cela correspond à un degré de tirage supplémentaire). Indépendamment du mode de tirage des ménages, il ne s'agit donc pas d'un tirage d'individus par un plan de sondage aléatoire simple au sein d'une base de sondage individu. De plus, il peut y avoir un effet « grappe » sur certaines variables si on sélectionne plusieurs individus par ménage (les individus d'un même ménage pouvant avoir un comportement similaire).

Si on néglige ce possible « effet de grappe », on peut supposer que l'on a réalisé un sondage aléatoire simple d'individus. Dans ce cas, il suffit d'utiliser au niveau « individu » les formules présentées ci-dessus en les adaptant.

En revanche, si l'on cherche à prendre en compte le mode de tirage dans les calculs d'intervalle de confiance, il est alors nécessaire de « remonter » les variables « individuelles » au niveau de chaque ménage, puis de calculer une variable « synthétique » pour chaque ménage \hat{u}^k et de calculer la variance d'échantillonnage du total de cette variable qui sera une approximation de la variance de la proportion estimée.

De façon plus précise, soit $\bar{p} = \frac{\hat{X}}{\hat{N}_D}$ une proportion estimée dans l'échantillon sur la sous population

D à partir des effectifs estimés \hat{X} et \hat{N}_D .

Par exemple \hat{N}_D est la population estimée de 20-29 ans, \hat{X} le nombre estimé de 20-29 ans ayant subi une agression et \bar{p} le pourcentage de 20-29 ans agressés.

On calcule au niveau de chaque ménage la variable $u^k = \frac{1}{\hat{N}_D} (x_k - \bar{p}y_k)$ où x_k est la valeur de la variable X pour le ménage k et y_k est le nombre d'individus dans le ménage appartenant à la sous population. Si N est le nombre de ménages total (pondéré) et n le nombre de ménages enquêtés, nous avons $V(\bar{p}) = V(\hat{U}) = \frac{N^2}{n} S_U^2$. Il suffit donc de calculer la variance S_U^2 .

Calcul pour un pourcentage sur une année

Nous obtenons l'intervalle de confiance :

$$\bar{p} \pm 2 \sqrt{\frac{N^2}{n} S_U^2}$$

où N est le nombre de ménages total (pondérés) et n le nombre de ménages enquêtés.

² Au sein d'un ménage, un « individu Kish » est un individu choisi au hasard pour répondre à l'enquête.

Calcul sur deux années consécutives

Si l'échantillon est renouvelé par moitié, $k = \frac{1}{2}$, $n_1 = n_2 = n$ d'où :

$$(\bar{p}_2 - \bar{p}_1) \pm 2\sqrt{(2 - \rho) \frac{N^2}{n} S_U^2}$$

En cas d'extension ponctuelle, on a :

$$(\bar{p}_2 - \bar{p}_1) \pm 2\sqrt{N^2 \left(\frac{1}{n_1} + \frac{1}{n_2} - \frac{\rho}{n_c} \right) S_U^2}$$

Calcul sur deux années non consécutives

Dans ce cas on a :

$$(\bar{p}_2 - \bar{p}_1) \pm 2\sqrt{N^2 \left(\frac{S_U^2}{n_1} + \frac{S_U^2}{n_2} \right)}$$

Exemples de calculs

Les exemples ci-dessous sont tirés du fichier historique 1996-2003, indicateurs sociaux d'octobre « Participation et contacts sociaux ».

% d'individus qui sont allés au cinéma au cours des douze derniers mois pour les années 2002 et 2003

Zone d'étude et d'aménagement du territoire (ZEAT)	2002			2003			2002-2003	
	%	Intervalle de confiance à 95 % (SAS)	Effectif brut ayant répondu à la question	%	Intervalle de confiance à 95 %	Effectif brut ayant répondu à la question	Variation en points	Intervalle de confiance à 95 % (SAS)
Sud-Ouest	48,0	{ 45,8 ; 50,1 }	652	46,9	{ 45,2 ; 48,5 }	1 115	-1,1	{ -3,8 ; 1,6 }
France métropolitaine	52,6	{ 51,7 ; 53,5 }	5 823	51,6	{ 50,9 ; 52,2 }	10 272	-1,1	{ -2,1 ; 0 }

Lecture : dans la ZEAT Sud-Ouest, 46,9% des 1 115 individus interrogés dans l'enquête ont déclaré, pour l'année 2003, être allés au cinéma au cours des douze derniers mois. Le pourcentage « vrai », dans l'ensemble de la population de la ZEAT, a 95 % de chances d'être compris en 45,2 % et 48,5 % ; il est donc estimé à + ou - 1,7 points.

Pour la France métropolitaine, où 10 272 individus ont répondu à l'enquête, le pourcentage « vrai » (estimé à 51,6 %) a 95 % de chances d'être compris entre 50,9 % et 52,2 % ; il est donc estimé à + ou - 0,6 point.

Par rapport à l'ensemble du fichier historique, l'année 2003 est une année atypique car l'échantillon sélectionné avait bénéficié d'une importante extension. En comparant avec l'échantillon 2002, nous pouvons constater que plus l'effectif enquêté est important, plus l'intervalle de confiance se resserre. Inversement, on doit toujours considérer que plus l'effectif est minime, moins les conclusions tirées des résultats obtenus sont fiables (plus les intervalles de confiance sont « larges »).

En réalité, la longueur des intervalles, obtenus par les formules précédentes et publiés par ailleurs, est un peu sous estimée du fait de l'hypothèse d'un sondage aléatoire simple (SAS). Si, à titre d'exemple, on tient compte du fait que le plan de sondage est à plusieurs degrés stratifié par catégories de communes, on obtient des résultats légèrement différents. A titre d'exemple, pour la variable cambriolage, variable des EPCV de janvier, au niveau ménage on obtient exactement le même intervalle. En revanche, pour la variable agression au niveau individu, si pour la France l'intervalle est identique, pour la ZEAT Nord le « vrai » pourcentage est estimé à + ou – 2,1 points, au lieu de + ou - 1,9 points dans le cas d'un SAS.

% en 2004 de ménages cambriolés au cours des deux dernières années

Zone d'étude et d'aménagement du territoire (ZEAT)	%	Intervalle de confiance à 95% (SAS)	Intervalle de confiance à 95% (sondage plusieurs degrés, stratifié)	Effectif brut ayant répondu à la question
Nord	2,5	{ 1,0 ; 4,0 }	{ 1,0 ; 4,0 }	400
France métropolitaine	2,5	{ 2,1 ; 2,9 }	{ 2,1 ; 2,9 }	6 440

Lecture : dans la ZEAT Nord comme en France métropolitaine, 2,5 % de ménages ont déclaré en 2004 avoir été cambriolés au cours des années 2002-2003. Pour la France, où 6 440 ménages ont répondu à l'enquête, le pourcentage « vrai », dans l'ensemble de la population de la ZEAT, a 95 % de chances d'être compris entre 2,1 % et 2,9 %, il est donc estimé à + ou – 0,4 point et ce, que ce soit en faisant l'hypothèse d'un sondage aléatoire simple ou non. Dans la ZEAT Nord, où seulement 400 ménages ont répondu, ce même pourcentage « vrai » a 95 % de chances d'être compris en 1 % et 4 %, il est donc estimé à + ou – 1,5 points.

% en 2004 d'individus agressés au cours des deux dernières années

Zone d'étude et d'aménagement du territoire (ZEAT)	%	Intervalle de confiance à 95% (SAS)	Intervalle de confiance à 95% (sondage plusieurs degrés, stratifié)	Effectif brut ayant répondu à la question
Nord	7,8	{ 5,9 ; 9,7 }	{ 5,7 ; 9,9 }	833
France métropolitaine	6,7	{ 6,2 ; 7,2 }	{ 6,2 ; 7,2 }	11 701

Lecture : dans la ZEAT Nord 7,8 % des 83 individus interrogés dans l'enquête ont déclaré en 2004 avoir été agressés au cours des années 2002-2003. Le pourcentage « vrai », dans l'ensemble de la population de la ZEAT, a 95 % de chances d'être compris en 5,9 % et 9,7 % ; il est donc estimé à + ou – 1,9 points si l'on fait l'hypothèse d'un SAS. Sous l'hypothèse d'un sondage à plusieurs degrés stratifié, ce pourcentage « vrai » est estimé à + ou - 2,1 points. Pour la France métropolitaine où 11 701 individus ont répondu à l'enquête, le pourcentage « vrai » (estimé à 6,7 %) a 95 % de chances d'être compris entre 6,2 % et 7,2 %, il est donc estimé à + ou – 0,5 point (avec les mêmes résultats sous l'hypothèse d'un sondage aléatoire simple ou non).

La différence peut être plus importante, notamment pour des questions d'opinion comme la préoccupation pour le manque de sécurité dans le quartier, où l'effet de grappe est plus important dans le tirage des individus répondants au niveau des ménages.

% en 2004 d'individus pour qui le manque de sécurité est un problème dans leur quartier

Zone d'étude et d'aménagement du territoire (ZEAT)	%	Intervalle de confiance à 95% (SAS)	Intervalle de confiance à 95% (sondage plusieurs degrés, stratifié)	Effectif brut ayant répondu à la question
Nord	17,6	{ 14,8 ; 20,4 }	{ 14,0 ; 21,2 }	833
France métropolitaine	13,0	{ 12,4 ; 13,6 }	{ 12,1 ; 13,9 }	11 701

Lecture : pour la France, les 13 % sont estimés à + ou - 0,6 point en faisant l'hypothèse d'un SAS, à + ou - 0,9 point avec l'hypothèse d'un sondage à plusieurs degrés stratifié. Pour le Nord, les 17,6 % sont estimés à + ou - 2,8 points en faisant l'hypothèse d'un SAS, à + ou - 3,6 points avec l'hypothèse d'un sondage à plusieurs degrés stratifié.

Glossaire

Échantillon maître

L'échantillon maître de l'Insee, issu du recensement de 1999, est une liste de logements établie à partir de ce recensement de la population. Elle contient les logements achevés à la date du recensement. L'échantillon maître ne concerne que la France métropolitaine.

Il convient de compléter cet échantillon maître grâce à la base de sondage des logements neufs.

Base de sondage des logements neufs

La base de sondage des logements neufs (BSLN) complète l'échantillon maître et permet sa mise à jour annuelle pour le tirage des échantillons de logements utilisés dans les enquêtes ménages de l'Insee. Elle regroupe une partie des logements neufs achevés après le recensement de 1999 et situés dans les mêmes communes que celles de l'échantillon maître. Cette base est établie à partir du fichier des permis de construire du Ministère en charge de l'équipement (fichier national de permis de construire déclarés achevés) et d'une enquête d'achèvement sur le terrain.

Intervalle de confiance

L'analyse statistique et économétrique ne procure pas des résultats exacts à 100 %, mais les résultats les plus probables. L'intervalle de confiance est une estimation de l'écart entre la plus petite valeur (limite inférieure) et la plus grande valeur (limite supérieure) de résultat possible. Le véritable résultat se situe à l'intérieur de cet intervalle de confiance, avec une certaine probabilité. Dans les publications, on présente généralement un intervalle de confiance à 95 %, ce qui signifie qu'il y a 95 % de chances que le véritable résultat se situe à l'intérieur des limites identifiées.