



**Conseil économique
et social**

Distr.
GÉNÉRALE

ECE/CES/AC.6/2008/4
29 février 2008

FRANÇAIS
Original: ANGLAIS

COMMISSION ÉCONOMIQUE POUR L'EUROPE

CONFÉRENCE DES STATISTICIENS EUROPÉENS

Réunion commune CEE/Eurostat sur les recensements
de la population et des habitations

Onzième réunion
Genève, 13-15 mai 2008
Point 2 de l'ordre du jour provisoire

ASSURANCE ÉVALUATION DE LA QUALITÉ DES RECENSEMENTS

**Précision attendue des résultats obtenus à partir d'échantillons lors
du recensement italien de la population et des logements**

Note présentée par l'Institut national italien de statistique (Istat) (Italie)

Résumé

L'Istat envisage d'utiliser, parmi de nombreuses autres procédures novatrices à l'étude pour le prochain recensement, des techniques de sondage (fondées sur l'emploi de questionnaires longs et de questionnaires courts). Dans les grandes communes (dont la population dépasse 10 000 habitants), des informations détaillées seraient recueillies pour un échantillon de la population seulement (à l'aide du questionnaire long), tandis que pour le reste de la population seules des informations limitées, principalement d'ordre démographique, seraient collectées (à l'aide du questionnaire court). Dans les communes plus petites, une démarche traditionnelle est prévue, le questionnaire long étant communiqué à l'ensemble de la population.

Des essais ont été effectués et des études ont été menées en vue d'évaluer la qualité des estimations pouvant être obtenues à partir d'un sondage. Quelques résultats préliminaires des essais sont présentés ci-après. Ils semblent inciter à l'utilisation de techniques de sondage, et donc à l'adoption, pour le recensement de la population italienne de 2011 d'une stratégie de collecte de données à l'aide de questionnaires courts ou longs selon les cas.

I. INTRODUCTION

1. L'Institut national italien de statistique (Istat) a récemment commencé à évaluer des procédures novatrices de recensement, qui visent à rationaliser les opérations d'enquêtes et à alléger la charge de travail statistique. Ces activités consistent à mettre au point des stratégies de sondage devant permettre, à l'occasion du recensement de la population de 2011 en Italie, d'obtenir des informations sur un sous-ensemble de variables pertinentes pour le recensement.
2. La principale stratégie vise à scinder l'ensemble des variables prises en compte dans le recensement en deux sous-ensembles, le premier contenant les variables démographiques et le second les autres variables (degré d'instruction, situation dans la profession, migrations pendulaires). Les versions courte et longue du questionnaire s'appliqueraient donc, respectivement, au premier sous-ensemble de variables et à l'ensemble complet des variables. Pour les communes de plus de 10 000 habitants, un échantillon de ménages serait interrogé à l'aide du questionnaire long tandis que pour le reste des ménages, on emploierait la version courte. Dans les communes plus petites, par contre, le questionnaire long serait utilisé pour toutes les unités de population.
3. S'agissant de la stratégie de sondage¹, l'Istat doit étudier avec soin les nombreuses difficultés se posant, telles que la constitution d'échantillons à partir des listes de population, la définition d'échantillons pour certaines zones suburbaines ou non urbaines, qui ont vu leur population augmenter ou diminuer considérablement entre deux recensements consécutifs, et la définition de secteurs qui regroupent plusieurs districts de recensement et pour lesquels les estimations doivent être établies.
4. La principale contrainte affectant la définition de la stratégie de sondage² est la précision des estimations obtenues à différentes échelles territoriales. Plus le territoire de référence est vaste, plus grande devrait être la précision, d'une part, des estimations concernant les variables sur lesquelles porte le questionnaire long et, d'autre part, de leur classification croisée avec d'autres variables appartenant à la même classe ou au domaine démographique (hors échantillonnage).

II. STRATÉGIES DE SONDAGE

5. Deux études de nature différente ont été menées. Dans la première, on a analysé la qualité d'un sondage aléatoire simple des ménages choisis dans les registres administratifs tenus par les communes. On a considéré différents taux d'échantillonnage (10, 15, 20 et 33 %) afin d'évaluer l'amélioration de la précision des estimations avec l'augmentation des fractions sondées. Dans la seconde étude, on a examiné une technique de sondage aréolaire fondée sur l'échantillonnage aléatoire simple des districts de recensement, les unités prélevées représentant environ un tiers de la population. Cette démarche incluait une collecte de données complète sur le logement des ménages dans les districts de recensement choisis (à partir d'une base de données numériques géoréférencées).

¹ Carbonetti G., De Vitiis C. (2007) «*Efficienza di stime campionarie relative ad un sottoinsieme di variabili di censimento*», Conférence italienne de statistique: «*Censimenti generali 2010-2011. Criticità e innovazioni*». CNR, Rome (Italie), novembre 2007.

² Särndal C.E., Swensson B., Wretman J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York.

6. Dans les deux cas, afin d'évaluer leurs propriétés en termes de grandeur du coefficient de variation, les stratégies de sondage ont été comparées l'une à l'autre en procédant à une réplique de l'échantillonnage selon la méthode de Monte Carlo, sur des données issues du recensement de 2001.

7. Les principaux concepts du plan de sondage sont résumés ci-après:

- a) Secteurs: «districts» rapportés à une division territoriale infracommunale³;
- b) Variables cibles: «variables» liées à la classification croisée du degré d'instruction, de la situation dans la profession et des migrations pendulaires avec les variables démographiques;
- c) Unités d'échantillonnage: «ménages» ou «districts de recensement», selon la stratégie adoptée;
- d) Estimateur: «estimateurs calibrés»⁴ employant des coefficients de pondération finals dûment modifiés⁵ dans le but de rendre l'échantillon plus représentatif.

III. ÉTUDE DE SIMULATION

8. L'étude de simulation a été effectuée sur les données du recensement de la population de 2001. On a pris en considération un ensemble de 40 communes italiennes, qui avaient des effectifs de population différents et étaient situées dans diverses régions du pays, afin de tenir compte des fortes différences entre les communes italiennes (tableau A).

Tableau A: Répartition des communes par zone géographique et par effectif démographique

Zone géographique	Classes d'effectif de la population ^a			Total
	10 000-20 000	20 000-100 000	plus de 100 000	
Nord	4	6	6	16
Centre	2	3	3	8
Sud	4	6	6	16
Total	10	15	15	40

^a Il a été considéré que la date légale (officielle) du relevé de la population était celle du recensement de la population de 2001.

³ Astorri P., Bianchi G., Di Pede F., Esposito N., Patruno E., Reale A., Ronchi I., Talice S. (2007) «*Metodi di determinazione delle aree di censimento a livello sub comunale*», vingt-huitième Conférence italienne sur les sciences régionales, Bolzano (Italie), septembre 2007.

⁴ Deville J.C., Särndal, C.E. (1992) Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, vol. 87, p. 367 à 382.

⁵ Logiciel Genesees v3.0 mis au point au sein de l'Istat, qui a servi à calculer les coefficients de pondération finals dans le processus de calibrage.

9. L'attention a été centrée sur une classification croisée du degré d'instruction, de la situation dans la profession, des migrations pendulaires et du sexe pour un nombre total de 90 cellules d'estimation simple.

10. Les contraintes de calibrage ont été définies par une classification croisée du sexe avec l'âge et du sexe avec l'état civil pour un nombre total de 40 fréquences de référence.

11. Un algorithme de calcul, utilisant le code SAS et comprenant les étapes suivantes, a été appliqué à chaque commune et à chaque variante du plan de sondage:

- 1) Choix d'un échantillon (ménages ou districts de recensement);
- 2) Calcul des coefficients de pondération finals;
- 3) Estimation des fréquences relatives p pour chaque cellule cible dichotomique;
- 4) Répétition, un nombre donné de fois, de l'exécution des étapes 1), 2) et 3) (pour des raisons pratiques, on a choisi de procéder à 1 000 réplifications de l'échantillonnage);
- 5) Pour chaque cellule dichotomique d'estimation X , calcul de la moyenne p_x et de l'erreur type $\sigma(\hat{p}_x)$ des estimations à partir de la distribution simulée d'échantillonnage.

12. Pour comparer les stratégies de sondage, on a examiné le critère d'évaluation du coefficient de variation $cv(\hat{p}_x) = \frac{\sigma(\hat{p}_x)}{p_x} \cdot 100$. En fait, il correspond à une mesure de la précision des stratégies de sondage, permettant d'estimer la valeur inconnue de la fréquence relative p pour chaque cellule X .

13. On a déterminé la distribution des coefficients de variation empiriques pour l'ensemble des 90 cellules cibles. Après avoir classé les cellules cibles en fonction des valeurs de leur fréquence relative p , on a étudié la distribution des coefficients de variation relatifs à la variable dans le même groupe.

14. Le tableau B fait état du nombre d'unités prises en compte dans l'étude de simulation.

Tableau B: Étendue de la simulation en termes de districts,
de districts de recensement, de ménages et de personnes

	Unités échantillonnées	Total	%
Districts	497	3 347*	14,85
Districts de recensement	30 890	382 534	8,08
Ménages	2 243 511	21 810 676	10,29
Personnes	5 537 582	56 594 021	9,78

* Nombre estimatif.

IV. QUELQUES RÉSULTATS

15. Le tableau 1 indique le coefficient de variation médian, en pourcentage, dans les districts, pour différentes classes de fréquence relative estimée p et pour chacun des quatre taux d'échantillonnage considérés. Afin que l'étude de simulation reste gérable, on a testé chaque taux d'échantillonnage sur 10 communes, 5 d'entre elles étant retenues dans tous les tests, tandis que les autres étaient choisies dans la même classe de population. Lorsqu'on adopte un taux d'échantillonnage de 33 %, des fréquences relatives p comprises entre 0,5 et 1 % peuvent être estimées avec un coefficient de variation médian d'environ 16 %. Ce pourcentage est multiplié par deux lorsque le taux d'échantillonnage est de 10 %. Ceci vaut pour toutes les classes de fréquence relative p . En général, on observe, pour presque toutes les classes de fréquence relative p , un gain relativement constant en termes de diminution du coefficient de variation pour les taux d'échantillonnage testés.

Tableau 1: Distribution du coefficient de variation médian pour les classes de fréquence relative p . Comparaison de quatre taux d'échantillonnage différents dans le sondage aléatoire simple des ménages choisis dans les registres administratifs

Classes de fréquence relative p	Taux d'échantillonnage = 10 %	Taux d'échantillonnage = 15 %	Taux d'échantillonnage = 20 %	Taux d'échantillonnage = 33 %
	170 districts	140 districts	111 districts	204 districts
< 0,05 %	220,51	157,20	142,00	98,21
0,05 % 0,1 %	111,48	87,22	74,20	51,14
0,1 % 0,25 %	75,57	59,83	49,97	34,76
0,25 % 0,5 %	50,70	39,92	33,97	23,44
0,5 % 1 %	35,54	28,10	23,74	16,56
1 % 2,5 %	23,62	18,56	15,33	10,68
2,5 % 5 %	15,50	12,29	10,09	7,04
5 % 10 %	10,46	8,26	6,93	4,82
10 % 15 %	7,06	5,40	4,40	3,13
15 % 20 %	5,57	4,27	3,54	2,42
20 % 30 %	4,50	3,48	2,84	1,93
≥ 30 %	3,20	2,42	1,94	1,34

16. Dans le tableau 2 figurent les mêmes résultats que dans le tableau 1, sous une forme affinée, les districts étant ventilés en trois classes de population (moins de 10 000 habitants; de 10 001 à 12 000 habitants; plus de 12 000 habitants). Il en ressort assez nettement que pour les districts de plus de 12 000 habitants, le coefficient de variation diminue sensiblement.

Tableau 2: Distribution du coefficient de variation médian pour les classes de fréquence relative p et trois classes de districts (déterminées en fonction de l'effectif de la population). Comparaison de quatre taux d'échantillonnage différents dans le plan de sondage aléatoire simple des ménages choisis dans les registres administratifs

Classes de fréquence relative p	Population par district (en milliers)	Taux d'échantillonnage = 10 %	Taux d'échantillonnage = 15 %	Taux d'échantillonnage = 20 %	Taux d'échantillonnage = 33 %
< 0,05 %	< 10	233,15	177,67	149,83	102,20
	10 ┤ 12	225,61	145,74	143,95	88,14
	≥ 12	184,05	141,61	117,45	82,77
0,05 % ┤ 0,1 %	< 10	136,29	109,01	88,57	61,14
	10 ┤ 12	107,97	85,49	72,90	50,61
	≥ 12	100,88	79,37	64,90	46,30
0,1 % ┤ 0,25 %	< 10	90,00	71,27	59,48	40,20
	10 ┤ 12	76,23	60,03	50,45	34,41
	≥ 12	66,65	53,04	43,51	30,58
0,25 % ┤ 0,5 %	< 10	60,88	47,23	40,24	27,29
	10 ┤ 12	50,87	39,73	34,10	23,40
	≥ 12	45,22	35,68	29,93	21,01
0,5 % ┤ 1 %	< 10	43,11	33,50	28,97	19,53
	10 ┤ 12	35,08	27,46	22,99	16,48
	≥ 12	31,25	24,95	20,97	14,85
1 % ┤ 2,5 %	< 10	28,85	21,82	18,11	12,42
	10 ┤ 12	23,37	18,00	14,92	10,62
	≥ 12	21,10	16,46	13,45	9,48
2,5 % ┤ 5 %	< 10	19,12	14,68	12,22	8,25
	10 ┤ 12	15,58	12,36	9,89	7,08
	≥ 12	14,00	10,98	9,06	6,35
5 % ┤ 10 %	< 10	12,70	9,77	8,14	5,65
	10 ┤ 12	10,36	8,01	6,90	4,82
	≥ 12	9,37	7,43	6,21	4,36
10 % ┤ 15 %	< 10	8,78	6,44	5,22	3,67
	10 ┤ 12	7,00	5,46	4,41	3,13
	≥ 12	6,29	4,79	3,89	2,83
15 % ┤ 20 %	< 10	6,79	5,03	4,31	2,85
	10 ┤ 12	5,47	4,18	3,41	2,31
	≥ 12	4,99	3,80	3,09	2,17
20 % ┤ 30 %	< 10	5,46	4,16	3,44	2,27
	10 ┤ 12	4,57	3,42	2,94	2,01
	≥ 12	4,05	3,20	2,59	1,77
≥ 30 %	< 10	3,85	2,84	2,32	1,56
	10 ┤ 12	3,14	2,30	1,88	1,32
	≥ 12	2,83	2,17	1,74	1,18

17. Puis sont donnés dans le tableau 3 les coefficients de variation médians pour trois communes particulières, situées dans différentes zones géographiques de l'Italie (Bologne dans le nord, Pérouge dans le centre et Trapani dans le sud). Même si d'autres analyses devraient être faites pour vérifier si le coefficient de variation estimé accuse éventuellement des différences d'un point de vue géographique, il ne semble pas qu'il y ait de grandes différences de cette nature entre les communes considérées ci-dessous.

Tableau 3: Distribution du coefficient de variation médian pour les classes de fréquence relative p . Comparaison de quatre taux d'échantillonnage (t. e.) différents dans le plan de sondage aléatoire simple des ménages choisis dans les registres administratifs.
Communes de Bologne, Pérouge et Trapani

Classes de fréquence relative p	Bologne (32 districts)				Pérouge (10 districts)				Trapani (5 districts)			
	t.e. = 10 %	t.e. = 15 %	t.e. = 20 %	t.e. = 33 %	t.e. = 10 %	t.e. = 15 %	t.e. = 20 %	t.e. = 33 %	t.e. = 10 %	t.e. = 15 %	t.e. = 20 %	t.e. = 33 %
< 0,05 %	206,09	164,99	137,07	96,52	199,94	164,00	137,12	94,68	223,25	173,66	144,99	102,00
0,05 % 0,1 %	107,70	85,78	72,14	50,67	105,78	83,96	70,74	49,14	120,35	94,89	78,63	56,37
0,1 % 0,25 %	76,59	60,19	50,74	35,00	76,12	61,17	52,04	36,44	68,70	51,88	44,30	31,03
0,25 % 0,5 %	52,27	41,09	34,45	24,17	49,04	38,20	32,20	21,87	49,69	38,13	32,47	22,49
0,5 % 1 %	36,50	28,93	24,35	16,85	34,32	27,24	22,77	15,97	32,52	25,37	20,83	14,86
1 % 2,5 %	23,28	18,27	15,28	10,74	22,44	17,41	14,53	10,24	24,20	18,58	15,95	10,96
2,5 % 5 %	15,47	12,03	9,98	7,07	14,86	11,48	9,66	6,76	15,57	12,13	10,15	7,05
5 % 10 %	10,43	8,16	6,76	4,75	10,50	8,21	6,76	4,77	10,62	8,40	6,98	4,83
10 % 15 %	6,83	5,28	4,38	3,09	6,54	5,11	4,22	2,97	6,68	5,17	4,44	2,96
15 % 20 %	5,19	4,02	3,41	2,38	5,45	4,26	3,53	2,41	5,01	3,94	3,29	2,31
20 % 30 %	4,23	3,26	2,78	1,92	4,20	3,26	2,74	1,92	4,63	3,67	3,03	2,15
≥ 30 %	2,84	2,25	1,86	1,32	2,62	2,07	1,77	1,21	3,30	2,51	2,09	1,41

18. Le tableau 4 reprend le tableau 2. Y sont comparés, d'une manière plus détaillée pour la commune de Bologne, les coefficients de variation médians pour différentes classes de fréquence relative p , différentes classes de populations de district et différents taux d'échantillonnage. Sachant que Bologne a été choisie parce que ses 32 districts échantillonnés permettent de calculer convenablement les données pour l'ensemble des cellules du tableau 4, il ne semble pas qu'il y ait de grandes différences entre les coefficients de variation indiqués dans ce tableau et ceux du tableau 2.

19. Dans le tableau 5 sont données les distributions, en pourcentage, des estimations de la fréquence relative p par classe de coefficient de variation pour chaque taux d'échantillonnage testé. Lorsque l'on passe d'un taux d'échantillonnage de 10 % à un taux de 33 %, on constate une augmentation uniforme de la distribution, en pourcentage, jusqu'à un seuil du coefficient de variation de 10 %, puis une diminution correspondante au-delà de cette valeur.

Tableau 4: Distribution du coefficient de variation médian pour les classes de fréquence relative p et trois classes de districts (en fonction de l'effectif de la population).

Comparaison de quatre taux d'échantillonnage différents dans le plan de sondage aléatoire simple des ménages choisis dans les registres administratifs.

Commune de Bologne

Classes de fréquence relative p	Population par district (en milliers)	Taux d'échantillonnage = 10 %	Taux d'échantillonnage = 15 %	Taux d'échantillonnage = 20 %	Taux d'échantillonnage = 33 %
< 0,05 %	< 10	219,16	174,43	147,14	98,84
	10-12	220,94	172,48	144,22	99,28
	≥ 12	180,79	140,52	118,01	83,60
0,05 %-0,1 %	< 10	138,44	101,63	88,69	60,32
	10-12	106,08	83,85	70,45	49,64
	≥ 12	100,56	79,05	64,11	46,23
0,1 %-0,25 %	< 10	89,74	69,56	58,36	40,89
	10-12	76,74	60,76	51,22	35,53
	≥ 12	66,63	52,92	44,46	30,78
0,25 %-0,5 %	< 10	58,78	46,06	39,06	27,05
	10-12	52,00	41,09	34,58	23,91
	≥ 12	45,87	36,52	30,67	21,71
0,5 %-1 %	< 10	45,08	34,37	29,01	20,22
	10-12	35,41	28,45	23,52	16,34
	≥ 12	31,89	24,85	20,86	14,88
1 %-2,5 %	< 10	27,27	21,54	18,01	12,53
	10-12	23,00	18,07	15,13	10,61
	≥ 12	20,74	16,39	13,90	9,63
2,5 %-5 %	< 10	18,54	14,49	11,88	8,29
	10-12	15,42	12,04	9,88	7,05
	≥ 12	13,70	10,74	8,99	6,23
5 %-10 %	< 10	12,12	9,35	7,76	5,44
	10-12	10,38	8,09	6,65	4,72
	≥ 12	9,43	7,42	6,10	4,29
10 %-15 %	< 10	8,27	6,29	5,21	3,64
	10-12	6,93	5,33	4,46	3,19
	≥ 12	6,10	4,71	3,90	2,77
15 %-20 %	< 10	6,38	4,98	4,18	2,88
	10-12	5,06	3,91	3,16	2,23
	≥ 12	4,54	3,59	3,02	2,13
20 %-30 %	< 10	5,12	3,95	3,34	2,33
	10-12	4,33	3,34	2,83	1,98
	≥ 12	3,72	2,93	2,45	1,72
≥ 30 %	< 10	3,46	2,69	2,23	1,50
	10-12	2,75	2,14	1,80	1,27
	≥ 12	2,73	2,12	1,78	1,23

Tableau 5: Distribution des estimations relatives à des districts de plus de 12 000 habitants pour les classes du coefficient de variation. Comparaison des fréquences, en pourcentage, pour quatre taux d'échantillonnage différents dans le plan de sondage aléatoire simple des ménages choisis dans les registres administratifs

Classes du coefficient de variation	Taux d'échantillonnage = 10 %	Taux d'échantillonnage = 15 %	Taux d'échantillonnage = 20 %	Taux d'échantillonnage = 33 %
< 2 %	0,57	2,69	6,39	13,14
2 % 5 %	13,04	17,53	18,40	23,64
5 % 10 %	16,18	18,02	26,28	28,64
10 % 20 %	29,14	30,16	23,54	16,20
20 % 50 %	25,09	19,71	16,75	13,32
50 % 100 %	9,32	7,21	5,69	3,44
100 % 200 %	4,40	3,65	2,00	1,61
≥ 200 %	2,25	1,03	0,95	-

20. Dans le tableau 6 sont comparés, pour un taux d'échantillonnage de 33 %, le plan d'échantillonnage pour le sondage aléatoire simple des ménages choisis dans les registres administratifs tenus par les municipalités (SRSHOU) et le plan d'échantillonnage pour le sondage aléatoire simple des districts de recensement (SRSENA). On a calculé les coefficients de variation médians pour les deux plans d'échantillonnage, pour diverses classes de fréquence relative p et pour chacune des quatre communes. En raison de l'effet de grappe, les erreurs d'échantillonnage dans les plans d'échantillonnage des ménages sont toujours plus faibles que celles observées dans les plans d'échantillonnage des districts. Toutefois, comme le plan SRSENA ne s'avère pas tellement moins performant que le plan SRSHOU lorsqu'on l'évalue en termes d'efficacité, son emploi, dans les cas où des registres fiables des ménages ne sont pas disponibles, pourrait permettre d'obtenir des estimations d'un niveau de qualité acceptable.

Tableau 6: Distribution du coefficient de variation médian pour les classes de fréquence relative p et pour les plans SRSHOU et SRSENA (taux d'échantillonnage de 33 %).
Comparaison de quatre communes

Classes de fréquence relative p	Milan (111 districts)		Bologne (32 districts)		Padoue (18 districts)		Livourne (13 districts)	
	SRSHOU	SRSENA	SRSHOU	SRSENA	SRSHOU	SRSENA	SRSHOU	SRSENA
< 0,05 %	97,78	94,12	96,52	94,31	99,65	98,34	102,21	101,61
0,05 % 0,1 %	51,61	51,59	50,67	49,54	51,70	54,13	50,69	52,06
0,1 % 0,25 %	34,67	34,92	35,00	35,20	35,37	36,03	35,08	35,67
0,25 % 0,5 %	22,96	24,38	24,17	24,73	25,58	26,45	23,70	24,37
0,5 % 1 %	16,86	18,71	16,85	18,32	16,95	17,81	17,16	18,72
1 % 2,5 %	10,61	12,21	10,74	11,95	11,07	12,00	11,34	12,90
2,5 % 5 %	7,02	8,53	7,07	8,25	7,35	8,48	7,17	9,00
5 % 10 %	4,84	5,97	4,75	5,74	5,05	5,85	4,88	6,39
10 % 15 %	3,17	4,41	3,09	4,09	3,19	4,37	3,06	4,82
15 % 20 %	2,44	3,46	2,38	3,12	2,44	3,14	2,44	3,39
20 % 30 %	1,89	2,61	1,92	2,48	2,08	2,73	2,05	2,88
≥ 30 %	1,35	1,78	1,32	1,60	1,39	1,72	1,40	2,00

21. Les estimations portant sur des zones territoriales plus grandes que le district (commune, niveaux Nuts3, Nuts2 et Nuts1 de la nomenclature des unités territoriales statistiques) ont été établies en agrégeant les districts. Par exemple, pour un grand secteur R comportant k districts, l'estimation de la fréquence relative p est obtenue en utilisant la formule suivante:

$$\hat{p}_R(x) = \sum_{a \in R} w_a \hat{p}_a \text{ où } w_a = N_a / N_R \text{ correspond au poids, en termes de population, du district}$$

générique a dans le grand secteur R . On peut montrer que pour une fréquence relative p donnée (concernant tant un district unique qu'un secteur plus grand), la formule $cv(\hat{p}_R) \equiv \frac{1}{\sqrt{K}} cv(\hat{p}_a)$

s'applique avec une réduction prévue, en pourcentage, du coefficient de variation y relatif, qui est donnée par la formule suivante:

$$réd.\% \equiv \left(1 - \frac{1}{\sqrt{K}}\right) \times 100.$$

Lorsqu'une partie du grand secteur R fait l'objet d'une enquête en totalité, la réduction prévue du coefficient de variation est donnée par la formule suivante: $réd.\% \equiv \left(1 - \frac{\gamma}{\sqrt{K}}\right) \times 100$ où $\gamma = N_C / N$ est la proportion de la population soumise à une enquête par sondage.

V. CONCLUSIONS

22. L'Istat envisage d'utiliser, parmi de nombreuses autres procédures novatrices à l'étude pour le prochain recensement, des techniques de sondage. Plus précisément, seules quelques informations, intéressant principalement les variables démographiques, seraient recueillies auprès de l'ensemble des répondants à l'aide d'un questionnaire court, tandis qu'un questionnaire plus détaillé (questionnaire long) contenant des informations complètes serait rempli par un échantillon de personnes interrogées. Par ailleurs, il est prévu d'inclure seulement dans l'échantillon les logements situés dans des communes de plus de 10 000 habitants, qui représentent environ 40 millions de personnes au total. Une démarche plus traditionnelle, consistant à soumettre le questionnaire long à l'ensemble de la population, est par contre envisagée pour les communes de moins de 10 000 habitants, qui représentent dans l'ensemble 18 millions de personnes.

23. Des essais ont été effectués et des expériences ont été menées en vue d'évaluer la qualité des estimations par sondage, s'agissant de fréquences de grandeurs très différentes. Des simulations selon la méthode de Monte Carlo des espaces d'échantillons ont été réalisées pour un sondage aléatoire simple des ménages avec différents taux d'échantillonnage et pour un sondage aréolaire.

24. Les premiers résultats semblent inciter à l'utilisation de telles techniques de sondage, donc à l'adoption pour le recensement de la population italienne de 2011 d'une stratégie de collecte de données à l'aide de questionnaires longs et de questionnaires courts. Les estimations les plus précises ont bien sûr été observées dans le cas du sondage aléatoire simple des ménages choisis dans les registres administratifs et pour le plus fort taux d'échantillonnage. Puisqu'il faut parvenir à concilier la dimension de l'échantillon et les coûts de la collecte de données, le choix

final dépendra de l'équilibre que l'on trouvera entre les considérations budgétaires et le degré de précision minimum requis aux diverses échelles territoriales.

25. Des estimations plus précises étant obtenues pour les districts les plus grands en termes d'effectif de la population, il pourrait être proposé, en vue de concevoir le plan de sondage, de définir des districts d'environ 15 000 habitants.

26. Même si le sondage aréolaire s'avère être moins performant que le sondage aléatoire simple des ménages, les résultats des simulations indiquent qu'il pourrait constituer une solution de remplacement lorsque des registres administratifs fiables ne sont pas disponibles.

27. On continuera d'envisager d'autres démarches, fondées sur des techniques d'estimation pour des petits districts, tant pour les échelles territoriales les plus petites que pour des populations rares.
