Rapport de méthodes

Estimation de la couverture du recensement de la population de l'an 2000

0

Statistische Grundlagen und Übersichten
Bases statistiques et produits généraux
Basi statistiche e presentazioni generali

Echantillon pour l'estimation de la sur-couverture (E-sample)

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica

OFS BFS UST

Neuchâtel, 2003

Statistik der Schweiz

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Medien, Zeitverwendung
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Rechtspflege
20 Einkommen und Lebensqualität der Bevölkerung
21 Nachhaltige Entwicklung und regionale Disparitäten

Statistique de la Suisse

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

0 Bases statistiques et produits généraux
1 Population
2 Espace et environnement
3 Vie active et rémunération du travail
4 Economie nationale
5 Prix
6 Industrie et services
7 Agriculture et sylviculture
8 Energie
9 Construction et logement
10 Tourisme
11 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, médias, emploi du temps
17 Politique
18 Administration et finances publiques
19 Droit et justice
20 Revenus et qualité de vie de la population
21 Développement durable et disparités régionales

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Estimation de la couverture du recensement de la population de l'an 2000

Echantillon pour l'estimation de la sur-couverture (E-sample)

Auteur **Anne Renaud**
Office fédéral de la statistique

Editeur **Office fédéral de la statistique**

OFS BFS UST

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica

Neuchâtel, 2003

Préambule

Le projet «estimation de la couverture du recensement» fait partie des projets visant à évaluer la qualité du recensement de la population de l'an 2000 (RFP2000). Ce rapport présente le plan d'échantillonnage de l'échantillon permettant d'estimer la composante de sur-couverture du recensement. Le travail a été réalisé par Anne Renaud du service de méthodes statistiques (METH).

Les résultats décrits dans ce rapport ont été obtenus grâce à l'apport de plusieurs personnes. Un grand merci à Philippe Eichenberger de METH pour les discussions méthodologiques aux étapes clés du projet et la relecture du rapport. On remerciera également MM. Rajendra Singh et David Whitford du US Census Bureau pour leur contribution lors de leur visite en mars 2002 en Suisse, et leur appui régulier par email depuis leur retour aux Etats-Unis. Merci aussi à Rachel Fritschi, Françoise Rossé et Stéphane Rey du projet RFP2000 pour le travail effectué sur les données de l'enquête de couverture et celles du recensement. Merci encore à Jann Potterat de METH pour la pondération provisoire du P-sample.

Résumé

Les défauts de couverture du recensement de la population de l'an 2000 (RFP2000) sont estimés selon le système dual, c'est-à-dire sur la base de deux échantillons de personnes: l'échantillon d'Enumeration (E-sample) et l'échantillon de Post-enumeration (P-sample).

Le E-sample est un échantillon parmi les enregistrements du recensement. Le P-sample est un échantillon indépendant du recensement. Le E-sample et le P-sample permettent d'estimer les composantes de sur-couverture, respectivement de sous-couverture, dans l'estimateur dual.

Le plan d'échantillonnage du E-sample a deux niveaux: (1) communes au Tessin, respectivement numéros postaux dans le reste de la Suisse, et (2) personnes. L'allocation du E-sample se base sur le P-sample. Elle permet d'obtenir des tailles similaires dans différents groupes d'intérêt et une faible variabilité des poids dans les strates des unités du premier niveau.

Mots-clé

rapport de méthodes; recensement de la population; VZ2000; RFP2000; enquête de couverture; EC2000; sur-couverture; estimations; couverture; plan d'échantillonnage; plusieurs niveaux; stratification; allocation; post-enumeration; coverage.

Complément d'information: Anne Renaud, tél. 032 713 62 65
Anne.Renaud@bfs.admin.ch

Réalisation: Service de méthodes statistiques, OFS

Diffusion: Office fédéral de la statistique
CH-2010 Neuchâtel
Tél. 032 713 60 60 / Fax 032 713 60 61
Order@bfs.admin.ch

Internet: http://www.statistik.admin.ch

Numéro de commande: 338-0019

Prix: gratuit

Série: Statistique de la Suisse

Domaine: 0 Bases statistiques et produits généraux

Langue du texte original: Français

Graphisme/Layout: OFS

Copyright: OFS, Neuchâtel 2003
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.

ISBN: 3-303-00265-7

Table des matières

1 Introduction **5**

2 L'estimateur dual : recensement, P-sample et E-sample **5**

3 Données **6**

3.1 Données du recensement 6

3.2 Données liées au P-sample 7

4 Méthode d'échantillonnage du E-sample **13**

4.1 Niveau des PSUs 13

4.2 Niveau des personnes 14

5 Cadre de sondage des personnes **14**

5.1 Données du recensement 15

5.2 Distribution des personnes dans les PSUs 17

5.3 Extrait des personnes dans les PSUs 19

6 Allocation des personnes **19**

7 Tirage **20**

8 Pondération **20**

9 Un peu plus sur le E-sample **21**

9.1 Répartition dans les types de communes 21

9.2 Tailles 21

9.3 Poids 22

9.4 Comparaison avec le P-sample et le recensement 22

10 Qualité du cadre de sondage **25**

11 E-sample final **26**

12 Conclusion **26**

1 Introduction

L'estimation de la couverture du recensement fédéral de la population de l'an 2000 (RFP2000) repose sur la méthode duale (dual system method).

Pour déterminer l'estimateur dual, on a besoin des données du recensement dans leur ensemble et de deux échantillons : le E-sample (échantillon d'*Enumeration*) et le P-sample (échantillon de *Post-enumeration*).

Le recensement donne les totaux de base pour le calcul de l'estimateur.

Le E-sample est un échantillon dans les données du recensement. La recherche, sur la base de cet échantillon, d'enregistrements erronés tels des doubles ou des personnes nées après le jour du recensement permet d'estimer la composante de sur-couverture du recensement.

Le P-sample est un échantillon indépendant du recensement. Dans notre cas, il correspond à un sous-ensemble des personnes interrogées lors de l'enquête de couverture (EC2000) déjà réalisée. L'appariement entre le P-sample et le recensement permet d'estimer la composante de sous-couverture du recensement.

Les composantes de sur-couverture et de sous-couverture sont intégrées dans l'estimateur dual afin d'obtenir une estimation du total de la population qui tient compte de ces deux types d'erreur.

Après une introduction sur l'estimateur dual, ce rapport documente le plan d'échantillonnage et le tirage du E-sample. Il fait suite au rapport décrivant le cadre général du projet, ainsi que la procédure et le plan d'échantillonnage de l'EC2000 ; voir Renaud et Eichenberger (2002).

2 L'estimateur dual : recensement, P-sample et E-sample

L'estimateur DSE du total de la population selon la méthode duale peut prendre diverses formes. On choisit celle du bureau du recensement aux États-Unis (US Census Bureau) :

$$DSE = (C - II)\frac{CE}{N_e}\left(\frac{M}{N_p}\right)^{-1} \tag{1}$$

avec C le total recensé, II le nombre d'enregistrements "non définis" (voir ci-dessous) dans le recensement, CE le total pondéré des enregistrements correctement énumérés - non erronés - dans le E-sample, N_e le total pondéré des enregistrements dans le E-sample, M le total pondéré des personnes appariées dans le P-sample et N_p le total pondéré des personnes dans le P-sample.

La notion de "défini" (data defined), respectivement "non défini" (non data defined), est utilisée dans le but de différencier les enregistrements utilisables, respectivement pas utilisables, dans les procédures d'appariement et de recherche des enregistrements erronés. L'idée est d'exclure de la procédure les quelques enregistrements qui n'ont que très peu d'informations afin d'éviter des biais dans les estimations.

Les composantes de sur-couverture et sous-couverture sont inclues dans l'estimateur DSE. Il s'agit respectivement des fractions CE/N_e, basée sur le E-sample, et M/N_p, basée sur le P-sample. Ces deux types d'erreurs peuvent en effet apparaître dans le RFP2000. Les termes C et

II ne sont pas estimés. Ils sont déduits des données du recensement.

Les principes et hypothèses servant de base à l'estimateur dual sont décrits par exemple dans Wolter (1986) pour la méthode et Fienberg (1992) pour une bibliographie dans le cadre de divers recensements. L'estimateur choisit par le US Census Bureau pour le recensement de 2000 est décrit dans Griffin (2000) ou encore Hogan (2000).

Notons que l'estimateur du système dual décrit dans le rapport Renaud et Eichenberger (2002) a été modifié suite aux discussions avec MM. Rajendra Singh et David Whitford du US Census Bureau. La notion de E-sample étendu est mise de côté et le terme II est ajouté à l'estimateur ; en gardant la possibilité de le retirer dans les estimations finales.

La population d'intérêt pour les estimations est la population résidante en Suisse, selon le domicile économique[1], vivant dans un ménage privé et dans l'état au jour du recensement (référence exacte : minuit dans la nuit du 4 au 5 décembre 2002).

3 Données

La base de sondage du E-sample est un sous-ensemble des données des personnes du RFP2000. Le plan d'échantillonnage du E-sample se base sur les données du RFP2000, ainsi que sur des données liées à la construction et au relevé du P-sample.

3.1 Données du recensement

Les données du recensement disponibles au moment du développement du plan d'échantillonnage et du tirage du E-sample sont dans l'état au 6 novembre 2002 pour la liste des personnes. Les caractéristiques des personnes, ménages et bâtiments sont dans l'état de différentes dates entre fin octobre et mi novembre 2002.

Les données ne sont pas définitives. La liste des personnes ne devrait subir que peu de changements mais les caractéristiques et les liens entre les personnes et les ménages ou les bâtiments vont subir des modifications.

Une personne est répétée dans la liste autant de fois qu'elle a de domiciles différents. Une personne avec un domicile secondaire différent de son domicile économique se trouvera donc deux fois dans la liste. Le code de domicile WSKAT permet de différencier les types d'enregistrements et d'extraire des listes de personnes sans doubles.

Seules les personnes avec code de domicile correspondant à des domiciles civils et/ou économiques sont extraits de la base de données (WSKAT $\in \{1, 2, 3, 4\}$; voir Table 1). Les données des personnes au domicile d'appoint et de séjour ne sont pas extraites (WSKAT=5).

Environ 3.36% des personnes sont assignées à des ménages administratifs ("Sammelhaushalte"). Une telle personne est assignée à un bâtiment mais pas à un logement dans ce bâtiment (les ménages n'ont pas pu être formés). De plus, 1.04% des personnes sont assignées à des bâtiments

[1] La notion de domicile principal est également utilisée à l'OFS au lieu de celle de domicile économique. Le terme de domicile économique est utilisé dans les résultats officiels.

TAB. 1 – Code de domicile WSKAT.

WSKAT	Type de domicile
1	un unique domicile (économique et civil)
2	domicile économique et civil
3	domicile économique de séjour
4	domicile d'appoint et civil
5	domicile d'appoint et de séjour

fictifs ("Sammelgebäude") et non pas à des bâtiments réels (dont 71% de ménages administratifs). Il y a un et un seul ménage administratif par bâtiment -vide ou rempli - et un et un seul bâtiment fictif par commune - vide ou rempli.

Dans le cadre du recensement, la localisation géographique d'une personne est donnée par l'adresse du bâtiment dans lequel elle est assignée. Cette adresse peut être différente de l'adresse indiquée par la personne dans le questionnaire. Les personnes sont assignées à des bâtiments fictifs lorsque le lien personne - bâtiment n'est pas réglé. Un numéro postal a été affecté aux bâtiments fictifs mais il n'est pas fiable car différents numéros postaux peuvent coexister sur le territoire d'une commune. Par conséquent, les personnes dans les bâtiments fictifs sont localisées de façon fiable dans une commune politique mais pas dans un numéro postal.

On note également que la liste des numéros postaux (NPA) dans les données du recensement est différente de celle dans la base de sondage de l'EC2000 utilisée pour le P-sample (35 nouveaux et 26 disparus sur un total de env. 3700 en Suisse sans le Tessin) et que certains bâtiments ont une valeur non valide dans les 2 variables permettant de construire le numéro postal à 6 chiffres.

Les données finales du recensement - état au 23 mai 2003 - ne sont pas disponibles au moment de la construction du plan et du tirage du E-sample. Elles sont utilisées pour déterminer le E-sample final et faire une analyse de la qualité de la liste utilisée comme cadre de sondage. Les problèmes potentiels sont les changements ayant une influence sur la définition de la population d'intérêt (ex. types de ménage et de domicile).

3.2 Données liées au P-sample

Le P-sample est l'ensemble des personnes interrogées lors de l'enquête de couverture (EC2000) et faisant partie de la population d'intérêt. Les données du P-sample sont prises comme base pour déterminer l'allocation du E-sample, notamment les poids d'échantillonnage et les tailles par variante de relevé, langue et taille de commune. Les unités primaires du plan d'échantillonnage du P-sample sont également reprises pour la construction du E-sample.

3.2.1 Échantillonnage

Le plan d'échantillonnage de l'EC2000 est un plan à plusieurs niveaux avec deux bases d'échantillonnage indépendantes ; voir Renaud et Eichenberger (2002).

On a tiré 285 parmi les 3699 numéros postaux (NPA) à 6 chiffres du NORD et 20 parmi les 250 communes/secteurs du Tessin. L'échantillon au premier niveau est donc composé de 305 PSUs (Primary Sampling Units) dans 28 strates (variable : stradap).

Les 22 strates du NORD sont construites sur la base de la variante de relevé de la commune majoritaire dans le PSU, la taille du PSU et une mesure de la difficulté de relevé dans le PSU. La commune majoritaire dans le PSU est la commune comportant le plus d'habitants parmi toutes les communes présentes sur le territoire du PSU. Les 6 strates du Tessin sont basées sur les tailles des communes et les problèmes d'adresse des bâtiments.

Deux PSUs du NORD n'ont pas participé à l'EC2000 : Founex (stradap=9, NPA=129700) et Mayens-de-Riddes (stradap=13, NPA=191800). Un ensemble de tournées de facteurs puis de bâtiments ont été tirés dans les 303 PSUs restants. L'échantillon des bâtiments est de taille 15'877.

Un total de 27'407 ménages ont été listés dans les bâtiments de l'échantillon. On considère qu'il s'agit d'une liste exhaustive des ménages dans ces bâtiments. Parmi ceux-ci, 24'024 ménages (88%) ont été contactés. Un ensemble de 21'350 ménages (78%) ont répondu à l'interview de l'EC2000 et sont des ménages privés avec au moins un membre faisant partie de la population d'intérêt (domicile économique, résidant au jour du recensement, pas un double). Ces 21'350 ménages comportent 50'818 personnes, dont 50'070 (98.5%) font partie de la population d'intérêt et définissent le P-sample provisoire - état au 13 septembre 2002 - utilisé pour la construction du E-sample.

Les valeurs de 21'350 ménages et 50'070 personnes sont provisoires. Elles seront modifiées en fonction de nouvelles informations recueillies durant la procédure de match entre le P-sample provisoire et le recensement ; notamment la découverte de doubles dans le P-sample.

3.2.2 Schéma de pondération

La pondération des 50'070 personnes du P-sample provisoire tient compte de la probabilité d'inclusion de la personne dans l'EC2000, d'une correction pour la non-réponse des deux PSUs manquants et d'une correction pour la non-réponse au niveau des ménages. Toutes les personnes sont interviewées dans les ménages de l'échantillon et il n'y a pas de non-réponse au niveau des personnes. Le poids de la personne du P-sample est donc identique au poids de son ménage.

Les données des 147 interviews partielles n'ont pas été fournies par l'institut de sondage. Seul le numéro de la dernière question saisie était disponible. Les réponses partielles sont considérées comme des non-réponses complètes avec indication parfois sur le fait d'être ou pas un vrai ménage. On considère par exemple avoir atteint un vrai ménage si l'information a été saisie pour 2 personnes et que l'interview a été interrompue au milieu de la troisième. Un ensemble de 109 parmi les 147 interviews partielles sont considérées comme des vrais ménages.

Le poids w_p de la personne p dans le ménage q du bâtiment k dans le PSU i de la strate h a la forme :

$$w_p = w_{kih}\, f_{ih}\, g_q \tag{2}$$

avec $w_{kih} = 1/\pi_{kih}$ l'inverse de la probabilité d'inclusion du bâtiment k, f_{ih} l'ajustement pour la non-réponse au niveau des PSUs et g_q l'ajustement pour la non-réponse au niveau des ménages.

Il n'y a pas de correction supplémentaire des poids. En effet, il n'y a aucune information auxiliaire fiable et indépendante du recensement sur le nombre de personnes, ménages ou bâtiments en Suisse.

3.2.3 Ajustement pour la non-réponse au niveau des PSUs

Soient s_h l'échantillon de taille m_h tiré parmi les M_h PSUs de la strate h et $\tau_h \in s_h$ le sous-échantillon de taille r_h formé des PSUs ayant participé à l'enquête.

Deux ajustements f_{ih} sont envisagés, tous deux dépendants uniquement de la strate h dans laquelle se trouve le PSU i. L'ajustement simple $f_{ih}^{(1)}$ fait l'hypothèse que l'ensemble des répondants τ_h est un échantillon aléatoire simple parmi s_h. L'ajustement pondéré $f_{ih}^{(2)}$ correspond à une répartition des poids des non-répondants dans les répondants.

$$f_{ih}^{(1)} = f_h^{(1)} = \frac{\sum_{i \in s_h} 1}{\sum_{i \in \tau_h} 1} = \frac{m_h}{r_h} \tag{3}$$

$$f_{ih}^{(2)} = f_h^{(2)} = \frac{\sum_{i \in s_h} w_{ih}}{\sum_{i \in \tau_h} w_{ih}} \tag{4}$$

avec $w_{ih} = 1/\pi_{ih}$ l'inverse de la probabilité d'inclusion π_{ih} du PSU i de la strates h dans s_h.

Dans le cas de l'EC2000, avec 2 NPA manquant sur les 283 NPA du NORD. On obtient :

$$f_{ih}^{(1)} = \begin{cases} 1.07 & \text{pour les 14 NPA de la stradap 9} \\ 1.33 & \text{pour les 3 NPA de la stradap 13} \\ 1 & \text{pour les 266 NPA restants} \end{cases} \tag{5}$$

et

$$f_{ih}^{(2)} = \begin{cases} 1.03 & \text{pour les 14 NPA de la stradap 9} \\ 1.24 & \text{pour les 3 NPA de la stradap 13} \\ 1 & \text{pour les 266 NPA restants} \end{cases} \tag{6}$$

La différence entre les deux corrections est due au fait que les deux PSUs manquants ont des probabilités d'inclusion inférieures à la moyenne des PSUs de leur strate. La correction $f_{ih}^{(2)}$ est choisie (valeurs plus précises : 1.03186 et 1.24418 respectivement). Elle est moins grande que $f_{ih}^{(1)}$ et permet de conserver la somme des poids au niveau des PSUs.

3.2.4 Ajustement pour la non-réponse au niveau des ménages

La correction pour la non-réponse au niveau des ménages est effectuée dans des groupes de ménages estimés homogènes du point de vue de la probabilité de réponse. Comme au niveau des PSUs, un ajustement simple est comparé à un ajustement pondéré. L'ajustement pondéré est conservé.

Chaque ménage non-répondant peut faire partie ou non de la population d'intérêt. Cette information n'existe cependant pas dans la plupart des cas. La part des ménages hors population

est importante dans l'échantillon (env. 11% parmi les contactés). Il s'agit par exemple d'entreprises, de logements de vacances, de ménages arrivés en Suisse après le jour du recensement ou de ménages collectifs. On ne peut donc pas faire l'hypothèse que tous les non-répondants font partie de la population comme cela est souvent fait lorsque la part des ménages hors population est négligeable. Une estimation d'appartenance à la population d'intérêt est donc intégrée dans la pondération.

3.2.5 Pondération provisoire

La pondération provisoire des personnes a été construite par Jann Potterat (METH) ; voir la Table 2 pour une statistique des poids par strate. Elle sera revue et documentée lorsque le P-sample final sera disponible.

3.2.6 Un peu plus sur les tailles et la variabilité des poids

Selon les résultats du RFP2000, la population résidante en Suisse au 5 décembre 2000 est de 7'288'010, y compris les ménages collectifs. En estimant la proportion des personnes dans des ménages collectifs à 2.3% de la population résidante, le total des poids du P-sample (7'343'140) est globalement 3.1% supérieur au recensement 2000. La différence est explicable par la variabilité d'échantillonnage de l'estimateur du total basé sur le P-sample. Cette variabilité est relativement importante dans le cas du P-sample (coefficient de variation approximatif de 3.3%). Notons que cette faible précision dans l'estimation du total n'est pas a priori dommageable pour l'estimation duale. Le P-sample n'est pas construit dans le but d'estimer un total. Il intervient dans l'estimateur dual sous la forme du rapport M/N_p.

La répartition des personnes du P-sample provisoire dans les variantes de relevé, tailles de communes et langues est donnée dans la Table 3. Ces trois variables étaient utilisées comme référence pour le choix des tailles d'échantillon du P-sample dans l'EC2000.

Les classes de tailles sont : petites (moins de 500 habitants), moyennes (entre 500 et 10'000 habitants) et grandes (plus de 10'000 habitants). La variante, la langue et la classe de taille correspondent aux données finales du recensement et sont inchangées pour les PSUs de l'EC2000. On reprend les informations pour la commune majoritaire déterminée dans la construction de la base de sondage de l'EC2000 ; c'est-à-dire la commune comportant le plus d'habitants parmi toutes les communes présentes sur le territoire du PSU.

Les objectifs de taille concernant le P-sample sont atteints en grande partie. Exemple : au moins 10'000 personnes dans les grandes communes, dans les petites et moyennes communes, et dans la variante TRANSIT. Dans la variante CLASSIC, la taille est légèrement plus petite que l'objectif des 10'000.

Dans la plupart des cas, la variabilité des poids dans des domaines transversaux aux strates - voir Table 4 - est supérieure à celle dans les strates - voir Table 2. Elle est assez importante dans les groupes de langues par exemple. Cette variabilité peut avoir une influence sur la précision des estimateurs.

TAB. 2 – Statistiques de base des poids provisoires des personnes du P-sample. Strates (stradap) 1 à 42 pour le plan NORD et 101 à 106 pour le plan TICINO. Nombre de cas n, moyenne $mean$, coefficient de variation cv, minimum min, médiane P_{50} et maximum max.

stradap	n	$mean$	cv	min	P_{50}	max
1	2211	22.22	8.93	19.23	21.27	36.53
2	2605	25.24	4.26	22.44	25.32	30.31
3	1415	27.85	0.50	27.49	27.90	27.90
5	471	42.15	7.16	40.26	40.90	48.56
6	408	40.02	1.22	38.19	40.15	40.15
7	275	51.62	2.36	48.66	51.73	53.09
8	284	35.60	4.92	34.64	34.64	38.80
9	2409	35.83	7.17	5.19	36.09	46.14
10	299	45.26	0.77	45.11	45.11	46.09
11	1163	37.71	7.13	35.67	37.09	74.63
12	332	37.54	4.18	32.90	38.07	38.07
13	300	62.80	5.83	50.94	63.93	63.93
14	65	29.48	2.22	29.08	29.08	30.53
15	10779	184.12	12.08	89.65	185.61	275.42
16	7920	190.40	3.74	183.28	187.8	216.51
17	6893	195.85	14.72	51.08	195.55	369.94
18	6701	202.37	6.49	195.32	199.15	263.86
19	945	371.85	11.34	339.36	339.36	474.44
20	61	201.25	7.61	191.03	191.03	223.83
21	733	82.70	5.66	81.89	81.89	109.96
41	1835	20.67	5.80	18.21	20.40	26.24
42	421	62.83	1.03	62.68	62.68	65.53
101	211	118.60	28.84	79.26	101.38	214.98
102	96	160.51	18.46	149.37	153.64	322.28
103	119	124.71	17.02	83.03	119.28	171.12
104	553	221.84	7.25	199.87	210.76	248.18
105	164	290.21	16.20	249.74	282.10	378.50
106	402	96.16	4.95	87.37	95.58	100.51
Total	50070	146.66	55.84	5.19	186.87	474.44

TAB. 3 – Nombre de personnes dans le P-sample, selon la variante, la langue et la taille de la commune.

	Langue			Taille commune			
Variante	D	F	I	Pet.	Moy.	Grand	Total
CLASSIC	6172	3100	369	4372	5269		9641
SEMI-CLASSIC	2711	2074	67	915	3937		4852
TRANSIT	23996	9303		135	18157	15007	33299
FUTURE	733				287	446	733
TESSIN			1545	313	544	688	1545
Total	33612	14477	1981	5735	28194	16141	50070

TAB. 4 – Statistiques des poids provisoire des personnes du P-sample, par variante, langue et taille de commune. Nombre de cas n, coefficient de variation cv, minimum min et maximum max.

		n	cv	min	max
Variante	CLASSIC	9641	37.03	18.21	65.53
	SEMI-CLASSIC	4852	18.51	5.19	74.63
	TRANSIT	33299	18.77	51.08	474.44
	FUTURE	733	5.66	81.89	109.96
	TESSIN	1545	41.43	79.26	378.50
Langue	D	33612	53.72	5.19	474.44
	F	14477	59.87	19.23	378.09
	I	1981	62.10	19.23	378.50
Taille commune	Petite	5735	89.92	19.23	322.28
	Moyenne	28194	61.63	5.19	474.44
	Grande	16141	13.98	81.89	263.86

4 Méthode d'échantillonnage du E-sample

Le E-sample est tiré selon un plan d'échantillonnage à deux niveaux.

Le premier niveau reprend les unités primaires (PSUs) définies pour l'EC2000. Il s'agit donc sont des numéros postaux à 6 chiffres (NPA) pour le plan NORD et des communes pour le plan TICINO.

Le deuxième niveau est le niveau des personnes. Contrairement au plan du P-sample, les informations au niveau des personnes sont disponibles dans le recensement. On évite ainsi l'effet de grappe produit par les niveaux tournées et bâtiments utilisé dans la construction du P-sample.

Aucune base objective n'existe pour définir la taille du E-sample. On choisit donc de tirer un échantillon de taille similaire à celui du P-sample, avec une répartition également semblable dans les variantes, classes de taille et langues de la commune majoritaire. Un objectif complémentaire est d'avoir une variabilité des poids d'échantillonnage dans les strates plus faible que dans le P-sample afin de limiter la variance de l'estimateur dual.

Notons que des alternatives au plan à deux niveaux "PSU-personne" ont été envisagées. Ces alternatives, qui utilisaient notamment le niveau bâtiment, n'ont pas été conservées pour des raisons de difficulté pour le calcul des probabilités d'inclusion ou encore de tailles d'échantillon trop grandes pour les ressources à disposition dans le projet. Limiter le nombre d'éléments dans le E-sample permet de mettre plus de ressources dans la recherche des enregistrements erronés et ainsi espérer avoir une meilleure qualité des données pour l'estimateur dual.

4.1 Niveau des PSUs

L'échantillon au premier niveau est composé des PSUs échantillonnés pour le P-sample et qui ont participé à l'EC2000. Il s'agit donc de 303 PSU, dont 283 NPA au NORD et 20 communes/secteurs au Tessin.

Le choix de reprendre les PSUs du P-sample a plusieurs avantages. Premièrement, certaines particularités des PSUs de l'EC2000 sont connues grâce au travail déjà effectué sur le terrain et sur les données du P-sample. Deuxièmement, on peut utiliser l'hypothèse que les personnes du E-sample appariées à des personnes du P-sample sont correctement énumérées et ainsi réduire la recherche des enregistrements erronés. Troisièmement, un premier niveau commun au P-sample et au E-sample peut être utilisé dans l'estimation de la variance de l'estimateur dual par une méthode de reéchantillonnage. Quatrièmement, les domaines de recherche des appariements et des enregistrements erronés peuvent être définis de manière identique ; évitant ainsi des biais dans l'estimation duale ("balancing error", voir par exemple Navarro, 2000).

Cependant, la liste des PSUs dans le recensement n'est pas identique à celle utilisée comme base de sondage pour le premier niveau du P-sample ; voir Section 3.1. Elle nécessite des adaptations afin que chaque personne du recensement faisant partie de la population d'intérêt ait une probabilité non nulle et calculable d'être sélectionnée dans le E-sample. Il s'agit notamment d'éviter qu'une personne dans un PSU existant dans le recensement mais pas dans la base de l'EC2000 ait une probabilité nulle d'être sélectionnée. De même, une personne dans un bâtiment fictif ou sans adresse devrait être assignée à un PSU réaliste.

Pour répondre au problème des listes, on reprend la base de sondage au premier niveau du P-sample pour définir celle du premier niveau du E-sample. Les personnes hors des PSUs de cette liste sont redistribuées dans des PSUs existants. Les personnes concernées se voient donc assignée à un PSU différent pour les besoins de l'échantillonnage. Cet artifice permet de garder les mêmes PSUs que l'EC2000 et d'obtenir des probabilités d'inclusion non nulles et calculables pour chaque personne de la population d'intérêt du recensement.

Les déplacements sont effectués dans des PSUs proches géographiquement ; voir la Section 5. Seuls les PSU touchant la commune de la personne sont papables pour un déplacement. Notons qu'il n'y a aucun déplacement hors des domaines utilisés pour la recherche des enregistrements erronés et des appariements. Ces domaines sont en effet définis par les PSUs et les communes environnantes.

La procédure a pour conséquence que les PSUs et leurs probabilités d'inclusion sont repris tels quels de l'EC2000. Aucun tirage n'est nécessaire.

La probabilité d'inclusion du PSU $i = 1, .., m_h$ de la strate $h = 1, .., H$ est notée π_{ih}. L'échantillon des PSUs est noté $s_{(PSU)}$. Il est de taille m.

Le poids du PSU i ayant participé à l'EC2000 est adapté pour tenir compte de la non-réponse de 2 PSUs : $w_{ih} = f_{ih}^{(2)}/\pi_{ih}$, ; voir la Section 3.2.3 pour la définition de $f_{ih}^{(2)}$.

4.2 Niveau des personnes

Le tirage des personnes est effectué selon un algorithme aléatoire simple sans remise, après avoir distribué les personnes dans les PSUs du premier niveau ; voir ci-dessus et la Section 5.

Soit $U_{(pers,ih)} = \{1, .., p, .., N_{ih}\}$ l'ensemble des personnes dans le PSU $i = 1, .., m_h$ de la strate h. Au deuxième niveau, on tire n_{ih} personnes parmi les N_{ih} personnes du PSU i de la strate h. L'échantillon de personnes (E-sample) est noté $s_{(pers)}$. Il est de taille $n = \sum\sum n_{ih}$.

La probabilité d'inclusion π_p de la personne p du PSU i de la strate h dans $s_{(pers)}$ est donnée par :

$$\pi_p = \frac{\pi_{ih}}{f_{ih}^{(2)}}\,\pi_{p|ih} = \frac{\pi_{ih}}{f_{ih}^{(2)}}\,\frac{n_{ih}}{N_{ih}} \tag{7}$$

avec π_{ih} la probabilité d'inclusion du PSU i de la strate h , $f_{ih}^{(2)}$ la correction pour la non-réponse au niveau des PSUs et $\pi_{p|ih}$ la probabilité d'inclusion de la personne p du PSU i de la strate h, sachant que le PSU i fait partie de l'échantillon $s_{(PSU)}$.

5 Cadre de sondage des personnes

Le cadre de sondage des personnes est un extrait des enregistrements du recensement, avec restriction à la population d'intérêt et aux éléments "définis" (data defined).

5.1 Données du recensement

Les données du recensement comportent 7'458'688 enregistrements avec un type de domicile économique et/ou civil (WSKAT $\in \{1, 2, 3, 4\}$, voir Table 1).

5.1.1 Type de domicile et type de ménage

Le type de domicile est indiqué dans la variable VZ_WSKAT_A (référence à l'adresse A sur le questionnaire) et le type de ménage dans la variable VZ_HH_TYP. On note que la population résidante comporte 7'288'010 personnes (VZ_WSKAT_A $\in \{1, 2, 3\}$), dont 3.5% avec deux domiciles ; voir Table 5. De plus, 3.4% des enregistrements ne sont pas clairement distribués dans les types de ménages privés ou collectifs (VZ_HH_TYP='S').

TAB. 5 – Types de domicile et de ménage dans le recensement.

		Nombre	Répartition
VZ_WSKAT_A	1 (unique)	7'044'285	94.4%
	2 (économique civil)	74'882	1.0%
	3 (économique séjour)	168'843	2.3%
	4 (appoint civil)	170'678	2.3%
VZ_HH_TYP	'P' (privé)	7'026'882	94.21%
	'K' (collectif)	180'871	2.4%
	'S' (administratifs)	250'935	3.4%
Total		7'458'688	100%

5.1.2 Enregistrement défini ou pas défini

Dans le E-sample tout comme dans le P-sample, on exclu les enregistrements dits "non data defined" ("non définis"). Les éléments non data defined du recensement sont en effet traités de manière particulière dans l'estimateur dual.

La particularité pour un enregistrement d'être data defined ou non data defined est définie sur la base d'une variable combinant le nom et le prénom (concaténation), et du nombre de variables manquantes - avant imputation - parmi les variables sexe, date de naissance, état civil et nationalité.

On notera que nous n'avons aucun champ vide dans les données mais que certaines valeurs ont été imputées. On considère que les données sont manquantes uniquement si la valeur avant imputation était manquante, et non pas si la valeur a été modifiée car non cohérente.

La définition du code DATADEF est donnée dans la Table 6. On dit qu'un enregistrement est "défini" si le code DATADEF est égal à 1 ou 2. Il est "non défini" si DATADEF=3.

Le code DATADEF=3 est particulièrement important pour l'estimateur dual et la définition de la population. La différence entre les codes DATADEF=1 et 2 l'est moins. Elle est utilisée pour la sélection des personnes dans le match automatique ou manuel.

TAB. 6 – Code DATADEF.

		Imputation	
		0-2 variables	3-4 variables
Nom	0 lettre	DATADEF=3	
prénom	1 lettre	DATADEF=2	DATADEF=3
	>=2 lettres	DATADEF=1	DATADEF=2

Exemple (valeur avant imputation, manquants notés NULL) :

Nom	prénom	sexe	date naissance	état civil	nationalité	DATADEF
Jean	Muller	1	07.04.45	NULL	NULL	1
Jean	Muller	1	NULL	NULL	NULL	2
J	NULL	1	04.07.45	NULL	NULL	2
NULL	NULL	1	07.04.54	2	1	3
NULL	M	1	NULL	NULL	NULL	3

La variable VZ_DATADEF est répartie comme suit : 7'456'211 cas avec VZ_DATADEF=1 (99.97%), 633 cas avec VZ_DATADEF=2 (0.01%) et 1844 cas avec VZ_DATADEF=3 (0.02%). Seuls 0.02% des enregistrements sont donc non définis.

5.1.3 Population d'intérêt

On extrait les données au domicile économique (VZ_WSKAT_A $\in \{1, 2, 3\}$), dans des ménages privés ou des ménages administratifs (VZ_HH_TYP $\in \{P, S\}$) et data defined (VZ_DATADEF $\in \{1, 2\}$). Ce sous-ensemble, définissant la population d'intérêt, comporte 7'124'434 enregistrements ; voir Table 7.

TAB. 7 – Types de domicile et de ménage dans la population d'intérêt. Voir les définitions dans les Tables précédentes.

VZ_WSKAT_A	VZ_HH_TYP	Nombre	Répartition
1	P	6745802	94.7%
1	S	182888	2.6%
2	P	61596	0.9%
2	S	5630	0.1%
3	P	83897	1.2%
3	S	44621	0.6%
Total		7124434	100%

La population d'intérêt contient seulement 372 enregistrements avec DATADEF=2.

On notera que les personnes assignées aux ménages collectifs - 2.2% des personnes au domicile

économique - sont exclues de la base de sondage. Si, dans les données finales du recensement, certaines d'entre elles sont dans des ménages privés ou administratifs, elles ont échappé à une possibilité de sélection. Heureusement, nous n'avons que très peu de ces cas ; voir la Section 10 pour l'analyse du cadre de sondage.

5.2 Distribution des personnes dans les PSUs

Chaque personne de la base de sondage extraite du recensement doit être assignée à un PSU de la base de sondage du premier niveau du P-sample ; voir la Section 4.1. On différencie le traitement des 6'824'275 personnes du NORD et des 300'159 personnes du TICINO.

Le partage entre le NORD et le TICINO est effectué sur la base de la commune de la personne recensée. Il n'y a donc aucun chevauchement entre les deux régions.

5.2.1 Personnes du NORD

On distingue 4 types de personnes au NORD :

1. la personne est recensée dans un des 3699 NPA de la base du P-sample, n'est pas dans un bâtiment fictif et a un NPA valable (6'770'869 enregistrements) ;
2. la personne est recensée dans un des 3699 NPA de la base du P-sample, dans un bâtiment non fictif mais sans NPA valable (16 enregistrements) ;
3. la personne est recensée dans un des 35 NPA qui n'appartenait pas à la liste pour le P-sample et n'est pas dans un bâtiment fictif (661 enregistrements) ;
4. la personne est recensée dans un bâtiment fictif (52'774 enregistrements).

La commune est considérée correcte pour toutes les personnes de la base. Le NPA des personnes du type 1 est également considéré correct.

Les 53'451 personnes des types 2 à 4 sont redistribués dans les PSUs de la base du plan NORD de l'EC2000 (0.78% de la base de sondage du NORD).

La distribution est basée sur un processus aléatoire. Elle assigne chaque personne à un des NPA touchant la commune dans laquelle elle est recensée.

La probabilité d'être assigné à un NPA dépend du nombre de personnes correctement assignées dans le croisement entre le NPA et la commune. Dans l'exemple de la figure 1, la probabilité π_A qu'une personne de la commune désignée par l'"ovale" soit assignée au NPA A est donnée par $\pi_A = P_A/(P_A + P_B + P_C)$, avec P_A, P_B et P_C le nombre de personnes dans l'intersection de la commune avec les NPA A, B et C, respectivement. Il s'agit d'une attribution proportionnelle à la taille.

Algorithme Soit une commune c contenant un ensemble Γ de P personnes faisant partie de la population d'intérêt. Γ est composé d'un ensemble Γ_{base} de P_{base} personnes correctement assignées et d'un ensemble Γ_{dis} de P_{dis} personnes à répartir dans les NPA touchant la commune. L'algorithme de distribution procède comme suit :

1. numérotation des personnes de Γ_{base} ; après un tri univoque des données. La variable créée est notée NUM et prend des valeurs entre 1 à P_{base} ;

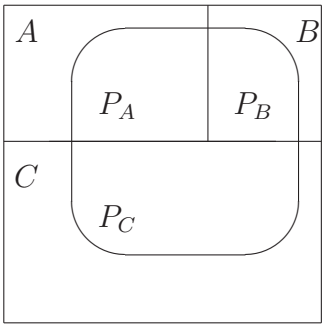

FIG. 1 – Schéma pour la répartition des personnes dans les NPA A, B et C. Le nombre de personnes dans l'intersection de la commune ("ovale") avec les NPA est noté P_A, P_B et P_C, respectivement.

2. génération, pour chacune des P_{dis} personnes à assigner, d'un nombre aléatoire entier RAN entre 1 et P_{base} selon une distribution uniforme ;
3. assignation pour chaque personne de Γ_{dis}, du NPA de la personne de Γ_{base} avec la variable NUM correspondant à la variable RAN.

Notons que l'assignation est réalisée dans un ensemble de 284 NPA. Cet ensemble est formé des 283 NPA du plan NORD auquel est joint un NPA fictif noté N999999 correspondant à l'ensemble de données hors des 283 NPA de l'EC2000.

Au total, 2'651 communes sont touchées par Γ_{dis}. Parmi les 53'406 personnes traitées, 8'738 (16%) sont assignées aux NPA de l'EC2000 et le reste aux NPA hors EC2000.

On note que dix combinaisons "Numéro de commune x NPA" présentes dans les données du recensement au moment de l'extrait des données et touchant des PSUs de l'EC2000 ne sont en fait pas valables. Ces 10 combinaisons contiennent 20 observations de la population d'intérêt (dans 4 NPA). Ni les 10 combinaisons ni les 20 observations n'ont été traitées selon une procédure spéciale dans l'algorithme de distribution. Leur influence est minime et la chance de les retrouver dans le E-sample est très faible.

5.2.2 Personnes du Tessin

Les PSUs de base pour le plan TICINO sont formés de 242 communes politiques et 8 secteurs cadastraux (parties de communes). Les PSUs échantillonnés correspondent à 19 communes politiques entières et 1 secteur à Lugano. Le secteur de Lugano sélectionné contient près de 99% des bâtiments de la commune. De plus, les variables dans le recensement ne permettent pas d'extraire les données au niveau des secteurs cadastraux. On fait donc l'hypothèse que la commune de Lugano a été tirée dans son ensemble lors de l'EC2000. Par conséquent, la base de sondage utilisée pour le plan TICINO est considérée identique à celle présente dans le recensement. La liste des 245 communes au Tessin est identique à celle pour l'EC2000 et aucune redistribution n'est nécessaire.

5.3 Extrait des personnes dans les PSUs

L'ensemble des 1'247'514 personnes de la population d'intérêt, et se trouvant dans les 303 PSUs de l'EC2000 après l'attribution des cas à distribuer, forme la base de sondage pour le niveau des personnes du l'E-sample. Il s'agit de 17.5% de la population d'intérêt, correspondant à 17.2% du NORD et 25.5% du Tessin.

6 Allocation des personnes

Dans chaque PSU de l'EC2000, et premier niveau du plan pour le E-sample, on sélectionne des personnes de la population d'intérêt selon un algorithme aléatoire simple.

Le but est d'obtenir des probabilités d'inclusion peu variables dans chaque strate (stradap) et un nombre de personnes similaire à celui dans le P-sample dans les combinaisons "variante x langue" et "variante x taille de commune". Il n'est pas nécessaire d'augmenter les tailles allouées car il n'y a pas de non-réponse dans le E-sample.

Les variables variante, langue et taille de commune sont reprises de la commune majoritaire dans le PSU d'échantillonnage. On fait donc une approximation par rapport aux informations de la commune effective de la personne recensée. Ce choix n'a pas une grande influence sur les résultats et permet de considérer la même définition que celle utilisée pour le P-sample. Les erreurs dues aux personnes se trouvant dans des PSU d'échantillonnage différents de leur PSU effectif ne sont pas non plus très nombreuses. La question de la définition de la commune de référence sera abordée plus en détails dans la Section 9.

L'allocation est déterminée en trois étapes.

Le première étape permet de déterminer le nombre de personnes par PSU tel que le poids w_p de la personne p du E-sample dans la stradap h soit égal à la médiane des poids des personnes du P-sample dans la même stradap.

On alloue $n_{ih}^{(1)}$ personnes dans le PSU i de la strate h tel que le poids w_p satisfasse l'équation :

$$w_p = w_{ih}\frac{N_{ih}}{n_{ih}^{(1)}} = med(\tilde{w}_h) \tag{8}$$

avec $med(\tilde{w}_h)$ la médiane des poids des personnes du P-sample dans la stradap h (voir la colonne P_{50} dans la Table 2), w_{ih} le poids du PSU i de la strate h (voir Section 4.1) et N_{ih} le nombre de personnes dans le PSU i de la strate h (base de sondage). On a donc :

$$n_{ih}^{(1)} = \frac{w_{ih}}{med(\tilde{w}_h)}\, N_{ih} \tag{9}$$

La deuxième étape adapte $n_{ih}^{(1)}$ de façon à obtenir des tailles par variante correspondant à celles du P-sample. L'idée est d'ajuster l'allocation sur la base des variantes uniquement, et non pas des combinaisons "variante x taille de commune" et "variante x langue". On conserve ainsi les poids constants par stradap (il y a une variante par stradap). On contrôlera tout de même que la répartition de l'échantillon dans les combinaisons "variante x taille de commune" et "variante x langue" est proche de celle du P-sample.

Soient $n_\theta = \sum_{h\in\theta}\sum_i n_{ih}^{(1)}$ le nombre alloué après la première étape dans la variante $\theta \in \{$CLASSIC, SEMI-CLASSIC, TRANSIT, FUTURE et TESSIN$\}$ et $\tilde{n}_\theta$ le nombre correspondant dans le P-sample, voir Table 3 et le rapport Renaud et Eichenberger (2002) pour la liste des stradap par variante. Dans le PSU i de la strate h avec la variante θ, on obtient $n_{ih}^{(2)}$ donné par :

$$n_{ih}^{(2)} = n_{ih}^{(1)}\ \frac{\tilde{n}_\theta}{n_\theta} \tag{10}$$

La troisième étape majore de 10% l'allocation déterminée dans la deuxième étape. Le but est d'obtenir des tailles similaires ou un peu supérieures à celles du P-sample pour les combinaisons "variantes x langue" et "variante x taille de commune".

L'allocation finale $n_{ih}^{(3)}$, majorée par la taille du PSU i de la stradap h, devient :

$$n_{ih}^{(3)} = \min[1.1\ n_{ih}^{(2)}, N_{ih}] = N_{ih}\ \min\left[1.1\ \frac{w_{ih}}{med(\tilde{w}_h)}\ \frac{\tilde{n}_\theta}{n_\theta},\ 1\right] \tag{11}$$

7 Tirage

Le tirage des personnes du E-sample est réalisé avec la procédure SURVEYSELECT de SAS. Un algorithme de choix d'un échantillon aléatoire simple sans remise est appliqué pour la sélection (method=SRS).

L'échantillon contient 55'469 personnes.

Le nombre de personnes par PSU varie de 12 à 815, avec une moyenne de 183.

Trois personnes ont été tirées parmi les 20 dans des combinaisons "NPA x commune" non valables. Cela ne pose cependant pas de problème pour la suite car ces combinaisons n'ont été utilisées que pour la définition de la base de sondage et la probabilité d'inclusion. Les domaines pour la détermination des enregistrements erronés seront basés sur les données finales du recensement.

8 Pondération

La pondération du E-sample est basée sur les poids des PSU - repris du P-sample - ainsi que que sur l'allocation et l'algorithme de tirages décrits dans les Sections 6 et 7. Aucune adaptation complémentaire n'est nécessaire.

Soit une personne p du PSU i de la strate h avec la variante θ. Son poids w_p est défini comme suit :

$$\begin{aligned} w_p &= w_{ih}\frac{N_{ih}}{n_{ih}^{(3)}} = w_{ih}/\min\left[1.1\ \frac{w_{ih}}{med(\tilde{w}_h)}\ \frac{\tilde{n}_\theta}{n_\theta},\ 1\right] \\ &= \max\left[\frac{med(\tilde{w}_h)}{1.1}\ \frac{n_\theta}{\tilde{n}_\theta},\ w_{ih}\right] \end{aligned} \tag{12}$$

avec w_{ih} le poids du PSU i de la strate h, $med(\tilde{w}_h)$ la médiane des poids des personnes du P-sample dans la strate h, $\tilde{n}_\theta$ le nombre de personnes du P-sample dans la variante θ et n_θ le nombre correspondant après la première étape de l'allocation du E-sample. La deuxième ligne de l'équation découle de la propriété suivante : Soient a et $b \in \mathbb{R}$ tels que $ab > 0$, alors $1/\min(a,b) = \max(1/a, 1/b)$.

On note que w_p dépend uniquement de la strate, excepté dans le cas où tous les éléments d'un PSU sont choisis ; c'est-à-dire $n_{ih}^{(3)} = N_{ih}$ et $w_p = w_{ih}$. Les poids sont constants dans les PSUs.

Notons que le jeu final des données du recensement permet de contrôler si tous les éléments du E-sample font partie de la population d'intérêt. Dans le cas contraire, les éléments hors population seront exclus ; voir Section 11. La pondération n'a pas besoin d'être adaptée.

9 Un peu plus sur le E-sample

9.1 Répartition dans les types de communes

La répartition des personnes dans les variantes, classes de taille et langues officielles peut être réalisée de deux manières. La première prend comme référence la commune majoritaire dans le PSU de la personne (lien par le PSU, *i.e.* le NPA pour le NORD) et la deuxième prend comme référence la commune politique où la personne a été recensée (lien par la personne).

Ces deux types de liens sont possibles pour les données du recensement, et donc du E-sample. Par contre, seul le premier lien est possible pour la partie NORD du P-sample car on a une information sur l'adresse des personnes mais pas sur leur commune politique.

La répartition du E-sample selon les deux types varie peu dans les combinaisons "variante x langue" et "variantes x taille de commune" si l'on passe du lien par PSU au lien par personne. Un ensemble de 75 personnes se déplacent des moyennes et grandes communes aux petites communes ; la plupart dans la variante CLASSIC. On note également le passage d'une trentaine de personnes de la variante SEMI-CLASSIC à la variante TRANSIT. Seules 5 personnes d'une unique commune changent de langue si on considère les groupes de langues : allemand + romanche, français et italien. Les différences relatives sont inférieures à 1% dans le tableau "variantes x langue". Elles sont inférieures à 1.5% dans le tableau "variante x taille de commune", avec une exception dans le cas des petites communes TRANSIT (passage de 130 à 150 personnes, i.e. augmentation de 15%).

Le choix du lien par les PSUs utilisé pour définir les valeurs de base de l'allocation du E-sample ne devrait donc pas créer de grosses surprises pour la suite des analyses.

9.2 Tailles

Le nombre de personnes dans le E-sample par variante, langue et taille de commune est supérieur à celui dans le P-sample dans tous les cas excepté 2 (petites communes, variantes TRANSIT et TESSIN) ; voir les Tables 8 et 3. Les objectifs de tailles sont donc presque complètement atteints.

TAB. 8 – Nombre de personnes dans le E-sample, selon la variante, la langue et la taille de la commune (lien par le PSU).

	Langue			Taille commune			
Variante	D	F	I	Pet.	Moy.	Grand	Total
CLASSIC	7005	3570	429	4826	6178		11004
SEMI-CLASSIC	2851	2415	71	950	4387		5337
TRANSIT	26108	10518		130	19218	17278	36626
FUTURE	806				303	503	806
TESSIN			1696	228	463	1005	1696
Total	36770	16503	2196	6134	30549	18786	55469

9.3 Poids

L'allocation choisie produit des poids presque constants dans les strates (stradap) ; voir Table 9. Ils varient plus dans les groupes d'intérêt, voir Table 10. La variabilité dans les groupes est du même ordre de grandeur que celle dans le P-sample, avec des valeurs inférieures et supérieures selon les cas ; voir Table 4.

9.4 Comparaison avec le P-sample et le recensement

La distribution des éléments du E-sample dans les domaines tels que les variantes de relevé, les langues et les tailles de communes peut être comparée à celles dans la population de base du recensement et dans le P-sample ; voir Table 11.

Les intervalles de confiance à 95% ont été calculés avec la procédure SURVEYMEANS de SAS. Les poids d'échantillonnage, ainsi que les grappes et les strates du premier niveau sont intégrées dans le calcul de la variance des estimateurs. Il s'agit d'intervalles approximatifs.

Les intervalles approximatifs du P-sample et du E-sample sur la base du lien par les personnes ($E1$) englobent tous la valeur dans le recensement. La seule exception se trouve dans le cas du E-sample dans les petites communes si on fait le lien par le PSU ($E2$).

On ne voit donc pas de problème majeur dans la répartition des personnes dans les trois variables variante, langue et taille de commune ; que cela soit dans le E-sample ou dans le P-sample.

TAB. 9 – Statistiques de base des poids provisoires des personnes du E-sample. Strates (stradap) 1 à 42 pour le plan NORD et 101 à 106 pour le plan TICINO. Nombre de cas n, moyenne $mean$, coefficient de variation cv, minimum min, médiane P_{50} et maximum max

stradap	n	$mean$	cv	min	P_{50}	max
1	2591	18.17	0.24	17.99	18.17	18.22
2	2931	21.63	0.16	21.54	21.63	21.70
3	1451	23.82	0.32	23.69	23.84	23.97
5	471	34.97	0.22	34.86	35.02	35.04
6	408	34.27	0.30	34.11	34.32	34.35
7	313	44.03	0.47	43.75	44.11	44.25
8	285	28.85	0.44	28.79	28.79	29.13
9	2757	30.06	0.16	29.99	30.09	30.12
10	311	37.61	0.03	37.6	37.6	37.63
11	1353	30.91	0.18	30.76	30.93	30.99
12	354	31.73	0.25	31.66	31.66	31.82
13	225	53.1	0.34	52.80	53.03	53.3
14	52	24.43	0.72	24.34	24.34	24.75
15	10935	165.08	0.17	164.31	165.09	165.70
16	9503	167.01	0.09	166.74	166.99	167.34
17	8077	173.97	0.19	173.16	174.01	174.80
18	7060	177.12	0.11	176.72	177.11	177.54
19	993	301.64	0.16	300.83	301.73	302.22
20	58	169.58	1.30	167.88	167.88	172.36
21	806	78.29	0.10	78.21	78.22	78.39
41	2205	17.44	0.17	17.36	17.44	17.51
42	634	53.55	0.14	53.46	53.56	53.63
101	176	137.72	0.73	135.67	137.44	139.21
102	61	209.37	1.13	206.84	210.05	212.83
103	80	162.5	1.05	160.86	160.86	164.54
104	878	285.50	0.12	285.33	285.33	286.42
105	185	382.64	0.25	381.28	381.98	383.64
106	316	129.66	0.01	129.65	129.65	129.67
Total	55469	130.95	56.12	17.36	165.48	383.64

TAB. 10 – Statistiques des poids des personnes du E-sample, par variante, langue et taille de commune. Nombre de cas n, coefficient de variation cv, minimum min et maximum max.

		n	cv	min	max
Variante	CLASSIC	11004	39.65	17.36	53.63
	SEMI-CLASSIC	5337	15.41	24.34	53.30
	TRANSIT	36626	12.62	164.31	302.22
	FUTURE	806	0.10	78.21	78.39
	TESSIN	1696	34.30	129.65	383.64
Langue	D	36770	51.75	17.36	302.22
	F	16503	60.35	17.43	301.74
	I	2196	61.35	17.99	383.64
Taille	Petite	6134	103.41	17.99	212.83
commune	Moyenne	30549	62.34	17.36	383.64
	Grande	18786	15.99	78.22	285.33

TAB. 11 – Comparaison de la distribution [%] dans le recensement VZ, dans le E-sample (lien par la personne $E1$ et lien par le PSU $E2$) et dans le P-sample P. Les colonnes $IC95(E1)$, $IC95(E2)$ et $IC95(P)$ indiquent un intervalle de confiance approximatif correspondant. (* : allemand sans le romanche.)

		VZ	$E1$	$IC95(E1)$	$E2$	$IC95(E2)$	P	$IC95(P)$
Variante	CLASSIC	3.72	3.83	3.3 - 4.4	3.60	3.3 - 3.9	3.66	3.3 - 4.0
	SEMI	2.44	2.31	2.0 - 2.7	2.33	2.0 - 2.7	2.55	2.2 - 2.9
	TRANSIT	88.79	87.32	83.4 - 91.3	87.52	83.6 - 91.5	89.37	88.3 - 90.4
	FUTURE	0.83	0.86	0.6 - 1.1	0.87	0.62 - 1.12	0.83	0.6 - 1.1
	TESSIN	4.21	5.68	1.5 - 9.9	5.68	1.5 - 9.9	3.60	2.8 - 4.4
Langue	D+R	72.01	66.56	58.7 - 73.4 *	66.56	59.2 - 73.9	68.69	61.7 - 75.7
	F	23.59	27.61	20.7 - 34.6	27.62	20.7 - 34.6	27.55	20.6 - 34.5
	I	4.40	5.83	1.7 - 10.0	5.83	1.7 - 10.0	3.76	3.0 - 4.5
Taille	Petite	3.52	3.02	2.3 - 3.8	2.78	2.2 - 3.4	3.1	2.4 - 3.8
comm.	Moy.	54.57	52.42	47.9 - 56.9	52.65	48.2 - 57.1	55.0	51.2 - 58.9
	Grande	41.92	44.56	40.0 - 49.1	44.57	40.0 - 49.1	41.87	38.0 - 45.8

10 Qualité du cadre de sondage

Le jeu de données disponible pour la construction du plan et le tirage du E-sample est un jeu de données provisoire du recensement ; voir Section 3. Les imperfections et problèmes de couverture des cadres de sondage peuvent induire des biais dans les estimations ; voir par exemple Särndal et al. (1992). Il est donc intéressant de comparer le cadre de sondage utilisé pour le E-sample avec le cadre de sondage que l'on aurait eu avec les données finales du recensement.

Dans cette section, la population d'intérêt est définie par l'ensemble des personnes recensées dans un ménage privé ou administratif et à leur domicile économique ; sans tenir compte du critère "défini" ou pas (data defined ou pas). On ne considère que les défauts de couverture du cadre et non pas les imperfections telles que les erreurs dans les variables de stratification.

Le fichier du recensement disponible pour la construction du plan et le tirage du E-sample comporte 7'458'688 enregistrements, dont 7'125'080 dans la population d'intérêt (*cadre provisoire*). Le fichier final du recensement comporte 7'452'075 enregistrements, dont 7'121'626 dans la population d'intérêt (*cadre final*).

L'intersection entre le cadre provisoire et le cadre final comporte 7'112'186 enregistrements. Les personnes passent dans 1.4% des cas d'un ménage administratif à un ménage privé et dans 0.02% des cas d'un ménage privé à un ménage administratif. Le nombre de personnes dans des ménages administratifs a diminué de moitié entre le cadre provisoire et le cadre final.

Sur-couverture du cadre de sondage Un groupe de 12'894 personnes, *i.e.* 0.18% du cadre provisoire, ne font pas partie du cadre final. Il s'agit donc d'éléments hors population d'intérêt. Parmi ceux-ci, 9'763 sont retrouvés dans le fichier final du recensement mais ne font plus partie de la population et 3'131 ne sont plus retrouvés. Les 9'763 éléments retrouvés ont eu des changements de type de ménage. Ils passent de privé (804), resp. administratif (8959), à collectif. Les éléments disparus sont principalement des cas de doubles domiciles exclus de la base de données car ils étaient localisés dans le même bâtiment.

Sous-couverture du cadre de sondage Un groupe de 9'440 personnes, *i.e.* 0.13% du cadre final, ne font pas partie du cadre provisoire. Il s'agit de personnes n'ayant eu aucune chance d'être tirées dans le E-sample. Parmi ceux-ci, 8'919 sont retrouvés dans la liste provisoire mais ne faisaient pas partie de la population car ils étaient dans des ménages collectifs (6454) et/ou à un domicile non économique (2499). De plus, 521 enregistrements n'existaient pas dans la liste provisoire (478 dans des ménages privés et 43 dans des ménages administratifs).

Effets combinés des défauts du cadre de sondage La sur-couverture du cadre de sondage est très faible. Le biais qu'elle pourrait provoquer est donc négligeable. Les éléments du E-sample se trouvant hors population sont simplement exclus de l'échantillon. La sous-couverture du cadre de sondages est également faible. Le biais potentiel lié à l'exclusion d'une partie de la population d'intérêt est négligeable dans notre cas. Les défauts de couverture du cadre de sondage n'ont donc aucune influence notable sur les estimations basées sur le E-sample.

11 E-sample final

Parmi les 55’469 éléments du E-sample, 21 ne se trouvent plus dans le jeu de données final du recensement. Ils sont exclus de l’échantillon car hors population (sur-couverture du cadre de sondage). La proportion des personnes hors population dans le E-sample est égale à 0.04%, c’est-dire inférieure à la proportion de 0.18% observée dans le cadre de sondage.

Le E-sample final contient 55’448 éléments. Les poids des éléments du E-sample final restent tels que calculés dans la Section 9.

12 Conclusion

Le E-sample est un échantillon d’enregistrements dans le recensement.

La recherche des enregistrements erronés tels que des doubles ou des enregistrements fantaisistes dans le E-sample (sur-couverture), combinée avec la recherche des appariements parmi le P-sample (sous-couverture), permet de faire une estimation de la couverture du recensement selon le système dual.

Le plan d’échantillonnage du E-sample est construit de manière à atteindre un nombre de personnes similaire ou un peu supérieur à celui du P-sample dans des groupes d’intérêt tels que les variantes de relevé du recensement, les régions linguistiques ou encore les tailles des communes. L’allocation choisie conduit à des poids presque constants dans les strates des unités du premier niveau ; avantage notable pour le calcul des variances.

Le E-sample est composé de 55’469 enregistrements parmi les personnes recensées à leur domicile économique, dans un ménage privé ou administratif et ayant suffisamment d’informations pour être considérées comme des personnes réelles ("définies").

Le jeu de donnée final du recensement est légèrement différent de celui utilisé pour le tirage du E-sample. Après exclusion de 21 enregistrements hors population, le E-sample final comporte 55’448 enregistrements. Les objectifs de taille d’échantillon et de limitation de la variabilité des poids sont atteints.

Annexe : Programmes SAS et support informatique

Les travaux décrits dans le présent rapport ont été effectués dans SAS v8.2 sur le réseau PC du recensement et sur le réseau Unix de l'OFS. Les données du recensement ont été anonymisées lorsqu'il s'est révélé nécessaire de les extraire du réseau du recensement.

Les programmes SAS pour la construction du E-sample et les comparaisons avec le P-sample sont les suivants :

- `...\Estimations\ProgSAS\divers.sas`
- `...\Estimations\ProgSAS\Esamp\frameEsamp.sas`
- `...\Estimations\ProgSAS\Esamp\allocEsamp.sas`

Le programme SAS pour le contrôle de la qualité du cadre de sondage est disponible dans :

- `...\DSE\ProgSAS\VzsurUnix\testframeE.sas`

References

Fienberg, S. E. (1992). Bibliography on capture-recapture modelling with application to census undercount adjustment. *Survey methodology*, 18(1) :143–154.

Griffin, R. (2000). Accuracy and coverage evaluation : Dual system estimation. DSSD Census 2000 Procedures and Operation Memorandum Series #Q-20, U.S. Census Bureau.

Hogan, H. (2000). Accuracy and coverage evaluation : Theory and application. 2000 DSE Workshop of the Academy of Science Panel to Review the 2000 Census.

Navarro, A. (2000). Accuracy and coverage evaluation : Targeted extended search plans. DSSD Census 2000 Procedures and Operation Memorandum Series #Q-18.

Renaud, A. and Eichenberger, P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Rapport de méthodes 338-0009, OFS.

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

Wolter, K. M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81(394) :338–346.

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodological reports of the Statistical Methods Unit of SFSO

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

Diffusionsmittel	*Kontakt* *N° à composer*	*Moyen de diffusion*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch	Service de renseignements individuels
Das BFS im Internet	www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch>>News>>Neuerscheinungen.

La Liste des publications mise à jour régulièrement, donne davantage de détails sur les divers moyens de diffusion. Elle se trouve sur Internet à l'adresse www.statistique.admin.ch >>Actualités>>Nouvelles publications.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodological reports of the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte, werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site internet de l'OFS.

Les défauts de couverture du recensement de la population de l'an 2000 (RFP2000) sont estimés selon le système dual, c'est-à-dire sur la base de deux échantillons de personnes: l'échantillon d'Enumeration (E-sample) et l'échantillon de Post-enumeration (P-sample).

Le E-sample est un échantillon parmi les enregistrements du recensement. Le P-sample est un échantillon indépendant du recensement. Le E-sample et le P-sample permettent d'estimer les composantes de sur-couverture, respectivement de sous-couverture, dans l'estimateur dual.

Le plan d'échantillonnage du E-sample a deux niveaux: (1) communes au Tessin, respectivement numéros postaux dans le reste de la Suisse, et (2) personnes. L'allocation du E-sample se base sur le P-sample. Elle permet d'obtenir des tailles similaires dans différents groupes d'intérêt et une faible variabilité des poids dans les strates des unités du premier niveau.

N° de commande: 338-0019
Commandes: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 3-303-00265-7