

Estimation de la couverture du recensement de la population 2000

Méthodes et résultats

Anne Renaud (OFS/METH)

Journées suisses de la statistique 2004

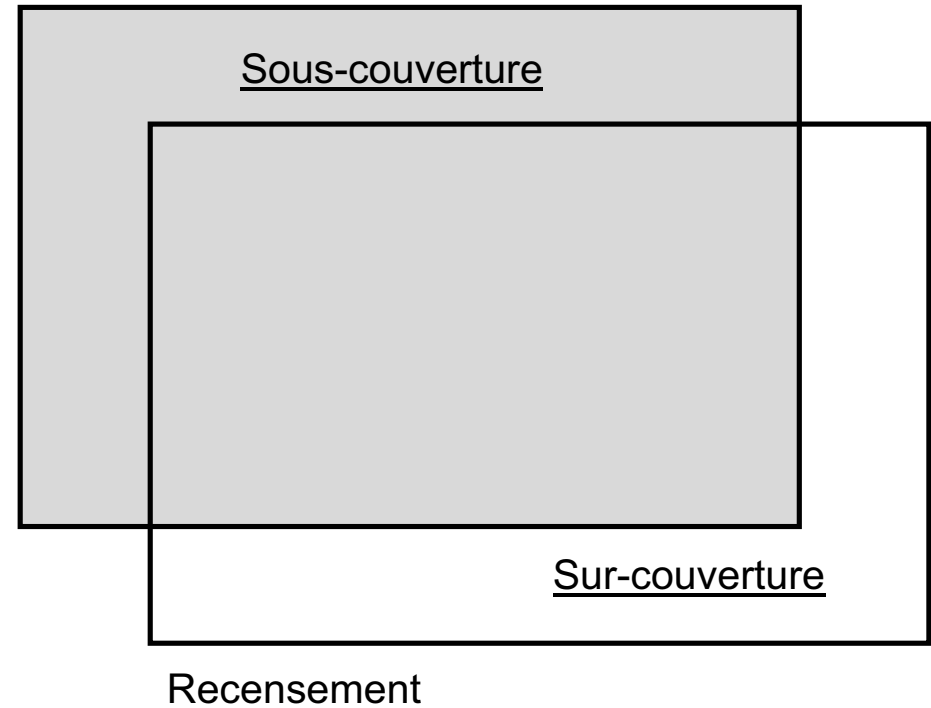
18 novembre 2004

Anne.Renaud@bfs.admin.ch

Présentation

- Introduction
- Sur-couverture
- Sous-couverture
- Couverture nette
- Conclusion

Population à recenser



Introduction

But du projet en Suisse :

- estimations de la sur-couverture, de la sous-couverture et de la couverture nette du RFP2000 ;
- résultats globaux et dans différents sous-groupes de la population ;
- information sur un aspect de la qualité du RFP2000 ;
- information pour la préparation des prochains recensements.

Exemples à l'étranger :

		Sur-couv.	Sous-couv.	Sous-couv. nette
Canada	1996	0.7%	3.2%	2.5%
	2001	0.9%	4.1%	3.2%
USA	1990	3.1%	4.7%	1.6%
	2000	-	-	-0.5%
Australie	1996	0.2%	1.8%	1.6%
	2001	0.9%	2.7%	1.8%

Sous-couverture nette dans des domaines :

Australie (1996) 1.1 - 3.1% selon état/territoire,
 2.0% hommes et 1.1% femmes,
 4.3% pour hommes 20-24 ans.

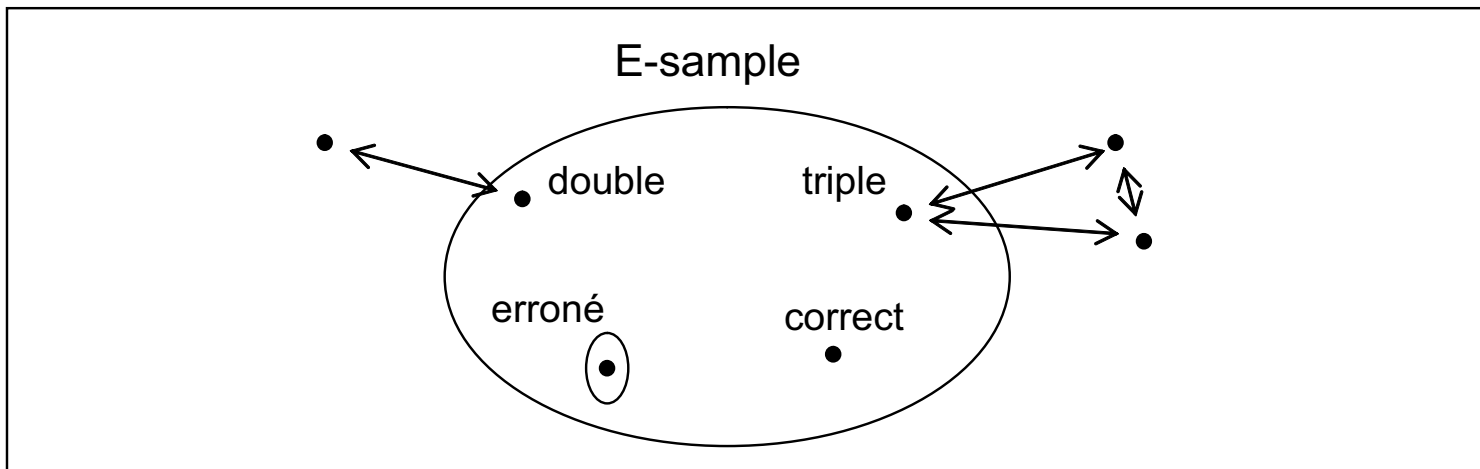
USA (1990) 0.7% White, 5.0% Hispanic, 4.6% Black

USA (2000) -1.1% White, 0.7% Hispanic, 1.8% Black

Sur-couverture : méthode

1. Échantillon dans le recensement : E-sample de taille $n_e \cong 55'000$.
2. Recherche des doubles/triples dans le recensement et des erronés
→ statut de "pertinence" $P_{ce,j}$, $j = 1, \dots, n_e$ (sans interviews).
(correct : 1, erroné : 0, double : 1/2, triple : 1/3)

Recensement



3. Estimation du taux de sur-couverture global R_{sur} :

$$\hat{R}_{sur} = 1 - \hat{R}_{ce} = 1 - \frac{\sum_j w_{e,j} P_{ce,j}}{\sum_j w_{e,j}} \quad (1)$$

avec $w_{e,j}$: poids de la personne j du E-sample.

4. Estimation pour un domaine d :

$$\hat{R}_{sur,d} = 1 - \hat{R}_{ce,d} = 1 - \frac{\sum_j w_{e,j} P_{ce,j} I_{jd}}{\sum_j w_{e,j} I_{jd}} \quad (2)$$

avec $I_{jd} = 1$ si $j \in d$ et $I_{jd} = 0$ sinon.

5. Estimation de la variance : jackknife stratifié.

Sur-couverture : résultats

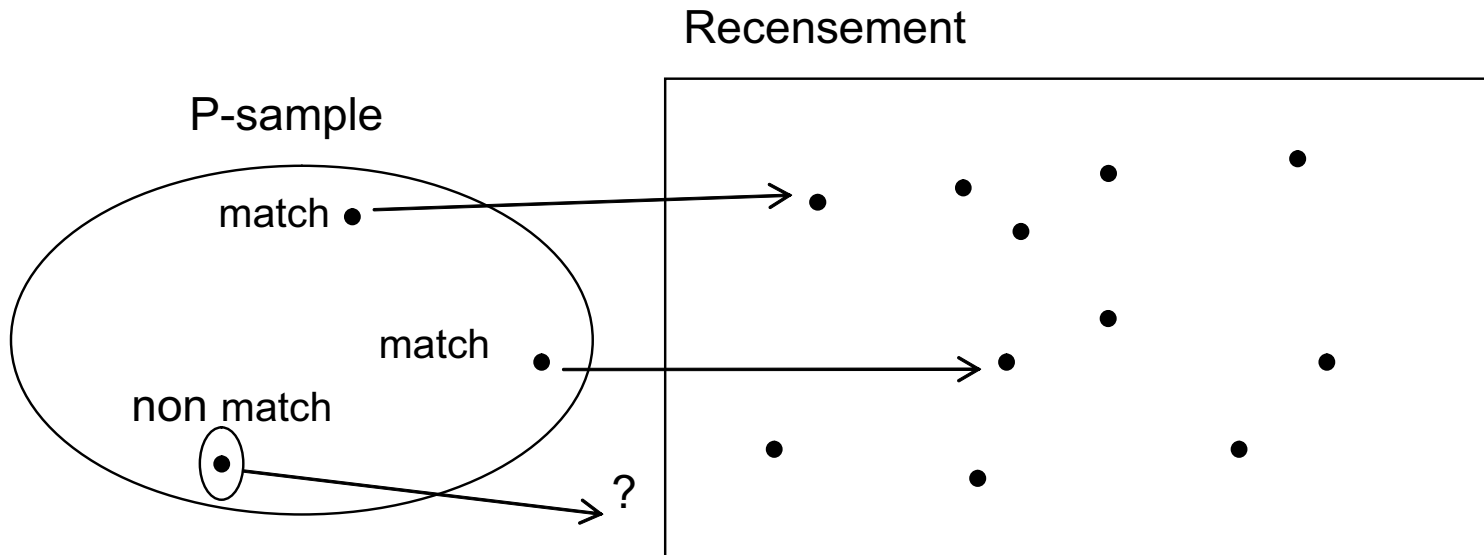
Sur-couverture globale : $\widehat{R}_{sur} = 0.4\%$ (écart-type=0.03%)

Valeurs pour 40 domaines : [0.1% (60-79 ans) ; 1% (20-31 ans)].

→ *faible sur-couverture estimée dans le RFP2000*

Sous-couverture : méthode

1. Relevé pour un échantillon indépendant du recensement (enquête de couverture) : P-sample de taille $n_p \cong 50'000$.
2. Appariement entre le P-sample et le recensement
→ statut de match $P_{m,j}$, $j = 1, \dots, n_p$ (match : 1, non match : 0)



3. Estimation du taux de sous-couverture global R_{sous} :

$$\hat{R}_{sous} = 1 - \hat{R}_m = 1 - \frac{\sum_j w_{p,j} P_{m,j}}{\sum_j w_{p,j}} \quad (3)$$

avec $w_{p,j}$: poids de la personne j du P-sample.

4. Estimation pour un domaine d :

$$\hat{R}_{sous,d} = 1 - \hat{R}_{m,d} = 1 - \frac{\sum_j w_{p,j} P_{m,j} I_{jd}}{\sum_j w_{p,j} I_{jd}} \quad (4)$$

avec $I_{jd} = 1$ si $j \in d$ et $I_{jd} = 0$ sinon.

5. Estimation de la variance : jackknife stratifié.

Sous-couverture : résultats

Sous-couverture globale : $\widehat{R}_{sous} = 1.6\%$ (écart-type=0.11%)

Valeurs pour 40 domaines :

- minimum : 0.9% (s.e.=0.13%) (60-79 ans) ;
- maximum : 8.0% (s.e.=0.85%) (permis B, A,...).
- par groupes d'âges (max 3.5% pour 20-31 ans)
- par états civils (max 2.1% pour célibataires)
- par régions NUTS (max 2.2 - 2.4% pour Lémanique et Tessin)
- non significatif entre variantes de relevé, tailles et types de communes

→ *sous-couverture globale raisonnable en comparaison internationale mais non négligeable dans certains domaines*

Autres résultats sur les cas matchés :

- Différences enquête - recensement : faibles (sexe, état civil, CH - permis B - autres) ou plus grandes (vie active, position dans le ménage, taille du ménage).
- Décalage 5 décembre 2000 - jour relevé effectif : 55% des déménagements autour nouvelle adresse et pas adresse au jour du RFP2000.

(Sous-)couverture nette : méthode

Taux de sous-couverture nette $\hat{R}_{SousNet}$:

$$\hat{R}_{SousNet} = 1 - \hat{R}_{net} = 1 - \frac{C}{\hat{N}} \quad (5)$$

avec C le nombre recensé et \hat{N} le total estimé.

Base de l'estimation de \hat{N} :

$$\hat{N} = C \cdot \widehat{CCF} = C \cdot \hat{R}_{ce} \cdot \hat{R}_m^{-1} = C \frac{\hat{R}_{ce}}{\hat{R}_m} \quad (6)$$

avec \widehat{CCF} le "census correction factor", \hat{R}_{ce} le taux de corrects (E-sample) et \hat{R}_m le taux de match (P-sample).

Application de l'estimation de \widehat{N} dans des cellules homogènes :

$$\widehat{N} = \sum_{\ell=1}^L \widehat{N}_{\ell} = \sum_{\ell=1}^L C_{\ell} \widehat{CCF}_{\ell} = \sum_{\ell=1}^L C_{\ell} \frac{\widehat{R}_{ce,\ell}}{\widehat{R}_{m,\ell}} \quad (7)$$

Estimation globale de $\widehat{R}_{SousNet}$:

$$\widehat{R}_{SousNet} = 1 - \frac{C}{\widehat{N}} = 1 - C \cdot \left[\sum_{\ell=1}^L C_{\ell} \frac{\widehat{R}_{ce,\ell}}{\widehat{R}_{m,\ell}} \right]^{-1} \quad (8)$$

Estimation pour un sous-groupe d (synthétique) :

$$\widehat{R}_{SousNet,d} = 1 - \frac{C_d}{\widehat{N}_d} = 1 - C_d \cdot \left[\sum_{\ell=1}^L C_{\ell,d} \frac{\widehat{R}_{ce,\ell}}{\widehat{R}_{m,\ell}} \right]^{-1} \quad (9)$$

Estimation variance : jackknife stratifié combinant les 2 échantillons

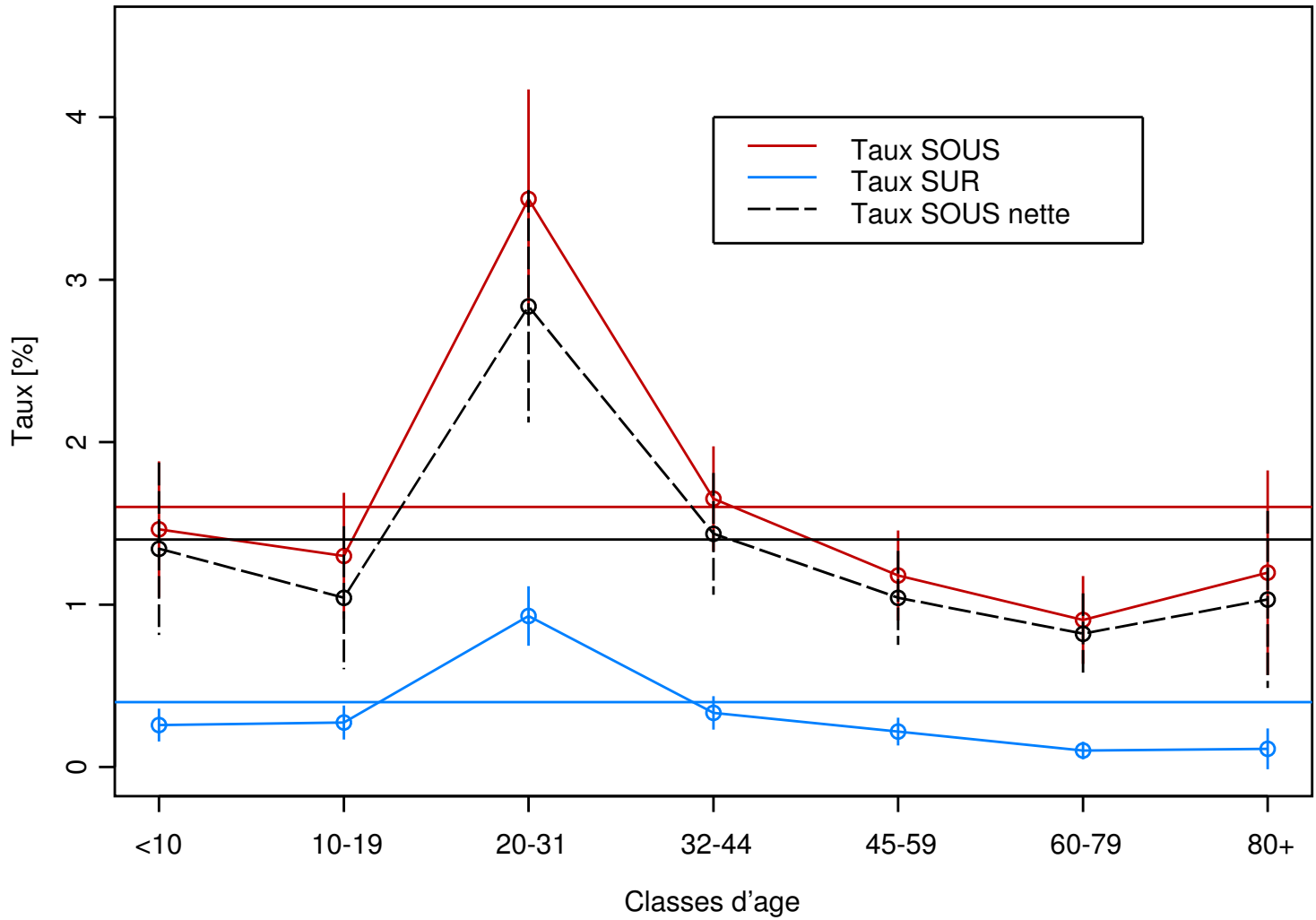
Couverture nette : résultats

Effet combiné sur et sous-couverture : sous-couverture nette

$$\widehat{R}_{SousNet} = 1.4\% \text{ (écart-type}=0.12\%)$$

Valeurs pour 40 domaines :

- minimum : 0.8% (s.e.=0.1%) (60-79 ans et veufs) ;
- maximum : 3.5% (s.e.=0.4%) (permis B, A,..) ;
- par groupes d'âges (max 2.8% pour 20-31 ans) ;
- par états civils (max 1.7% pour célibataires) ;
- par régions NUTS (max 1.8 - 1.6% pour Lémanique et Tessin) ;
- par types de commune (max 1.8% pour centres urbains) ;
- non significatif entre sexes et variantes de relevé.



Conclusion

Couverture RFP2000 raisonnable au niveau international.
Non négligeable dans certains domaines.

Informations annexes : erreurs de mesure potentielles, décalages temporels, etc.

Améliorations possibles côté estimations et côté réduction des défauts de couverture.

Documentation : rapports de méthodes sur site internet de l'OFS.