

Estimation sur petits domaines par scission des poids

Toky RANDRIANASOLO()(**), Yves TILLÉ(**), Jimmy ARMOOGUM(*)*

()Université Paris-Est, IFSTTAR, DEST*

*(**) Université de Neuchâtel*

Résumé

Une nouvelle méthodologie d'estimation sur petits domaines est proposée ; celle-ci se base sur l'utilisation d'une matrice de poids. Chacun des vecteurs de poids permet d'estimer le total des variables d'intérêt dans un domaine. L'intérêt de cette méthode est qu'elle respecte automatiquement la propriété de cohérence des estimations dans les domaines avec les estimations globales.

Lohahevitra

Paika vaovao hentina haminavinana eo anivon'ny faritra kely no atolotra ; paika izay mifototra amy ny fampiasana voho-danja. Voavina amy ny alalan'ny tsilon-danja tsirairay ny tambatry ny ova dinihina eo anivon'ny faritra iray. Ny tombont-soan'ity paika ity dia ny fahazoana fitoviana eo amy ny fitambaran'ny faminavinana isam-paritra sy ny faminavinana ankapobe.

1 Introduction

L'approche classique pour réaliser une estimation sur petits domaines consiste à utiliser un modèle. On utilise généralement un modèle linéaire mixte qui prend en compte les domaines dans la partie aléatoire du modèle. Un des problèmes que nous pouvons rencontrer dans ce type d'approche est qu'elle ne fournit pas nécessairement des estimations locales cohérentes avec des estimations globales. On voudrait en effet que la somme des estimations au niveau de ces petits domaines coïncide avec l'estimation au niveau du domaine global.

Le modèle le plus connu se basant sur les modèles linéaires mixtes est le modèle de [Fay et Herriot \(1979\)](#). En effet, ils partent d'une modélisation d'une fonction de la moyenne dans un domaine quelconque qui va être expliquée d'une part par l'information auxiliaire, et d'autre part par une partie aléatoire expliquant la variabilité entre domaines non traduite par l'information auxiliaire. Plus tard, en utilisant un modèle de régression linéaire à erreurs emboîtées, [You et Rao \(2002\)](#) ont développé une méthode qui fournit des estimations cohérentes grâce à un habile changement de variable dans la régression et à l'emploi des poids d'échantillonnage. En effet, la somme des estimateurs des domaines coïncide bien avec l'estimateur du domaine global.

Nous proposons une nouvelle approche qui consiste à scinder les poids de l'estimateur global. On veut d'emblée respecter la propriété de cohérence. L'idée de l'approche est de créer un poids dépendant à la fois de l'individu et du petit domaine. Chaque individu pouvant ainsi contribuer à tous les domaines. Pour un individu donné, la somme de ses poids relatifs à chaque domaine doit être égale au poids global. Ce poids global peut être l'inverse de la probabilité d'inclusion dans l'échantillon ou un poids résultant d'un calage. De plus, on impose que le système de poids pour un domaine particulier satisfasse des propriétés de calage sur des variables auxiliaires connues pour ce domaine.

On construit ainsi une matrice notée \mathbf{Q} dont le nombre de lignes est égal au nombre d'individus dans l'échantillon et dont le nombre de colonnes est égal au nombre de petits domaines. Cette matrice représente la manière dont le poids global de chaque individu est scindé entre les domaines. Cette matrice est stochastique car la somme de toutes ses lignes vaut 1. Il existe deux types de contributions : celle de l'individu à son propre domaine (que nous appelons "auto-contribution") et celle de l'individu aux autres domaines que le sien (que nous appelons "extra-contribution"). Un individu qui contribue plus aux autres domaines qu'au sien, contribuera moins à son propre domaine et inversement. Notre approche consiste à construire un estimateur composite ayant une partie d' "extra-contribution" construite à l'aide de la matrice de probabilité \mathbf{Q} et une partie "domaine" avec une estimation directe (en utilisant directement les poids globaux pour les unités du domaine). Ces deux parties sont pondérées par un paramètre obtenu en minimisant la dispersion de la variable d'intérêt au niveau de chaque petit domaine de telle sorte que lorsqu'un petit domaine n'est pas représenté dans l'échantillon, la partie "auto-contribution" ne soit pas considérée et qu'inversement lorsque la taille d'un petit domaine est très grande dans l'échantillon, la partie estimée par d'autres unités que celles du domaine soit négligeable.

On testera cette méthodologie sur la base de données sur les communes Suisses du package "sampling" de R, notamment dans l'estimation de la po-

population par tranche d'âge par canton.

2 Notation

Soit une population finie U composée de N individus appartenant à D domaines disjoints notés $\{A_1, \dots, A_d, \dots, A_D\}$. On note également

$$\{N_1, \dots, N_d, \dots, N_D\}$$

les effectifs de chacun des domaines. Par simplicité, les individus sont représentés par leurs numéros. On peut dès lors écrire

$$U = \{1, \dots, k, \dots, N\}.$$

On considère également J caractères auxiliaires notés $x_1, \dots, x_j, \dots, x_J$. Les valeurs prises par le j^e caractère sur les individus de la population sont notées $x_{1j}, \dots, x_{kj}, \dots, x_{Nj}$. Les J caractères du k^e individu sont représentés par le vecteur colonne de \mathbb{R}^J

$$\mathbf{x}_k = \begin{pmatrix} x_{k1} \\ \vdots \\ x_{kj} \\ \vdots \\ x_{kJ} \end{pmatrix}.$$

Les J caractères auxiliaires de toute la population sont alors représentés par la matrice à N lignes et à J colonnes

$$\mathbf{X}_U = \begin{pmatrix} x'_1 \\ \vdots \\ x'_k \\ \vdots \\ x'_N \end{pmatrix}.$$

Les totaux des J caractères pour toute la population sont représentés par le vecteur ligne à J colonnes

$$\mathbf{t}_x = \left(\sum_{k \in U} x_{k1} \quad \dots \quad \sum_{k \in U} x_{kj} \quad \dots \quad \sum_{k \in U} x_{kJ} \right).$$

Les totaux des J caractères pour toute la population appartenant au domaine A_d sont représentés par le vecteur ligne à J colonnes

$$\mathbf{t}_x^d = \left(\sum_{k \in U \cap A_d} x_{k1} \quad \dots \quad \sum_{k \in U \cap A_d} x_{kj} \quad \dots \quad \sum_{k \in U \cap A_d} x_{kJ} \right).$$

Les totaux des J caractères pour toute la population appartenant à chacun des domaines sont représentés par la matrice à D lignes et à J colonnes.

$$\mathbf{t}_x^{\cup A_d} = \begin{pmatrix} \mathbf{t}_x^1 \\ \vdots \\ \mathbf{t}_x^d \\ \vdots \\ \mathbf{t}_x^D \end{pmatrix}.$$

Un échantillon s est un sous-ensemble de U , et un plan de sondage $p(s)$ est une loi de probabilités sur tous les échantillons possibles, telle que

$$p(s) \geq 0, \text{ et } \sum_{s \subset U} p(s) = 1.$$

Un échantillon aléatoire S prend comme valeur un échantillon s selon le plan de sondage $p(s)$, c'est-à-dire $Pr(S = s) = p(s)$ pour tout $s \subset U$. On note $\{n_1, \dots, n_d, \dots, n_D\}$ les effectifs de chaque domaine $\{S \cap A_1, \dots, S \cap A_d, \dots, S \cap A_D\}$. La probabilité d'inclusion d'ordre 1 du k^e individu dans l'échantillon est notée π_k .

3 Estimation directe

L'estimation directe au niveau d'un domaine consiste à construire un estimateur n'utilisant aucune information extérieure au domaine donné. Dans l'estimateur direct, un individu ne contribue donc qu'à son propre domaine. Par exemple, l'estimateur d'Horvitz-Thompson (ou π -estimateur) utilisant directement les poids de sondage $1/\pi_k$ est un estimateur direct. Un estimateur calé est également un estimateur direct.

On appelle "estimateur d'Horvitz-Thompson" ou " π -estimateur" du total de la variable quantitative y noté $t_y = \sum_{k \in U} y_k$, la quantité :

$$\hat{t}_{y,\pi} = \sum_{k \in S} \frac{y_k}{\pi_k}.$$

Au niveau du domaine A_d , l'estimateur d'Horvitz-Thompson du total de la variable quantitative y noté alors $t_y^d = \sum_{k \in U \cap A_d} y_k$ est donné par la quantité :

$$\hat{t}_{y,\pi}^d = \sum_{k \in S \cap A_d} \frac{y_k}{\pi_k}.$$

Les méthodes de calage ont été formalisées par [Deville et Särndal \(1992\)](#). Ils y donnent un cadre général pour l'estimation par calage et ainsi que les

propriétés de ces estimateurs. Soit \mathbf{t}_x un vecteur de totaux connus, l'estimateur par calage du total de la variable quantitative y s'écrit :

$$\hat{t}_{y,w} = \sum_{k \in S} w_k y_k.$$

Les poids w_k dépendent évidemment de l'échantillon S et satisfont à l'équation de calage :

$$\sum_{k \in S} w_k \mathbf{x}'_k = \mathbf{t}_x,$$

où les w_k doivent être proches des poids $\frac{1}{\pi_k}$.

La proximité entre les poids w_k et $1/\pi_k$ est définie en utilisant une pseudo-distance notée G_k supposée définie positive, dérivable et strictement convexe par rapport à w_k . Les poids w_k sont obtenus en minimisant la quantité

$$\sum_{k \in S} \frac{G_k(w_k, d_k)}{q_k}$$

sous la contrainte de l'équation de calage et où les q_k^{-1} sont des coefficients de pondération qui déterminent l'importance de chaque unité dans le calcul de la distance.

Deville et Särndal (1992) ont développé une fonction générale de la pseudo-distance qui dépend d'un paramètre α et qui ne dépend pas de k :

$$\begin{aligned} G_k^\alpha(\omega_k, d_k) &= G^\alpha(\omega_k, d_k) \\ &= \begin{cases} \frac{\omega_k^\alpha}{d_k^{\alpha-1}} + (\alpha - 1)d_k - \alpha\omega_k & \text{si } \alpha \in \mathbb{R} - \{0, 1\} \\ \omega_k \log \frac{\omega_k}{d_k} + d_k - \omega_k & \text{si } \alpha = 1 \\ d_k \log \frac{d_k}{\omega_k} + \omega_k - d_k & \text{si } \alpha = 0 \end{cases} \end{aligned}$$

Nous obtenons plusieurs pseudo-distances selon les valeurs de α . Par la suite, nous choisirons toujours la pseudo-distance pour laquelle $\alpha = 1$, appelée la méthode du raking ratio. Nous prendrons donc toujours une fonction de calage exponentielle. En effet, cette méthode a l'avantage de toujours fournir des poids positifs et de conserver les zéros.

Par cette méthode, les poids w_k s'obtiennent en résolvant en $\boldsymbol{\lambda}$ au moyen de la méthode de Newton le système d'équations (voir par exemple, Tillé, 2001) :

$$\mathbf{t}_x = \sum_{k \in S} \frac{1}{\pi_k} \mathbf{x}'_k \exp(\mathbf{x}_k \boldsymbol{\lambda}),$$

$\boldsymbol{\lambda}$ représentant le vecteur des J multiplicateurs de Lagrange. Finalement,

$$w_k = \frac{1}{\pi_k} \exp(\mathbf{x}_k \boldsymbol{\lambda}).$$

Le problème de l'estimation directe au niveau des petits domaines réside du fait que la variance de l'estimateur obtenu est trop importante. En effet, pour un domaine A_d d'effectif n_d , la variance de son estimateur directe est de l'ordre de $\mathcal{O}(1/n_d^2)$. Cette variance est donc grande quand l'effectif n_d est petit. La qualité des estimations directes au niveau des petits domaines reste donc presque toujours discutable. On tente alors d'améliorer l'estimation dans les domaines en cherchant de l'information au niveau des autres domaines lorsque l'effectif d'un domaine ne nous permet pas d'avoir une estimation directe satisfaisante. La technique que nous utilisons par la suite est la scission des poids.

La méthode développée consiste à scinder les poids en deux parties distinctes. Dans la première partie, seuls les poids des individus d'un domaine A_d peut contribuer à une estimation au niveau de ce même domaine, ce que nous appellerons "partie auto-contribution". Dans la seconde partie, n'importe quel individu peut contribuer avec un certain poids à l'estimation de n'importe quel domaine, ce que nous appellerons "partie extra-contribution". Bien évidemment, plus un individu contribuera à son propre domaine, moins il contribuera aux autres domaines et inversement. En effet, nous voulons respecter d'emblée la propriété de cohérence selon laquelle la somme des estimations au niveau des petits domaines doit coïncider avec une estimation au niveau du domaine global.

4 Estimation par scission des poids ou extra-contribution

4.1 Principe de scission des poids

Considérons d'abord le cas où n'importe quel individu peut contribuer à n'importe quel domaine. L'idée principale de cette approche consiste à construire un poids w_{kd} qui dépend à la fois du k^e individu et du domaine A_d . Ce poids sera défini comme étant le produit du poids w_k obtenu à partir d'un calage global sur le vecteur de totaux connus \mathbf{t}_x avec un nouveau poids q_{kd} tel que

$$w_{kd} = w_k q_{kd}.$$

Comme nous voulons une cohérence entre la somme des estimations au niveau des petits domaines et l'estimation au niveau du domaine global, ce nouveau

poids devra vérifier l'équation de calage sur le vecteur de totaux connus \mathbf{t}_x :

$$\sum_{d=1}^D \underbrace{\sum_{k \in S} w_{kd} \mathbf{x}'_k}_{\mathbf{t}_x^d} = \sum_{k \in S} w_k \mathbf{x}'_k = \mathbf{t}_x$$

ce qui équivaut à

$$\sum_{d=1}^D \sum_{k \in S} w_k q_{kd} \mathbf{x}'_k = \sum_{k \in S} w_k \mathbf{x}'_k = \mathbf{t}_x.$$

De plus, on imposera également les n contraintes

$$\sum_{d=1}^D q_{kd} = 1,$$

qui vont assurer que la somme des estimations locales sont égales aux estimateurs globaux pour n'importe quel variable d'intérêt.

Les poids q_{kd} peuvent être rassemblés dans une matrice à n lignes et à D colonnes notée \mathbf{Q} . La somme de chacune des lignes de cette matrice vaut 1 :

$$\mathbf{Q} = \left(\begin{array}{cccc|c} q_{11} & \cdots & q_{1d} & \cdots & q_{1D} & \sum_d q_{1d} = 1 \\ \vdots & & \vdots & & \vdots & \vdots \\ q_{k1} & \cdots & q_{kd} & \cdots & q_{kD} & \sum_d q_{kd} = 1 \\ \vdots & & \vdots & & \vdots & \vdots \\ q_{n1} & \cdots & q_{nd} & \cdots & q_{nD} & \sum_d q_{nd} = 1 \end{array} \right)$$

Comme chacune des D colonnes de la matrice est obtenue en calant sur le vecteur de totaux \mathbf{t}_x^d de chacun des domaines. Chaque colonne vérifie donc bien l'équation de calage au niveau de chaque domaine :

$$\mathbf{Q}' = \left(\begin{array}{cccc|c} q_{11} & \cdots & q_{k1} & \cdots & q_{n1} & \sum_{k \in S} w_k q_{k1} \mathbf{x}'_k = \mathbf{t}_x^1 \\ \vdots & & \vdots & & \vdots & \vdots \\ q_{1d} & \cdots & q_{kd} & \cdots & q_{nd} & \sum_{k \in S} w_k q_{kd} \mathbf{x}'_k = \mathbf{t}_x^d \\ \vdots & & \vdots & & \vdots & \vdots \\ q_{1D} & \cdots & q_{kD} & \cdots & q_{nD} & \sum_{k \in S} w_k q_{kD} \mathbf{x}'_k = \mathbf{t}_x^D \end{array} \right),$$

soit matriciellement :

$$\left(\begin{array}{cccc} q_{11} & \cdots & q_{k1} & \cdots & q_{n1} \\ \vdots & & \vdots & & \vdots \\ q_{1d} & \cdots & q_{kd} & \cdots & q_{nd} \\ \vdots & & \vdots & & \vdots \\ q_{1D} & \cdots & q_{kD} & \cdots & q_{nD} \end{array} \right) \left(\begin{array}{cccc} w_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & w_k & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & w_n \end{array} \right) \left(\begin{array}{c} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_k \\ \vdots \\ \mathbf{x}'_n \end{array} \right) = \left(\begin{array}{c} \mathbf{t}_x^1 \\ \vdots \\ \mathbf{t}_x^d \\ \vdots \\ \mathbf{t}_x^D \end{array} \right).$$

En résumé, les deux propriétés suivantes doivent être satisfaites :

1. la somme de chacune des lignes de \mathbf{Q} vaut 1,
2. $\mathbf{Q}' \text{diag}(w_1, \dots, w_k, \dots, w_n) \mathbf{X}_S = \mathbf{t}_x^{\cup A_d}$.

Chaque coefficient q_{kd} représente la contribution du k^e individu au domaine A_d . Nous cherchons donc à déterminer la matrice de poids \mathbf{Q} obtenue par un calage matriciel.

4.2 Calcul de la matrice \mathbf{Q}

Pour calculer la matrice \mathbf{Q} qui respecte à la fois la propriété de cohérence et les marges au niveau de chaque domaine, on utilisera un algorithme consistant à répéter deux calages successifs. On cale les colonnes sur les totaux des domaines, ensuite on ajuste les lignes de la matrice pour que chacune des lignes somme à 1. On répètera ces deux opérations jusqu'à l'obtention d'une configuration stable.

Plus précisément, on commence l'algorithme en initialisant la matrice par

$$\mathbf{Q}^{\{0\}} = \begin{pmatrix} \frac{N_1}{N} & \cdots & \frac{N_d}{N} & \cdots & \frac{N_D}{N} \\ \vdots & & \vdots & & \vdots \\ \frac{N_1}{N} & \cdots & \frac{N_d}{N} & \cdots & \frac{N_D}{N} \end{pmatrix}$$

La somme de chacune des lignes de $\mathbf{Q}^{\{0\}}$ vaut 1.

On répète ensuite les deux opérations suivantes :

- Chaque colonne de $\mathbf{Q}^{\{0\}}$ est calée sur le vecteur de totaux connus de chacun des domaines en résolvant pour chaque domaine A_d , $d = 1 \dots D$ le système d'équations :

$$\mathbf{t}_x^d = \sum_{k \in S \cap A_d} w_k q_{kd}^{\{0\}} \mathbf{x}'_k \exp(\mathbf{x}_k \boldsymbol{\lambda})$$

Nous obtenons alors les coefficients d'une nouvelle matrice $\mathbf{Q}^{\{0'\}}$:

$$q_{kd}^{\{0'\}} = q_{kd}^{\{0\}} \exp(\mathbf{x}_k \boldsymbol{\lambda}).$$

- Ensuite on recalcule chacune des lignes sur 1 :

$$q_{kd}^{\{t+1\}} = \frac{q_{kd}^{\{t'\}}}{\sum_{d=1}^D q_{kd}^{\{t'\}}}.$$

L'itération s'arrête lorsque la matrice reste stable, i.e. lorsque la somme des lignes vaut 1 après un calage des colonnes, ou plus précisément quand

$$\sum_{k=1}^n \sum_{d=1}^D |q_{kd}^{\{t'\}}| - n < \varepsilon,$$

où ε est un réel positif suffisamment petit.

Une fois la matrice \mathbf{Q} calculée, on peut alors estimer les totaux des domaines de n'importe quel variable. En effet, au niveau d'un domaine fixé A_d , l'estimateur de type extra-contribution du total de la variable quantitative y s'écrit :

$$\hat{t}_{y,\text{extra}}^d = \sum_{k \in S} w_k q_{kd} y_k.$$

La somme des estimateurs des domaines est toujours égale à l'estimateur dans la population.

5 Estimateur composite

L'estimation composite consiste à mélanger les deux estimateurs pour proposer une méthode où chaque domaine est estimé en partie par une auto-contribution construite à partir d'une estimation directe et en partie par une extra-contribution construite à l'aide de la matrice \mathbf{Q} . Pour ce faire, on commence par construire une matrice qui mélange l'estimateur direct et l'estimateur donné par la matrice \mathbf{Q} .

Pour cela, nous voulons construire une matrice \mathbf{C} de nouveaux poids composites pondérés par des poids α_d pour $d = 1, \dots, D$. On commence par construire une matrice

$$\mathbf{G} = \left(\mathbf{g}_1 \quad \dots \quad \mathbf{g}_d \quad \dots \quad \mathbf{g}_D \right)$$

où chaque colonne \mathbf{g}_d pour $d = 1, \dots, D$ est donné par :

$$\mathbf{g}_d = \begin{pmatrix} g_{1d} \\ \vdots \\ g_{kd} \\ \vdots \\ g_{nd} \end{pmatrix} = \begin{pmatrix} \alpha_d q_{1d} + (1 - \alpha_d) \mathbb{1}_{\{1 \in A_d\}} \\ \vdots \\ \alpha_d q_{kd} + (1 - \alpha_d) \mathbb{1}_{\{k \in A_d\}} \\ \vdots \\ \alpha_d q_{nd} + (1 - \alpha_d) \mathbb{1}_{\{n \in A_d\}} \end{pmatrix}.$$

On constate que les coefficients de pondération α_d dépendent du domaine. Plus le domaine est petit, plus α_d doit être grand. Il est en effet souhaitable que, dans les grands domaines, l'estimateur dépende plus des unités de ce domaine, tandis que, dans les petits domaines, l'estimateur dépende plus d'unités qui n'appartiennent pas au domaine.

Ensuite, la matrice \mathbf{G} est à nouveau calée sur les totaux des domaines et de manière à ce que la somme des lignes soit égale à 1 selon la méthode décrite en Section 4.2. On obtient ainsi la matrice $\mathbf{C} = (c_{kd})$ dont les éléments peuvent s'écrire $c_{kd} = g_{kd} h_{kd}$ où les h_{kd} sont les ajustements résultant du calage.

En considérant une variable d'intérêt quantitative y , l'estimateur composite du total de y au niveau du domaine A_d est construit en pondérant à l'aide du poids α_d :

$$\begin{aligned}\hat{t}_{y,\text{comp}}^d &= \sum_{k \in S} c_{kd} w_k y_k \\ &= \alpha_d \sum_{k \in S} h_{kd} q_{kd} w_k y_k + (1 - \alpha_d) \sum_{k \in S \cap A_d} h_{kd} w_k y_k.\end{aligned}$$

5.1 Choix des poids α_d

Si nous posons $\sigma_{d,1}^2/n$ la variance de la partie extra-contribution de notre estimateur composite, et $\sigma_{d,2}^2/n_d$ la variance de la partie domaine, la variance de notre estimateur composite est alors proportionnelle à la quantité suivante :

$$\begin{aligned}\text{Var}(\hat{t}_{y,\text{comp}}^d) &\propto \alpha_d^2 \frac{\sigma_{d,1}^2}{n} + (1 - \alpha_d)^2 \frac{\sigma_{d,2}^2}{n_d} + 2\alpha_d(1 - \alpha_d) \frac{\sigma_{d,12}}{n_d n} \\ &\propto \alpha_d^2 \frac{\sigma_{d,1}^2}{n} + (1 - \alpha_d)^2 \frac{\sigma_{d,2}^2}{n_d},\end{aligned}$$

le terme associé à la covariance étant supposé négligeable (Rao, 2003).

Si nous minimisons cette quantité en fonction de α_d , nous dérivons :

$$\alpha_d^2 \frac{\sigma_{d,1}^2}{n} + (1 - \alpha_d)^2 \frac{\sigma_{d,2}^2}{n_d} \simeq 0$$

Ce qui nous donne un poids optimal dépendant de n_d :

$$\alpha_d(n_d) \simeq \frac{1}{1 + \frac{\sigma_{d,1}^2}{\sigma_{d,2}^2} \frac{n_d}{n}}.$$

Nous notons ϑ_d la valeur du quotient $\frac{\sigma_{d,1}^2}{\sigma_{d,2}^2}$. Nous voyons bien que si n_d tend vers 0 (resp. $+\infty$), alors $\alpha_d(n_d)$ tend vers 1 (resp. 0).

6 Calcul de précision

Nous voulons tester la performance de la méthodologie avancée. Nous allons donc calculer les racines relatives des erreurs moyennes quadratiques (ou RRMSE) de l'estimateur composite par ré-échantillonnage.

Définition 6.1. Racine relative de l'erreur quadratique moyenne. La "racine relative de l'erreur quadratique moyenne" (RRMSE) de l'estimateur \hat{t} d'un total t est donnée par

$$\text{RRMSE}(\hat{t}) = \frac{\sqrt{\text{b}(\hat{t})^2 + \text{Var}(\hat{t})}}{t}$$

où $\text{b}(\hat{t})$ est le biais de l'estimateur \hat{t} .

Chaque échantillonnage se fait en tirant un échantillon de taille d'espérance égale à n avec des probabilités inégales. Nous obtenons ensuite pour chaque échantillonnage une quantité

$$\varepsilon = \frac{\hat{t} - t}{t} \times 100$$

où \hat{t} représente le total estimé et t le vrai total.

Proposition 6.2. Le biais $\text{b}(\hat{t})$ du total t peut s'écrire en une certaine fonction f de ε i.e.

$$\text{b}(\hat{t}) = f(\varepsilon).$$

Démonstration. Par définition, le biais de l'estimateur \hat{t} du total t est donné par

$$\begin{aligned} \text{b}(\hat{t}) &= \mathbb{E}(\hat{t}) - t \\ &= \mathbb{E}(\hat{t} - t) \\ &= \frac{t}{100} \times \mathbb{E}(\varepsilon). \end{aligned}$$

Nous avons donc $f(\varepsilon) = \frac{t}{100} \times \mathbb{E}(\varepsilon)$. □

Proposition 6.3. La variance de l'estimateur \hat{t} du total t peut s'écrire en une certaine fonction g de ε i.e.

$$\text{Var}(\hat{t}) = g(\varepsilon).$$

Démonstration. La variance de l'estimateur \hat{t} est donnée par

$$\begin{aligned} \text{Var}(\hat{t}) &= \text{Var}(\hat{t} - t) \\ &= \left(\frac{t}{100}\right)^2 \times \text{Var}(\varepsilon). \end{aligned}$$

Nous avons donc $g(\varepsilon) = \left(\frac{t}{100}\right)^2 \times \text{Var}(\varepsilon)$. □

Corollaire 6.4. *Le $\text{RRMSE}(\hat{t})$ de l'estimateur \hat{t} du total t peut s'écrire en une certaine fonction h de ε i.e.*

$$\text{RRMSE}(\hat{t}) = h(\varepsilon).$$

Démonstration. Le biais et la variance de l'estimateur \hat{t} pouvant s'exprimer comme étant des fonctions de ε , nous avons alors :

$$\begin{aligned}\text{RRMSE}(\hat{t}) &= \frac{1}{100} \sqrt{\mathbb{E}(\varepsilon)^2 + \text{Var}(\varepsilon)} \\ &= \frac{1}{100} \sqrt{\mathbb{E}(\varepsilon^2)}.\end{aligned}$$

□

7 Application sur la population Suisse par tranche d'âge par canton

7.1 Données sur la population au niveau des communes Suisses

Dans le package “sampling” de R, nous disposons d'information sur la population Suisse au niveau des municipalités provenant du recensement de 2003. Nous disposons de 2896 communes comme unités d'observation suivant 22 variables :

Variables	Libellés
CT	Canton
REG	Région
COM	Numéro de la commune
Nom	Nom de la commune
HApoly	Surface de la commune
Surfacesbois	Surface boisée
Surfacescult	Surface cultivée
Alp	Surface de l'alpage
Airbat	Surface avec des immeubles
Airind	Surface industrielle
P00BMTOT	Nombre d'hommes
P00BWTOT	Nombre de femmes
Pop020	Nombre d'hommes et de femmes âgés entre 0 et 19 ans
Pop2040	Nombre d'hommes et de femmes âgés entre 20 et 39 ans
Pop4065	Nombre d'hommes et de femmes âgés entre 40 et 64 ans
Pop65P	Nombre d'hommes et de femmes de plus de 65 ans
H00PTOT	Nombre de ménages
H00P01	Nombre de ménages avec 1 personne
H00P02	Nombre de ménages avec 2 personnes
H00P03	Nombre de ménages avec 3 personnes
H00P04	Nombre de ménages avec 4 personnes
POPTOT	Population totale

Afin de tester notre méthodologie, nous considérerons comme variables auxiliaires : le nombre de communes par canton, **POPTOT**, **H00PTOT**, **Pop020+Pop2040**. Nous considérerons comme variables d'intérêt : **Pop020**, **Pop2040**, **Pop4065**, **Pop65P**.

7.2 Principe du ré-échantillonnage

Chaque échantillonnage se fait suivant un tirage poissonnien. Les probabilités d'inclusion sont considérées comme étant proportionnelles à la population totale de chaque municipalité.

7.3 Calage de matrice

Le calage sur la matrice se fait simultanément sur les lignes et les colonnes. La somme des éléments sur chaque ligne doit être égale à 1. Les éléments de chaque colonne doivent être obtenus par calage sur chaque domaine.

7.3.1 Algorithme du calage de matrice

- 1: **Initialisation** des coefficients de la matrice
- 2: Fixer un seuil d'arrêt et le nombre maximal d'itérations
- 3: **while** conditions d'arrêt non respectées **do**
- 4: **for** $d = 1$ **to** D **do**
- 5: Faire les calages sur les totaux des caractères connus au niveau du d^e petit domaine
- 6: **end for**
- 7: Mettre la contrainte sur les lignes : la somme des lignes doit être égale à 1
- 8: **end while**
- 9: **return** Q

7.4 Algorithme d'implémentation de la méthode

- 1: Fixer le nombre maximal d'itérations
- 2: **while** nombre maximal d'itérations non atteint **do**
- 3: Fixer la taille de l'échantillon
- 4: Calculer les probabilités d'inclusion pour chaque commune
- 5: Effectuer l'échantillonnage poissonnien
- 6: Faire un premier calage sur les poids d'inclusion au niveau du domaine global
- 7: Initialisation de la matrice Q
- 8: Calage de la matrice Q
- 9: Fixer les valeurs des $\vartheta = (\vartheta_1, \dots, \vartheta_d, \dots, \vartheta_D)$
- 10: Calcul des poids $\alpha = (\alpha_1, \dots, \alpha_d, \dots, \alpha_D)$
- 11: Initialisation de la matrice composite
- 12: Calage de la matrice composite
- 13: Calcul de la quantité ε

- 14: **end while**
- 15: Calcul du RRMSE

7.5 Résultats

Afin de calculer les RRMSEs, nous faisons 100 simulations pour des valeurs des constantes $\vartheta = (\vartheta_1, \dots, \vartheta_d, \dots, \vartheta_D)$ égales à $0 \times \mathbf{1}_D, 1 \times \mathbf{1}_D, \dots, 10 \times \mathbf{1}_D$ où $\mathbf{1}_D$ est un vecteur unité d'ordre D , pour un taux de sondage à 8%.

Nous remarquons que les RRMSEs diminuent lorsque les ϑ augmentent. Nous prenons alors des valeurs de ϑ très grandes égales à $10 \times \mathbf{1}_D, 20 \times \mathbf{1}_D, \dots, 100 \times \mathbf{1}_D$.

Pour chaque domaine, les RRMSEs en fonction de ϑ_d ont une certaine forme parabolique admettant un minimum. Il est intéressant de voir à quoi correspond la valeur de ϑ_d avec laquelle le minimum des RRMSEs est atteint.

0	1	2	3	4	5	6	7	8	9	10	ϑ	10	20	30	40	50	60	70	80	90	100
RRMSE pour la variable Pop020											RRMSE pour la variable Pop020										
0.65	0.60	0.57	0.53	0.53	0.45	0.47	0.47	0.43	0.43	0.44	ZH	0.44	0.41	0.39	0.46	0.44	0.42	9.59	11.28	16.19	11.58
0.93	0.78	0.78	0.74	0.70	0.70	0.67	0.68	0.81	0.64	0.65	BE	0.70	0.69	0.75	0.63	0.76	0.63	13.66	13.47	18.99	13.65
0.79	0.79	0.92	0.70	0.84	0.76	0.84	0.80	0.82	0.84	0.73	LU	0.69	0.77	0.88	0.83	0.83	0.85	14.17	11.34	17.78	14.15
1.50	1.62	1.67	1.69	1.66	1.45	1.64	1.45	1.38	1.61	1.46	UR	1.48	1.58	1.57	1.59	1.68	1.72	17.40	15.60	23.68	17.39
2.18	1.99	2.15	1.95	2.01	1.94	1.97	1.89	1.77	1.80	1.74	SZ	1.67	1.63	1.46	1.46	1.41	1.13	17.37	15.20	17.33	17.35
0.48	0.52	0.62	0.53	0.56	0.54	0.53	0.56	0.58	0.61	0.59	OW-NW	0.53	0.67	0.82	0.87	1.01	0.82	17.34	15.54	22.98	17.32
2.77	2.57	2.66	2.65	2.59	2.55	2.50	2.58	2.78	2.47	2.57	GL	2.57	2.51	2.36	2.35	2.24	2.17	17.48	17.16	23.84	17.42
6.83	6.64	6.83	6.70	6.67	6.26	6.29	6.10	6.15	5.93	5.82	ZG	6.14	5.35	4.85	4.65	4.50	4.16	17.80	15.04	23.75	17.68
2.28	2.42	2.28	2.13	2.20	2.04	2.34	2.03	1.82	2.04	1.83	FR	1.93	1.92	1.93	1.90	1.72	1.87	15.48	11.97	19.87	15.50
1.13	1.08	1.10	1.08	1.00	0.98	0.92	0.99	1.11	0.93	0.95	SO	0.98	1.05	0.89	1.02	0.88	1.09	17.35	17.12	24.29	17.29
0.50	0.47	0.46	0.46	0.45	0.48	0.44	0.57	0.51	0.44	0.53	BS-BL	0.50	0.61	0.55	0.60	0.62	0.73	11.87	9.70	11.85	11.84
3.06	2.92	2.99	2.93	2.90	2.91	2.80	2.80	2.95	2.70	2.73	SH	2.82	2.72	2.36	2.58	2.31	2.30	10.33	10.04	14.24	10.25
5.90	5.82	5.70	5.64	5.49	5.50	5.35	5.44	5.41	5.19	5.10	AR-AI	5.08	4.73	4.68	4.49	4.05	4.27	17.71	16.66	23.87	17.73
2.89	2.85	2.54	2.48	2.36	2.30	2.16	2.14	2.20	2.13	1.99	SG	2.04	1.64	1.52	1.37	1.29	1.42	12.47	13.56	20.25	15.07
2.30	2.42	2.22	2.05	2.15	2.08	2.46	2.24	2.03	2.26	2.08	GR	2.03	2.25	2.63	2.54	2.68	2.66	14.85	13.88	20.44	14.67
0.71	0.73	0.71	0.68	0.68	0.68	0.67	0.61	0.63	0.67	0.60	AG	0.66	0.70	0.74	0.78	0.74	0.72	17.34	15.91	23.94	17.30
4.37	4.19	4.01	3.86	3.80	3.64	3.54	3.41	3.58	3.43	3.23	TG	3.22	2.67	2.40	2.14	2.21	2.13	17.44	14.27	23.67	17.39
7.10	6.85	6.35	6.06	5.90	5.57	5.61	5.25	5.10	4.98	4.98	TI	5.00	4.34	3.79	3.53	3.60	3.13	17.41	17.54	24.55	17.57
1.97	1.71	1.72	1.63	1.51	1.55	1.35	1.40	1.50	1.29	1.21	VD	1.22	1.12	1.13	1.04	1.02	1.13	14.54	13.32	20.13	14.27
3.60	3.59	3.47	3.34	3.25	3.09	3.20	2.97	2.69	3.09	2.79	VS	2.71	2.66	2.45	2.36	2.15	2.33	15.70	14.58	21.33	15.49
1.77	1.71	1.70	1.69	1.66	1.62	1.57	1.62	1.67	1.49	1.49	NE	1.50	1.41	1.42	1.35	1.31	1.35	10.77	10.59	14.84	10.72
0.30	0.36	0.37	0.33	0.37	0.38	0.34	0.36	0.42	0.41	0.43	GE	0.37	0.46	0.54	0.58	0.55	0.68	9.68	7.91	9.58	
3.01	2.66	2.78	2.87	2.75	2.94	2.47	2.68	2.89	2.50	2.42	JU	2.58	2.47	2.42	2.46	2.42	2.25	17.47	16.64	23.84	17.32

0	1	2	3	4	5	6	7	8	9	10	ϑ	10	20	30	40	50	60	70	80	90	100
RRMSE pour la variable Pop2040											RRMSE pour la variable Pop2040										
0.42	0.39	0.37	0.35	0.34	0.29	0.31	0.31	0.28	0.28	0.29	ZH	0.29	0.27	0.25	0.30	0.28	0.27	8.05	9.48	13.60	9.68
0.73	0.61	0.61	0.58	0.55	0.55	0.53	0.53	0.63	0.50	0.51	BE	0.54	0.54	0.59	0.49	0.59	0.50	12.12	11.97	16.90	12.11
0.67	0.67	0.78	0.59	0.71	0.65	0.71	0.67	0.69	0.71	0.62	LU	0.58	0.65	0.74	0.70	0.71	0.72	12.90	10.39	16.33	12.87
1.36	1.47	1.52	1.54	1.52	1.32	1.49	1.32	1.25	1.47	1.32	UR	1.34	1.43	1.42	1.45	1.53	1.57	17.39	15.65	23.80	17.38
1.86	1.70	1.83	1.67	1.72	1.66	1.68	1.61	1.51	1.54	1.48	SZ	1.43	1.39	1.25	1.25	1.20	0.97	17.36	15.29	17.32	17.33
0.42	0.46	0.54	0.46	0.49	0.47	0.46	0.49	0.50	0.54	0.52	OW-NW	0.46	0.57	0.72	0.76	0.88	0.72	17.34	15.59	23.05	17.30
2.59	2.41	2.48	2.47	2.41	2.39	2.34	2.41	2.59	2.31	2.40	GL	2.40	2.35	2.21	2.20	2.09	2.03	17.46	17.15	23.85	17.40
4.96	4.83	4.84	4.68	4.56	4.55	4.57	4.43	4.34	4.31	4.23	ZG	4.26	3.92	3.50	3.38	3.18	3.03	17.57	14.88	23.69	17.49
1.89	2.01	1.90	1.78	1.84	1.69	1.94	1.69	1.49	1.69	1.52	FR	1.63	1.59	1.60	1.58	1.43	1.55	14.31	11.17	18.61	14.32
0.96	0.91	0.93	0.91	0.85	0.83	0.78	0.84	0.93	0.79	0.81	SO	0.82	0.89	0.76	0.87	0.74	0.92	17.34	17.13	24.33	17.28
0.36	0.34	0.33	0.33	0.32	0.34	0.31	0.41	0.36	0.31	0.38	BS-BL	0.35	0.43	0.39	0.43	0.45	0.52	10.37	8.47	10.35	10.34
2.64	2.52	2.57	2.52	2.49	2.51	2.41	2.41	2.53	2.32	2.35	SH	2.43	2.34	2.03	2.22	1.99	1.98	9.16	8.90	12.63	9.08
6.00	5.92	5.79	5.73	5.59	5.59	5.44	5.53	5.50	5.27	5.19	AR-AI	5.16	4.81	4.76	4.57	4.12	4.34	17.72	16.63	23.97	17.74
2.54	2.50	2.23	2.18	2.07	2.02	1.90	1.88	1.93	1.87	1.74	SG	1.79	1.44	1.34	1.20	1.14	1.25	11.60	12.71	18.98	13.91
1.86	1.96	1.81	1.67	1.75	1.69	2.00	1.82	1.64	1.83	1.69	GR	1.68	1.82	2.13	2.06	2.17	2.15	14.19	13.54	19.52	14.07
0.59	0.61	0.59	0.57	0.57	0.57	0.56	0.51	0.52	0.56	0.50	AG	0.55	0.59	0.61	0.65	0.62	0.61	17.33	15.96	24.04	17.29
4.12	3.95	3.78	3.63	3.58	3.42	3.33	3.21	3.37	3.23	3.04	TG	3.03	2.51	2.26	2.01	2.08	2.00	17.43	14.26	23.72	17.37
4.97	4.80	4.45	4.25	4.15	3.90	3.93	3.68	3.57	3.49	3.49	TI	3.52	3.04	2.66	2.47	2.52	2.19	17.25	17.40	24.44	17.42
1.53	1.34	1.34	1.26	1.17	1.20	1.05	1.09	1.15	1.00	0.94	VD	0.95	0.87	0.88	0.81	0.79	0.88	13.35	12.22	18.52	13.04
3.06	3.05	2.95	2.84	2.77	2.63	2.72	2.53	2.29	2.62	2.37	VS	2.31	2.26	2.08	2.01	1.83	1.98	15.64	14.46	21.25	15.43
1.40	1.35	1.34	1.33	1.31	1.28	1.24	1.28	1.31	1.18	1.18	NE	1.18	1.11	1.12	1.07	1.03	1.06	9.62	9.46	13.27	9.57
0.21	0.26	0.26	0.23	0.26	0.27	0.25	0.26	0.30	0.29	0.30	GE	0.27	0.33	0.38	0.41	0.39	0.49	7.91	6.47	7.81	
2.85	2.52	2.62	2.71	2.58	2.79	2.34	2.54	2.70	2.37	2.29	JU	2.43	2.33	2.29	2.33	2.28	2.13	17.46	16.69	23.93	17.34

0	1	2	3	4	5	6	7	8	9	10	ϑ	10	20	30	40	50	60	70	80	90	100
RRMSE pour la variable Pop4065											RRMSE pour la variable Pop4065										
2.98	2.70	2.49	2.36	2.23	2.08	1.97	1.88	1.85	1.77	1.64	ZH	1.71	1.40	1.13	1.09	0.98	1.01	9.18	10.86	15.54	11.06
1.37	1.20	0.95	0.87	0.97	0.84	0.85	0.92	0.76	0.83	0.83	BE	0.81	0.68	0.85	0.85	0.85	0.85	12.89	12.74	17.98	12.87
1.78	1.64	1.58	1.45	1.44	1.38	1.44	1.46	1.26	1.20	1.27	LU	1.27	1.21	1.37	1.22	1.19	1.34	13.03	10.65	16.57	13.02
6.27	6.16	5.93	5.75	5.96	5.63	5.75	5.74	5.60	5.56	5.39	UR	5.53	5.01	5.05	4.91	4.75	4.77	17.94	16.41	24.25	17.86
1.07	0.96	0.98	0.87	0.82	0.75	0.88	0.85	0.83	0.77	0.89	SZ	0.85	1.01	1.15	1.17	1.11	1.44	17.37	15.46	17.36	17.37
1.39	1.28	1.24	1.13	1.08	0.98	1.07	1.17	1.05	0.99	1.06	OW-NW	1.02	1.10	1.09	1.16	1.12	1.33	17.36	15.77	23.24	17.34
6.46	6.28	6.07	5.92	5.98	5.73	5.85	5.79	5.64	5.72	5.74	GL	5.49	5.02	4.81	4.75	4.85	4.60	17.88	17.59	24.25	17.80
2.42	2.47	2.49	2.48	2.54	2.41	2.32	2.37	2.43	2.44	2.22	ZG	2.42	2.16	1.90	2.05	1.99	1.87	17.42	14.74	23.71	17.38
1.19	1.09	1.02	1.09	1.02	0.95	1.12	0.97	1.05	1.04	0.92	FR	1.00	1.19	1.18	1.21	1.44	1.32	15.01	11.61	19.53	14.98
1.89	1.71	1.53	1.44	1.44	1.32	1.37	1.30	1.17	1.27	1.22	SO	1.26	1.21	1.15	1.28	1.39	1.40	17.36	17.17	24.35	17.31
1.37	1.42	1.43	1.43	1.36	1.40	1.28	1.22	1.28	1.25	1.19	BS-BL	1.20	1.13	1.06	1.04	0.94	0.91	11.24	9.19	11.22	11.21
0.85	0.77	0.70	0.61	0.72	0.65	0.77	0.63	0.60	0.68	0.70	SH	0.72	0.71	0.86	0.72	0.81	0.97	9.56	9.36	13.32	9.49
5.76	5.63	5.35	5.18	5.22	5.01	5.03	5.13	4.76	4.82	4.68	AR-AI	4.68	4.15	4.24	3.80	3.66	3.78	17.66	16.69	23.96	17.64
2.34	2.14	2.00	1.83	1.73	1.58	1.66	1.63	1.59	1.46	1.48	SG	1.49	1.22	1.25	1.20	1.15	1.12	12.02	13.27	19.78	14.49
1.97	1.82	1.63	1.60	1.66	1.43	1.84	1.57	1.52	1.67	1.60	GR	1.49	1.71	1.72	1.77	1.76	2.08	14.41	13.57	19.79	14.19
1.72	1.70	1.64	1.62	1.52	1.57	1.38	1.28	1.46	1.27	1.31	AG	1.36	1.23	1.27	1.07	1.13	1.12	17.35	16.10	24.12	17.34
2.71	2.49	2.39	2.19	2.14	1.96	2.05	2.04	1.84	1.78	1.75	TG	1.76	1.48	1.51	1.31	1.24	1.13	17.38	14.21	23.76	17.33
1.05	0.75	1.01	0.71	0.93	0.85	0.97	0.82	0.77	0.77	0.87	TI	0.86	1.03	1.14	1.08	1.09	1.20	17.13	17.30	24.36	17.32
0.52	0.50	0.47	0.55	0.53	0.57	0.61	0.54	0.66	0.64	0.63	VD	0.63	0.99	1.03	0.96	0.99	0.88	14.34	13.27	19.95	14.09
2.56	2.36	2.17	2.00	2.11	1.84	2.02	1.90	1.73	1.78	1.64	VS	1.71	1.45	1.47	1.15	1.29	1.30	15.67	14.72	21.46	15.46
1.56	1.46	1.36	1.26	1.31	1.20	1.29	1.25	1.14	1.21	1.24	NE	1.20	1.10	1.09	1.15	1.08	1.20	10.55	10.41	14.61	10.53
1.67	1.63	1.65	1.64	1.59	1.60	1.56	1.44	1.46	1.47	1.42	GE	1.38	1.27	1.14	1.13	1.05	0.97	8.96	7.33	8.87	
4.41	4.20	3.95	3.82	3.96	3.65	3.90	3.67	3.57	3.80	3.49	JU	3.54	3.25	3.48	3.33	3.03	3.21	17.54	16.82	24.08	17.41

0	1	2	3	4	5	6	7	8	9	10	ϑ	10	20	30	40	50	60	70	80	90	100
RRMSE pour la variable Pop65P											RRMSE pour la variable Pop65P										
6.53	5.92	5.46	5.16	4.88	4.56	4.31	4.11	4.05	3.88	3.60	ZH	3.74	3.08	2.48	2.38	2.14	2.22	8.00	9.45	13.35	9.56
2.62	2.29	1.82	1.67	1.85	1.61	1.62	1.77	1.46	1.60	1.60	BE	1.56	1.30	1.63	1.63	1.63	1.62	12.21	12.11	16.96	12.18
3.74	3.44	3.32	3.03	3.02	2.91	3.03	3.06	2.65	2.53	2.66	LU	2.66	2.53	2.88	2.56	2.49	2.80	11.90	9.92	14.94	11.94
11.89	11.67	11.25	10.91	11.30	10.68	10.90	10.88	10.61	10.53	10.22	UR	10.49	9.50	9.57	9.32	9.01	9.04	19.46	18.16	25.46	19.23
2.59	2.34	2.37	2.11	1.99	1.81	2.15	2.05	2.03	1.86	2.16	SZ	2.06	2.46	2.79	2.85	2.70	3.50	17.59	15.90	17.68	17.69
3.24	3.00	2.89	2.65	2.51	2.29	2.49	2.72	2.45	2.31	2.48	OW-NW	2.38	2.57	2.54	2.70	2.62	3.11	17.51	16.20	23.33	17.59
11.91	11.57	11.19	10.91	11.04	10.55	10.79	10.67	10.40	10.53	10.57	GL	10.13	9.26	8.87	8.76	8.95	8.47	19.16	18.89	25.40	19.01
6.47	6.61	6.71	6.74	6.75	6.48	6.22	6.34	6.32	6.52	5.96	ZG	6.42	5.69	5.11	5.48	5.18	5.01	18.01	16.08	24.34	17.97
2.83	2.58	2.43	2.58	2.41	2.25	2.65	2.31	2.56	2.47	2.17	FR	2.38	2.83	2.81	2.87	3.40	3.12	14.73	11.76	19.10	14.74
3.88	3.50	3.14	2.96	2.95	2.70	2.79	2.67	2.40	2.60	2.50	SO	2.57	2.48	2.36	2.63	2.85	2.87	17.49	17.36	24.47	17.47
2.58	2.68	2.69	2.69	2.55	2.62	2.41	2.30	2.40	2.34	2.24	BS-BL	2.26	2.12	1.99	1.96	1.76	1.70	10.13	8.32	10.13	10.06
1.58	1.43	1.31	1.15	1.35	1.20	1.44	1.17	1.12	1.27	1.31	SH	1.34	1.33	1.60	1.33	1.52	1.81	8.99	8.82	12.45	8.93
11.08	10.84	10.30	9.98	10.05	9.64	9.68	9.88	9.17	9.27	9.01	AR-AI	9.01	7.99	8.16	7.32	7.05	7.28	18.54	17.76	24.77	18.52
5.00	4.59	4.28	3.91	3.70	3.38	3.55	3.49	3.39	3.12	3.18	SG	3.19	2.62	2.67	2.56	2.45	2.41	11.61	12.89	19.00	13.86
4.07	3.76	3.39	3.32	3.45	2.96	3.81	3.25	3.19	3.45	3.31	GR	3.12	3.54	3.56	3.66	3.66	4.31	14.73	13.94	19.88	14.40
4.26	4.20	4.07	4.01	3.77	3.88	3.43	3.18	3.60	3.15	3.24	AG	3.36	3.04	3.15	2.66	2.80	2.77	17.52	16.53	24.35	17.58
6.07	5.58	5.36	4.90	4.80	4.39	4.60	4.57	4.13	3.98	3.93	TG	3.95	3.32	3.38	2.94	2.78	2.53	17.61	14.46	24.11	17.51
2.01	1.43	1.91	1.34	1.75	1.61	1.84	1.56	1.47	1.46	1.65	TI	1.63	1.96	2.18	2.05	2.08	2.29	17.10	17.41	24.45	17.44
1.11	1.08	1.01	1.16	1.13	1.21	1.30	1.15	1.44	1.37	1.34	VD	1.36	2.11	2.20	2.05	2.12	1.89	13.65	12.88	19.05	13.40
5.42	5.00	4.60	4.24	4.48	3.91	4.28	4.02	3.68	3.78	3.48	VS	3.64	3.07	3.12	2.44	2.74	2.75	15.83	15.11	21.66	15.58
2.90	2.70	2.53	2.35	2.44	2.23	2.40	2.33	2.13	2.26	2.30	NE	2.24	2.04	2.02	2.14	2.01	2.24	9.86	9.76	13.58	9.87
3.84	3.76	3.79	3.76	3.65	3.67	3.59	3.30	3.37	3.38	3.27	GE	3.16	2.93	2.61	2.61	2.41	2.23	8.37	6.92	8.26	
8.49	8.09	7.63	7.39	7.65	7.04	7.52	7.08	6.93	7.33	6.73	JU	6.85	6.27	6.70	6.41	5.86	6.18	18.13	17.43	24.55	17.97

Références

- DEVILLE, J.-C. et SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376 – 282.
- FAY, R. E. et HERRIOT, R. A. (1979). Estimates of income for small places : An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74:269–277.
- RAO, J. N. K. (2003). *Small Area Estimation*. Wiley, New-York.
- TILLÉ, Y. (2001). *Théorie des sondages : Échantillonnage et estimation en populations finies*. Dunod.
- YOU, Y. et RAO, J. N. K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, 30(3):431–439.