

Calcul de précision des principaux estimateurs de l'Enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux de la Drees

Audrey Isel, Emmanuel Gros

Contexte

- Des travaux réalisés dans le cadre du stage de Master 2 de Statistique-Econométrie, spécialité Statistique Publique, proposé par l'Ensaï et l'Université Rennes 1.
- Au coeur de l'article, l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux de la Drees.
- Un objectif de quantification du « reste-à-vivre » des bénéficiaires de minima sociaux rendant le calcul de précision nécessaire.
- Etude d'une population particulière de bénéficiaires de minima sociaux : les allocataires du minimum vieillesse relevant de la Cnav.

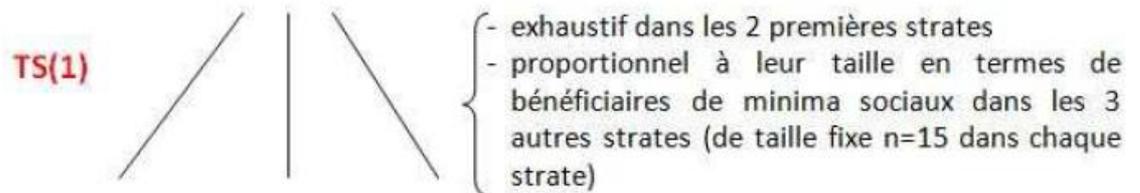
Plan de sondage de l'enquête

Un tirage à 3 degrés (avec des stratifications préalables à chaque degré) :

- 1^{er} degré : les départements

Univers : France métropolitaine hors Corse

Stratification préalable : 5 strates de départements

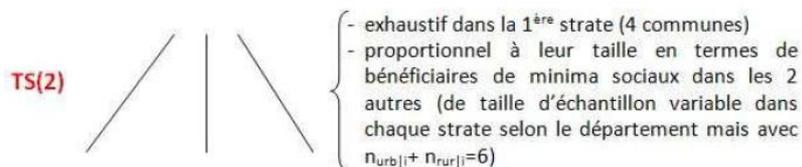


Échantillon s_i d'unités primaires : les départements

- 2^e degré : les communes ou fractions de cantons

Échantillon s_i d'unités primaires : les départements

Stratification préalable : 3 strates de communes et fractions de cantons



Échantillon s_{iq} d'unités secondaires : les communes et fractions de cantons

Stratification préalable par minimum social

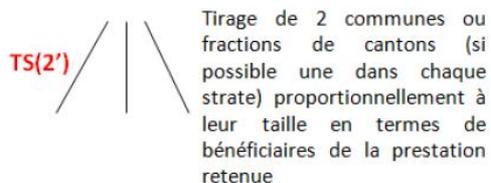
Dans le cas du minimum vieillesse, stratification par organisme gestionnaire

Test dans chacun des départements :

« Est-ce que le minimum social de l'organisme considéré y est un critère rare ? »

Si « oui » : Introduction d'une phase de tirage supplémentaire.

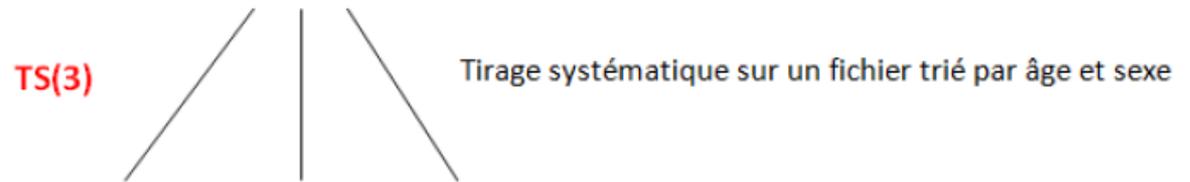
Dans le département considéré, hors strate exhaustive, parmi les 6 communes ou fractions de cantons sélectionnées dans les strates 2 et 3



Si « non » : Aucune opération supplémentaire n'est nécessaire.

- 3^e degré : les bénéficiaires, prestation par prestation

Dans chaque commune et fraction de cantons préalablement sélectionnée :



Échantillon final : les bénéficiaires du minimum vieillesse relevant de chacun des organismes

Estimation de la variance d'échantillonnage

1^{ère} étape : Estimation sans prise en compte du traitement de la non-réponse et du calage sur marges

⇒ Mise en oeuvre d'un système récursif basé sur la formule de Raj

Dans le cas d'un tirage à 2 degrés, on a :
$$\hat{V}(t_y) = f(t_y) + \sum_{i \in s} \frac{\hat{V}_i}{\pi_i}$$

avec $f(t_y)$ l'estimateur de la variance relative au 1^{er} degré de sondage

\hat{V}_i l'estimateur de la variance de t_i liée au 2^e degré de sondage au sein de l'unité primaire i

Univers : France métropolitaine hors Corse

TS(1)



$$\hat{V}(\hat{t}_y) = f^{TS(1)}(\hat{t}_i) + \sum_{i \in S} w_i \hat{V}(\hat{t}_i)$$

avec $\hat{t}_y = \sum_{i \in S} w_i \hat{t}_i$

et $w_i = \frac{1}{\pi_i}$

Échantillon s_i d'unités primaires : les départements

TS(2)



$$\hat{V}(\hat{t}_i) = f^{TS(2)}(\hat{t}_{iq}) + \sum_{q \in S_i} w_{q|i} \hat{V}(\hat{t}_{iq})$$

avec $\hat{t}_i = \sum_{q \in S_i} w_{q|i} \hat{t}_{iq}$

et $w_{q|i} = \frac{1}{\pi_{q|i}}$

Échantillon s_{iq} d'unités secondaires : les communes et fractions de cantons

TS(3)



$$\hat{V}(\hat{t}_{iq}) = f^{TS(3)}(y_k)$$

avec $\hat{t}_{iq} = \sum_{k \in S_{iq}} w_{k|i,q} y_k$

et $w_{k|i,q} = \frac{1}{\pi_{k|i,q}}$



Échantillon final : les bénéficiaires du minimum vieillesse relevant de chacun des organismes

Estimation de la variance d'échantillonnage : Tirage systématique au 3^e degré

Après avoir trié, dans chaque unité secondaire iq , les allocataires du minimum vieillesse relevant de la Cnav selon leur âge et leur sexe, on estime la variance liée à ce 3^e degré de sondage de la façon suivante :

$$\begin{aligned}\hat{V}(t_{iq}) &= f^{TS(3)}(y_k) \\ &= (w_{k|i q} n_{iq})^2 \left(1 - \frac{1}{w_{k|i q}}\right) \frac{1}{2n_{iq}(n_{iq} - 1)} \sum_{k=1}^{n_{iq}-1} (y_k - y_{k+1})^2\end{aligned}$$

avec $w_{k|i q}$ le poids de l'individu k qui réside dans la commune ou fraction de canton q du département i
 n_{iq} le nombre d'allocataires du minimum vieillesse relevant de la Cnav dans l'échantillon

Estimation de la variance d'échantillonnage : Tirages à probabilités inégales aux 1^{er} et 2^e degrés

Après stratification, les unités primaires et secondaires des strates non exhaustives ont été sélectionnées proportionnellement à leur taille en termes de bénéficiaires de minima sociaux.

L'approximation retenue pour estimer ces variances est celle de Deville :

$$\hat{V}(t_{y\pi}) = \frac{n}{n-1} \sum_{k \in s} (1 - \pi_k) \left(\frac{y_k}{\pi_k} - \frac{\sum_{k \in s} (1 - \pi_k) \frac{y_k}{\pi_k}}{\sum_{k \in s} (1 - \pi_k)} \right)^2$$

Cette formule est à ajuster selon le degré de sondage considéré et à injecter dans $f^{TS(1)}(\hat{t}_i)$ et $f^{TS(2)}(\hat{t}_{iq})$.

Un 2^e degré de tirage en 2 phases lorsque le minimum vieillesse est un "critère rare"

Pour tenir compte de cet ajustement majeur du plan de sondage, 300 000 tirages d'échantillons d'unités secondaires ont été simulés.

Idée : Assimiler ce tirage en 2 phases à un tirage en 1 phase

⇒ probabilités de sélection des unités secondaires = probabilités estimées par réplification

Résultats de ces premières estimations

| | Paramètre estimé | \hat{V} | \widehat{V}_{SAS} | \widehat{CV} | \widehat{Inf}_{IC} | \widehat{Sup}_{IC} | \widehat{Deff} |
|--|------------------|------------|---------------------|----------------|----------------------|----------------------|------------------|
| Montant mensuel moyen de minimum vieillesse perçu | 316,9 | 14,0 | 10,3 | 1,2 | 309,6 | 324,2 | 1,358 |
| Nombre total d'allocataires pauvres en conditions de vie | 158 295 | 34 038 892 | 8 752 497 | 3,7 | 146 860 | 169 730 | 3,889 |
| Proportion d'allocataires disposant d'un patrimoine immobilier | 14,7 % | 0,000 053 | 0.000 044 | 4,9 | 13,3 % | 16,2 % | 1,205 |

*Cadre : Sans prise en compte de la non-réponse totale et du calage sur marges
Pondération utilisée : poids de sélection initial des allocataires*

Prise en compte de la non-réponse totale

Le taux de réponse des allocataires du minimum vieillesse est de 33% (contre 50% en moyenne pour les bénéficiaires des autres minima sociaux).

Une modélisation par une phase de tirage supplémentaire : Tirage de l'échantillon des répondants au sein de l'échantillon sélectionné initialement.

Au final, cette phase de non-réponse peut être vue comme un degré de sondage supplémentaire \Rightarrow Utilisation de la formule de Rao pour le prendre en compte dans l'estimation de la précision :

$$\hat{V}(\hat{t}_{y,CNR}) = f\left(\frac{t_y \mathbb{I}_r}{\hat{p}_{rép}}\right) + \sum_{k \in s} \left(\left(\frac{1}{\pi_k} \right)^2 - q_k \right) (1 - \hat{p}_{k,rép}) \left(\frac{y_k \mathbb{I}_{i \in r}}{\hat{p}_{k,rép}} \right)^2$$

avec f la forme quadratique permettant d'estimer la variance de \hat{t}_y en l'absence de non-réponse
 q_k le coefficient associé aux termes diagonaux de la forme quadratique f

Résultats des estimations avec prise en compte de la non-réponse totale

Précision de l'estimateur de montant mensuel moyen de minimum vieillesse perçu :

| | Paramètre estimé | \hat{V} | \widehat{V}_{SAS} | \widehat{CV} | \widehat{Inf}_{IC} | \widehat{Sup}_{IC} | \widehat{Deff} |
|---|------------------|-----------|---------------------|----------------|----------------------|----------------------|------------------|
| Sans prise en compte de la non-réponse | 316,9 | 14,0 | 10,3 | 1,2 | 309,6 | 324,2 | 1,358 |
| Avec prise en compte de la non-réponse | 319,3 | 41,9 | 31,2 | 2,0 | 306,6 | 332,0 | 1,343 |

Cadre : Sans prise en compte du calage sur marges

Prise en compte du calage sur marges

Pour les allocataires du minimum vieillesse relevant de la Cnav, 9 variables ont été utilisées pour redresser les résultats de l'enquête :

- l'âge (en 6 tranches)
- le sexe
- être en couple ou isolé
- être né en France ou à l'étranger
- le type d'allocation du minimum vieillesse perçu
- être au minimum vieillesse depuis moins de 10 ans ou depuis 10 ans ou plus
- le montant mensuel de minimum vieillesse perçu (en 3 tranches)
- le type de droit perçu (droit propre, droit dérivé ou les 2)
- la strate de département à laquelle appartient la commune de résidence

Approximation de Deville et Särndal : $\hat{V}(\hat{t}_{y,calé}) \approx \hat{V}(\hat{t}_{\varepsilon,CNR})$

avec ε la variable de résidu de la régression de Y sur les variables de calage pondérée par le poids de non-réponse

On a donc :

$$\hat{V}(\hat{t}_{y,calé}) = f\left(\frac{t_{\varepsilon} \mathbb{I}_r}{\hat{p}_{rép}}\right) + \sum_{k \in s} \left(\left(\frac{1}{\pi_k}\right)^2 - q_k \right) (1 - \hat{p}_{k,rép}) \left(\frac{\varepsilon_k \mathbb{I}_{i \in r}}{\hat{p}_{k,rép}} \right)^2$$

Effet de calage

| Variable d'intérêt | Estimation de l'effet de calage |
|--|---------------------------------|
| Montant mensuel moyen de minimum vieillesse perçu | 0.17 |
| Total d'allocataires pauvres en conditions de vie | 0.41 |
| Proportion d'allocataires ayant connu de longues périodes d'activité ou n'ayant jamais travaillé | 0.73 |
| Proportion d'allocataires qui se déclarent en mauvaise santé | 0.85 |
| Proportion d'allocataires disposant d'un patrimoine immobilier | 0.80 |

Précision finale des estimateurs retenus

| | Montant mensuel moyen de minimum vieillesse perçu | Nombre total d'allocataires pauvres en conditions de vie | Proportion d'allocataires ayant connu de longues périodes d'activité ou n'ayant jamais travaillé | Proportion d'allocataires qui se déclarent en mauvaise santé | Proportion d'allocataires disposant d'un patrimoine immobilier |
|-------------------------|---|--|--|--|--|
| Paramètre estimé | 316.9 | 141 784 | 22,3 % | 46,6 % | 14,3 % |
| \hat{V} | 7.1 | 24 651 594 | 0,000 149 | 0,000 258 | 0,000 154 |
| \widehat{V}_{SAS} | 6.5 | 20 037 964 | 0,000 138 | 0,000 238 | 0,000 115 |
| \widehat{CV} | 0,8 % | 3,5 % | 5,5 % | 3,5 % | 8,7 % |
| \widehat{Inf}_{IC} | 311,6 | 132 052 | 19,9 % | 43,5 % | 11,8 % |
| \widehat{Sup}_{IC} | 322,1 | 151 515 | 24,7 % | 49,8 % | 16,7 % |
| \widehat{Def}_f | 1,092 | 1,230 | 1,079 | 1,087 | 1,340 |

Conclusion

Au vu des résultats obtenus, une enquête statistique de très bonne qualité.

Des travaux complémentaires peuvent être envisagés :

- Extension du champ de l'étude aux autres minima sociaux de l'enquête
⇒ Problème du tirage en 2 phases, la base de sondage utilisée étant elle-même un échantillon
- Extension à d'autres variables d'intérêt, en particulier aux variables quantitatives qui seront disponibles mi-2015

Merci de votre attention !