



---

**Commission économique pour l'Europe**

Conférence des statisticiens européens

**Groupe d'experts des recensements  
de la population et des habitations****Vingt et unième réunion**

Genève, 18-20 septembre 2019

Point 4 de l'ordre du jour provisoire

**Travaux de recherche sur l'utilisation des données  
administratives pour les recensements****Problèmes et solutions possibles en matière d'organisation  
d'un recensement de la population à partir de registres****Note de Statistics Estonia\****Résumé*

L'évolution de la société entraîne des changements dans de nombreux phénomènes et concepts qui leur sont liés et qui avaient longtemps été stables. Cela rend la production et l'utilisation de statistiques concernant de tels phénomènes compliquées et souvent ambiguës, car il y a rarement de la place dans le système de statistique traditionnel pour les phénomènes et concepts nouveaux.

Statistics Estonia a résolu ces problèmes grâce à une méthode innovante qui s'appuie sur les informations contenues dans les registres, ou plus précisément sur des indices créés par l'utilisation croisée d'un grand nombre de registres [1-4].

---

\* Établie par Ene-Margit Tiit.



## I. Introduction

1. Le nombre de personnes qui composent une population, c'est-à-dire le nombre de résidents permanents dans un pays, est en général déterminé par recensement. Récemment, toutefois, le sous-dénombrement est devenu un problème dans de nombreux pays, car il s'est avéré impossible durant le recensement de contacter toutes les personnes qui devaient être comptabilisées. Il y a plusieurs raisons à cela :

a) La première est que les gens sont beaucoup plus mobiles qu'auparavant ;

b) La seconde est d'ordre psychologique : les gens ne veulent pas communiquer leurs données personnelles et s'efforcent de préserver à tout prix leur vie privée.

2. Il s'agit d'un problème mondial qui, pour le moment, se manifeste surtout dans les pays développés mais qui prend de l'ampleur sous l'effet de la couverture croissante par les médias des problèmes de protections des données et des cas de fuite et de piratage. Cela pose un sérieux problème aux fournisseurs officiels de statistiques.

3. Depuis le milieu du XX<sup>e</sup> siècle, les statisticiens ont mis au point diverses méthodes pour vérifier la couverture du recensement, c'est-à-dire pour déterminer dans quelle mesure le nombre d'habitants comptabilisés par le recensement correspond à la réalité, par exemple en procédant à un suivi du recensement [5-6].

4. Ces méthodes permettent habituellement d'évaluer la taille exacte de la population mais s'avèrent moins efficaces si l'on a besoin de statistiques démographiques basées sur les personnes, par exemple celles qui donnent des informations précises sur la répartition des habitants en fonction de différents marqueurs.

5. Pour obtenir de telles informations il est nécessaire d'avoir recours à des sources de données supplémentaires afin d'ajuster les données sous-dénombrées ou sur-dénombrées produites par recensement.

## II. Registres et système de registres

6. Des méthodes d'évaluation relativement efficaces peuvent être utilisées dans les pays disposant d'un système de registres de haute qualité remplissant les principales conditions suivantes :

a) Tous les objets (personnes et logements) peuvent être identifiés de façon univoque et les identifiants sont universels pour tous les registres qui appartiennent au système ;

b) Les objets de différents types (personnes et logements) sont reliés entre eux par des identifiants ;

c) Les registres sont complets et contiennent les données de la population concernée ;

d) Les informations contenues dans les registres sont vérifiées et régulièrement actualisées (au moins une fois par an).

7. L'Estonie dispose d'un système qui respecte les exigences susmentionnées et qui, outre les personnes et les logements, contient aussi en tant qu'objets les entreprises avec leurs identifiants.

8. Bien que les informations rassemblées dans les registres soient généralement basées sur des documents, les risques d'erreurs ne sont pas exclus :

a) Les données peuvent être obsolètes ;

b) Elle peuvent contenir des erreurs accidentelles (pour cause de négligence ou d'erreur humaine) ;

c) Des données peuvent être systématiquement erronées dans l'intérêt de quelqu'un (le plus souvent de la personne qui les a soumises, ou parfois aussi du détenteur du registre).

9. Il est primordial pour disposer de statistiques basées sur des registres de vérifier continuellement la qualité des données contenues dans ces registres, et en particulier leur exactitude (correspondance entre les données et la situation réelle).

10. Un moyen relativement efficace de le faire est l'utilisation croisée des registres, lorsque les données contenues dans un registre sont vérifiées à l'aide d'autres registres contenant des données indépendantes.

11. Dans la plupart des pays, les possibilités de procéder à une telle utilisation croisée sont limitées par des lois de protection des données et sont réservées aux organisations qui produisent des statistiques, ce qui est le cas de Statistics Estonia. Statistics Estonia n'est pas autorisée à révéler à d'autres organisations les résultats de cette utilisation croisée (même pas les erreurs constatées) [7].

12. Puisque la loi et les bonnes pratiques ne permettent pas de demander plusieurs fois (pour différents registres) les mêmes informations à une personne, les données erronées se transmettent d'un registre à l'autre, ce qui arrive par exemple avec les adresses inscrites dans le registre de la population.

13. Du point de vue des utilisateurs de données il s'agit d'une grave source de préoccupation, car la précision des adresses constitue l'un de problèmes les plus sérieux auxquels se trouve confronté le système estonien des registres – les adresses qui figurant dans le registre de la population ne correspondent pas à la réalité ou sont imprécises [7].

14. Statistics Estonia a néanmoins réussi à élucider plusieurs problèmes liés aux statistiques démographiques en utilisant des registres.

### **III. Correction du sous-dénombrement lors du recensement en Estonie**

15. Le recensement le plus récent effectué en Estonie, celui de 2011, avait sous-dénombré la population. Cela a été signalé par les participants au recensement eux-mêmes. Il est également apparu que le nombre d'habitants comptabilisés était de près de 5 % inférieur au nombre de personnes résidant en Estonie selon le registre de la population.

16. Pour obtenir le nombre d'habitants réel (et une liste des personnes résidant effectivement en Estonie) il a été décidé d'utiliser plusieurs registres contenant des données personnelles. Il s'agissait de la première tentative de corriger les données concernant la population grâce à l'utilisation croisée de données issues de dossiers [10, 11].

17. À cette fin, les données concernant les habitants de l'Estonie ont été collectées à partir de plus de dix registres et sous-registres. En raison du fait que l'activité des gens dans les registres dépend beaucoup de leur âge, et dans une certaine mesure aussi de leur sexe, douze groupes d'âge et de sexe séparés ont été pris en compte.

18. Dans chaque groupe d'âge, plusieurs modèles différents ont été utilisés – analyses linéaire et logistique, ainsi qu'évaluations d'experts par comparaison. Les résultats des différents modèles ont largement coïncidé. Les erreurs des modèles ont été évaluées sur la base de résultats d'essais issus de recherches. Elles étaient inférieures à 5 % dans tous les groupes d'âge et de sexe.

19. En conclusion, il s'est avéré que le recensement de 2011 avait été marqué par une sous-couverture de 2,3 % [11]. À partir de l'année suivant ce recensement, les statistiques démographiques ont utilisé des données de recensement dûment corrigées.

20. Il est également apparu que le nombre de résidents en Estonie avait été surestimé d'environ 3% dans les registres de la population, la raison en étant que tous les départs du pays n'avaient pas été enregistrés.

## IV. Problèmes liés au pays de résidence

21. Bien qu'il existe un certain nombre de normes internationales concernant la manière de déterminer le pays de résidence d'une personne dans le cadre d'un recensement, ces normes et leurs exceptions posent plusieurs problèmes essentiels qui compliquent cette détermination.

22. En général, on ne devient résident permanent d'un pays qu'après y avoir vécu pendant douze mois, mais il y a des exceptions pour les personnes qui n'ont pas passé autant de temps dans le pays mais qui ont l'intention de le faire.

23. Une telle définition est inévitablement contradictoire, car il n'est pas possible de vérifier dans tous les cas l'intention de quelqu'un (une personne peut répondre pour une autre sans connaître pour autant les intentions de celle qu'elle représente). Il peut aussi arriver que la personne elle-même exprime une intention qui n'est pas sérieuse ou qui ne se matérialisera pas, pour une raison ou pour une autre.

24. En principe, les immigrants et les émigrants devraient être traités de manière symétrique. Or il est tout à fait irréaliste de demander à des gens qui ont quitté le pays depuis moins d'un an s'ils comptent rester à l'étranger au moins une année ou s'ils prévoient de rentrer avant.

25. Dans ce dernier cas, ces personnes doivent être comptées comme faisant partie de la population du pays, même s'il est pratiquement impossible d'obtenir des données les concernant s'ils n'ont pas de proches parents dans le pays de départ.

26. La relation entre résidents temporaires et résidents permanents est également source de confusion. Les premiers sont des gens qui vivent à un certain endroit (adresse, localité, pays) pendant plus de trois mois mais moins de douze mois par an, ce qui signifie que ce ne sont pas des résidents permanents. Les seconds, en revanche, sont des gens qui résident quelque part de manière permanente, mais qui peuvent séjourner temporairement ailleurs (y compris dans un autre pays) si la longueur de ce séjour ne dépasse pas douze mois.

27. Si donc quelqu'un séjourne entre trois et neuf mois par an dans un pays et le reste du temps, c'est-à-dire également entre trois et neuf mois, dans un autre (ce qui n'est pas du tout un cas rare), son statut de résident permanent doit être déterminé en fonction de son lieu de résidence avant son départ – si son intention de rester plus longtemps dans l'un de ces pays n'a pas été exprimée. En fait, une personne peut même séjourner ainsi dans trois pays et il n'est pas facile de déterminer dans lequel elle est résidente permanente.

28. Il ne semble pas y avoir de problème si l'immigrant précise son pays de résidence permanente lors d'un entretien dans le cadre d'un recensement traditionnel (soit en suivant les instructions du recenseur ou de la manière qu'il juge appropriée).

29. Telle était en général la pratique dans le passé. Toutefois, on ne sait pas exactement avec quelle précision les règlements et les délais étaient respectés dans de tels cas. Des membres de la famille pouvaient avoir compté le parent parti comme résident permanent même si le départ remontait à plus de douze mois.

30. Établir le pays de résidence est encore plus compliqué dans le cas des nombreuses exceptions internationalement reconnues. Ainsi, par exemple, le pays de résidence des étudiants de l'enseignement supérieur est le pays où ils étudient (malgré de possibles aller et retour entre leur pays natal et celui où ils font leurs études). Et il n'est pas simple non plus de déterminer sans ambiguïté ce qu'est une « migration pendulaire ».

## V. Déterminer la taille d'une population à l'aide d'indices

31. Il est assez facile de corriger la taille de la population après le recensement en se basant sur les registres, car on peut utiliser comme « ensemble de données d'entraînement » la population comptabilisée et enregistrée comme résidente permanente du pays par opposition aux personnes non comptabilisées et à celles qui sont enregistrées comme résidents étrangers. Il ne convient toutefois pas d'utiliser le même algorithme pour déterminer le nombre précis d'individus qui composent une population pendant toute la

période comprise entre les recensements, car les données recueillies dans le cadre des recensements deviennent vite obsolètes en raison de la mobilité de la population.

32. Le défi est de savoir comment tenir à jour la liste des résidents permanents, c'est-à-dire de déterminer à tout moment (en pratique chaque année), de manière nominative, le groupe des personnes qui résident de manière permanente dans le pays et le groupe de celles qui l'ont quitté au cours des années (et qui sont susceptibles d'y revenir). Additionnés, ces deux groupes forme l'ensemble de personnes que l'on appelle pour cette raison la population élargie ; il s'agit des personnes au sujet desquelles existent ou ont existé auparavant des données dans les registres estoniens (y compris dans le registre de la population).

33. Pour définir les résidents permanents, on a créé l'indice de résidence  $R_i(k)$  selon la formule :

$$R(k) = d \cdot R(k-1) + g \cdot X(k-1), \quad (1)$$

où (1) est la valeur de l'indice de résidence attribué à la personne  $i$  pour l'année  $k$  (valeur comprise entre zéro et un).

34. Dans la formule,  $R(k-1)$  est la valeur de l'indice de résidence de la personne pour l'année précédente, alors que  $X(k-1)$  est la somme pondérée des signes de vie [20] accumulés par la personne  $i$  au cours de l'année précédente :

$$X(k-1) = \sum a_j b_j \quad (2)$$

35. Le signe de vie  $b_j$  correspondant au registre  $j$  est une variable binaire, dont la valeur est égale à 1 si la personne  $i$  a été active dans le registre  $j$  au cours de l'année précédente et à 0 si elle ne l'a pas été ( $j = 1, 2, \dots, m$ ), où  $m$  (nombre de facteurs dans la formule (2)) est le nombre de registres et sous-registres utilisés. Ce nombre est actuellement supérieur à 20 en Estonie.

36. La personne se voit donc attribuer un signe de vie correspondant si, par exemple, elle va consulter un médecin, étudie dans une école estonienne, touche des prestations sociales de l'administration locale ou achète des médicaments sur ordonnance.

37. Le premier terme de la formule (1), appelé élément de stabilisation, relie le statut de résidence de la personne avec le passé, et son poids est déterminé par le paramètre  $d$  ; le deuxième terme est le signe de vie, qui est une mesure de l'activité de la personne au cours de l'année précédente [2].

38. Les poids  $a_j$  des signes de vie  $b_j$  sont déterminés sur la base des données des années précédentes en tant que ratio des probabilités conditionnelles d'occurrence du signe de vie dans le cas de résidents certains et de non-résidents certains.

39. La base pour déterminer les paramètres  $d$  et  $g$  ( $d + g = 1$ ) dans la formule (1) est une évaluation d'expert de la politique démographique pour savoir combien de temps il faut à un immigrant pour obtenir le statut de résident permanent et combien de temps une personne restera sur la liste des résidents permanents si elle ne montre aucune activité [2]. Dans le cas présent, les valeurs  $d = 0,8$  et  $g = 0,2$  sont utilisées. Une valeur seuil de 0,7 est utilisée, ce qui, selon les données reçues jusqu'à présent, minimise et équilibre les erreurs en termes tant d'inclusion que d'exclusion.

40. En vertu des règles internationales, la période de stabilité est de deux ans dans les deux sens (ce qui couvre le délai de douze mois même si les signes de vie apparaissent à la fin de l'année), mais des personnes particulièrement actives peuvent devenir résidents en moins d'un an, bien qu'une condition supplémentaire stipule qu'elles doivent avoir un domicile (enregistré) en Estonie.

41. L'indice de résidence est calculé chaque année pour toutes les personnes qui font partie de la population élargie et, l'année  $k$ , les personnes dont l'indice de résidence est supérieur à la valeur seuil  $c$ , qui est déterminée de manière empirique, sont considérées comme résidentes.

## VI. Avantages et inconvénients d'une méthode basée sur des indices pour évaluer le nombre de personnes dans une population par rapport au recensement traditionnel

42. Le nombre de personnes dans une population est calculé en Estonie depuis 2012 à l'aide d'une méthode basée sur des indices, et depuis 2016 toutes les statistiques démographiques reposent sur le concept des indices de résidence. Cette méthode a été vérifiée par des études d'échantillonnage au cours desquelles on questionne les personnes pour déterminer leur pays de résidence.

43. Il a été démontré que lorsqu'on compare la méthode basée sur des indices avec les données par questionnaire, les erreurs d'inclusion et d'exclusion ne dépassent pas 1 % [2].

44. Le fait de déterminer la résidence à l'aide de l'indice présente par rapport au recensement traditionnel basé sur des questions les avantages suivants :

a) Couverture : on obtient en principe des informations sur les personnes à comptabiliser où qu'elles se trouvent au moment du recensement ;

b) Uniformité : les données concernant toutes les personnes proviennent des mêmes registres et sont calculées à l'aide des mêmes algorithmes. Les informations ne dépendent donc pas de la question de savoir si la personne concernée répond elle-même ou si les données sont communiquées par d'autres membres de son foyer ;

c) Objectivité : le résultat ne dépend pas de ce que déclare la personne, ni de sa volonté ou pas de dévoiler ses intentions, ni de sa santé, de son état mental ou de sa mémoire ;

d) La probabilité d'erreurs accidentelles, notamment d'erreurs humaines, diminue. Dans le cas des recensements basés sur des questionnaires, de telles erreurs sont en revanche inévitables dans une mesure plus ou moins grande.

45. Le principal inconvénient de la méthode basée sur des indices réside dans le fait qu'elle est tributaire de la qualité des registres et notamment de la rigueur avec laquelle y sont appliquées les définitions de la résidence (par exemple si une administration locale octroie des prestations à des personnes qui ont en réalité quitté le territoire national).

46. En raison du grand nombre de signes de vie, les erreurs susceptibles d'exister dans un seul registre n'ont pas beaucoup d'importance lorsqu'il s'agit d'établir le lieu de résidence.

47. Toutefois, avec la méthode basée sur des indices, des résidents permanents totalement passifs peuvent aussi passer inaperçus s'ils n'ont aucune activité dans aucun des registres pendant plusieurs années, c'est-à-dire si la personne ne travaille pas, ne touche aucune prestation, ne fréquente aucun établissement de santé et n'enfreint aucune loi.

48. Ces personnes sont à l'abri d'une exclusion prématurée de la population permanente par l'élément stabilisateur de l'indice de résidence (voir la formule (1)).

49. Il est également possible qu'une personne résidant à l'étranger soit comptabilisée parmi les résidents permanents si elle a recours à des services lorsqu'elle se trouve en Estonie : visite chez le médecin, achat de médicaments, transactions commerciales, etc. C'est pour prévenir une inclusion prématurée parmi les résidents que l'on utilise un délai de temporisation.

## VII. Conclusion

50. En conclusion, des erreurs sont aussi possibles lorsqu'on applique la méthode basée sur des indices pour évaluer le nombre de personnes que compte une population, mais leur probabilité décroît si la qualité des registres s'améliore et si de nouvelles sources de données sont utilisées. La précision des estimations basées sur des indices est évaluée au moyen d'enquêtes supplémentaires et les résultats sont communiqués avec les valeurs d'erreur potentielles. L'ajout de nouvelles informations (de nouveaux signes) se traduira par une amélioration constante de la précision des estimations basées sur des indices.

## Bibliographie

- Tiit, Ene-Margit (2017). Residency Testing. Estimating the true population size of Estonia. *Statistics in Transition*, Warsaw, 2, 211-226.
- Maasing, Ethel ; Tiit, Ene-Margit; Vähi, Mare (2017). Residency index – a tool for measuring the population size. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 21 (1), 129-139.
- Tiit, E.-M.; Vähi, M. (2017). Indexes in demographic statistics: a methodology using nonstandard information for solving critical problems. *Papers on Anthropology*, 26 (1), 72-87.
- Tiit, Ene-Margit ; Maasing, Ethel (2016). RESIDENCY INDEX AND ITS APPLICATIONS IN CENSUSES AND POPULATION STATISTICS. *Eesti Statistika Kvartalikirj*, 3, 41-60.
- Tamm, Ebu (2001a). Sample Design for the Evaluation Study of the 2000 Year Census in Estonia. [http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/documents/60.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/60.pdf).
- Tamm, Ebu (2001c). Rahvaloenduse andmete kvaliteedi hindamisest. *ESS Teabevihik* nr 12, lk 8-14.
- Registripõhise rahva ja eluruumide loenduse (REGREL) meetoodika väljatöötamine. (2013). Koost A. Puur, L. Sakkeus, S. Aben. Tallinn. [www].
- Tiit, Ene-Margit. Rahvaloendused Eestis (2011), *Akadeemia*, nr 12, lk 2276-2312).
- Tiit, E.-M. (2013). Eesti rahvaarvu ja Eestist lahkunute arvu hinnangud. – *Riigikogu toimetised*, nr 27, lk 153-169).
- Tiit, E.-M., K. Meres, M. Vähi 2012. Rahvaloenduse üldkogumi hindamine. – *Eesti Statistika Kvartalikirj*, nr 3, lk 79-95.
- Tiit, Ene-Margit (2011). 2011.aasta rahva ja eluruumide loenduse alakaetuse hinnang. *Eesti Statistika kvartalikirj* 4, 12.