



Европейская экономическая комиссия

Конференция европейских статистиков

Группа экспертов по переписям населения и жилищного фонда

Двадцать первое совещание

Женева, 18–20 сентября 2019 года

Пункт 4 предварительной повестки дня

**Исследования по вопросам использования
административных данных в целях переписи**

Проблемы, возникающие при подготовке регистровой переписи населения, и возможные пути их решения

Записка Департамента статистики Эстонии*

Резюме

Развитие общества ведет к изменению многих явлений и связанных с ними понятий, которые оставались стабильными на протяжении длительного времени. Это осложняет получение и использование статистических данных о таких явлениях и зачастую делает их спорными, поскольку в традиционной статистической системе для новых явлений и концепций нет места.

Департамент статистики Эстонии решает эти проблемы, используя инновационный метод, в основе которого лежит информация из регистров, точнее говоря, индексы, получаемые посредством перекрестного использования большого числа регистров [1–4].

* Подготовлена Эне-Маргит Тийт.



I. Введение

1. Численность населения, т. е. число постоянных жителей страны, обычно определяется путем переписи населения. Однако в последнее время перед многими странами встала проблема неполного охвата населения переписью – в ходе переписи оказалось невозможным связаться со всеми теми, кого следует учитывать. Этому есть несколько причин:

a) во-первых, люди стали гораздо мобильнее, чем раньше;

b) вторая причина связана с психологией – люди не хотят делиться своими персональными данными и всеми способами пытаются защитить свою частную жизнь.

2. Хотя до сих пор эта проблема проявлялась в основном в развитых странах, она носит глобальный характер и встает со всей остротой, поскольку вопросы защиты данных и случаи утечки и взлома данных все чаще получают огласку в СМИ. Это служит серьезным вызовом для поставщиков официальной статистики.

3. С середины XX века статистики изобрели различные методы проверки охвата переписи, т. е. оценивания того, насколько точно численность населения по результатам переписи соответствует действительности, например с помощью проверочной переписи [5–6].

4. Обычно такие методы позволяют правильно оценить численность населения, но не столь полезны, если требуется демографическая статистика, в большей степени основанная на личных данных, т. е. содержащая точную информацию о распределении населения по различным признакам.

5. Для получения такой информации необходимо использовать дополнительные источники данных, позволяющие скорректировать сведения, полученные в результате переписи с неполным или избыточным охватом.

II. Регистры и система регистров

6. Сравнительно эффективные методы оценки численности населения могут применяться в странах с функциональной системой высококачественных регистров, отвечающей следующим основным требованиям:

a) все объекты (люди и жилища) могут быть индивидуально идентифицированы, и идентификаторы являются универсальными для всех регистров системы;

b) объекты различных типов (люди и жилища) связаны между собой через идентификаторы;

c) регистры являются полными, т. е. содержат данные о всей совокупности, которой посвящен конкретный регистр;

d) информация в регистрах проверяется и обновляется регулярно (не реже одного раза в год).

7. В Эстонии существует отвечающая вышеуказанным требованиям система регистров, в которой наряду с лицами и жилыми помещениями в качестве объектов регистрации выступают коммерческие единицы со своими идентификаторами.

8. Хотя собранная в регистрах информация, как правило, основана на документах, нельзя исключать возможность ошибок:

a) данные могут быть устаревшими;

b) они могут содержать случайные ошибки (которые могут быть вызваны небрежностью или человеческой ошибкой);

c) в регистры в интересах отдельных субъектов (чаще всего лиц, предоставляющих данные, а иногда и тех, кто отвечает за ведение регистров) могут вноситься систематически ошибочные данные.

9. Одной из предпосылок использования регистровой статистики является непрерывная проверка качества данных в регистрах, в частности их правильности (соответствие между данными и фактическим положением дел).
10. Сравнительно эффективным способом такой проверки является перекрестное использование регистров, т. е. проверка данных одного регистра через другие регистры, содержащие независимые данные.
11. Однако в большинстве стран возможности такого перекрестного использования ограничены законами о защите данных, и такое использование разрешено исключительно организациям, формирующим статистику, в том числе Департаменту статистики Эстонии. Департаменту статистики Эстонии не разрешается сообщать другим организациям результаты перекрестного использования (в том числе об обнаруженных ошибках) [7].
12. Поскольку законодательство и принципы надлежащей практики не позволяют несколько раз запрашивать у одного и того же лица одни и те же данные (для разных регистров), ошибочные данные одного регистра переносятся в другие регистры – например, это происходит с адресами, внесенными в регистр населения.
13. С точки зрения пользователей данных, это является серьезной причиной для беспокойства, так как точность адресов является одной из самых больших проблем в эстонской системе регистров – адреса в регистре населения часто не соответствуют действительности или не точны [7].
14. Тем не менее Департамент статистики Эстонии научился решать ряд задач демографической статистики с помощью регистров.

III. Корректировка на неполноту охвата переписи в Эстонии

15. Последняя перепись населения Эстонии, проведенная в 2011 году, страдала недостаточным охватом. Сигналы об этом поступали от самих участников переписи. Также оказалось, что численность переписанных лиц была почти на 5% меньше числа жителей Эстонии по данным регистра населения.
16. Для того чтобы узнать реальное число жителей (и список лиц, фактически проживающих в Эстонии), было принято решение использовать несколько регистров, содержащих персональные данные. Это была первая попытка скорректировать данные о населении путем перекрестного использования регистровых данных [10, 11].
17. С этой целью данные о жителях Эстонии были собраны более чем из десяти регистров и подрегистров. Учитывая, что указанная в регистрах активность лиц в значительной степени зависит от их возраста и в некоторой степени от пола, было выделено 12 отдельных возрастных групп.
18. К каждой возрастной группе применялось несколько различных моделей – линейный и логистический анализ, а также экспертные оценки в целях сравнения. Результаты применения различных моделей в значительной степени совпали. На основе контрольных данных, полученных исследовательским путем, была проведена оценка погрешностей в моделях. Во всех возрастных группах ошибки не превышали 5%.
19. В результате выяснилось, что неполнота охвата переписи 2011 года составила 2,3% [11]. Начиная с последующего после переписи года в демографической статистике использовались данные переписи, скорректированные на неполноту охвата.
20. Выяснилось также, что число жителей Эстонии в регистре населения было завышено почти на 3% поскольку не все покинувшие Эстонию жители зарегистрировали свой отъезд.

IV. Проблемы, связанные со страной проживания

21. Хотя существует ряд международных стандартов определения страны проживания человека для целей переписи, в связи с этими стандартами и исключениями из них возникает ряд существенных проблем, которые затрудняют определение страны проживания.
22. Как правило, лицо становится постоянным жителем только после 12 месяцев проживания в стране, но делается исключение, если человек не живет в стране так долго, но имеет такое намерение.
23. Такое определение неизбежно является противоречивым, поскольку невозможно проверить намерение человека во всех случаях (человек может ответить от имени другого человека, но он может не знать о намерениях того, кого он представляет). Также может случиться, что человек сам выражает намерение, которое не является серьезным или не материализуется по той или иной причине.
24. В принципе, учет иммигрантов и эмигрантов должен быть симметричным. Однако совершенно нереалистично спрашивать у людей, покинувших страну менее года назад, собираются ли они оставаться за границей по крайней мере на протяжении одного года или планируют вернуться раньше.
25. В последнем случае эти люди должны учитываться в составе населения страны, однако в случае отсутствия у них близких родственников в стране, которую они покинули, получить информацию о них практически невозможно.
26. Связь между временными и постоянными жителями также не всегда очевидна. Временными жителями считаются лица, проживающие в определенном месте (по конкретному адресу, в населенном пункте, в стране) свыше 3, но менее 12 месяцев, т. е. лица, не являющиеся постоянными резидентами. Постоянными резидентами в то же время считаются лица, проживающие в каком-либо месте постоянно, однако они могут временно проживать в другом месте (в том числе в другой стране), если срок их пребывания там не превышает 12 месяцев.
27. Таким образом, если лицо находится от трех до девяти месяцев в году в одной стране, а остальное время, также от трех до девяти месяцев, в другой стране (что происходит совсем нередко), его статус постоянного резидента должен определяться по его предыдущему, до миграции, месту жительства, если он не заявил о своем намерении оставаться в одной из этих стран в течение более длительного времени. На самом деле, человек может проживать таким образом даже в трех странах, и не так-то просто определить, в какой из них он проживает постоянно.
28. Как представляется, проблемы не возникают, если иммигрант определяет страну своего постоянного проживания в ходе опроса в рамках традиционной переписи (либо по указаниям переписчика, либо по собственному мнению).
29. Как правило, в прошлом дела обстояли именно таким образом. Однако неясно, как именно в таких случаях соблюдались правила и временные ограничения. Члены семьи могут считать уехавшего члена семьи постоянным жителем страны, даже если этот член семьи пробыл за границей более 12 месяцев.
30. Определение страны проживания еще более осложняется различными исключениями, согласованными на международном уровне. Например, страной проживания студентов высших учебных заведений считается страна, где они проходят обучение (несмотря на возможную миграцию из страны происхождения в страну обучения и обратно). Кроме того, не так просто однозначно локализовать «маятниковую миграцию».

V. Определение численности населения с помощью индексов

31. Переписную численность населения сравнительно легко скорректировать по данным регистров, поскольку население, переписанное и зарегистрированное в качестве постоянно проживающего в стране происхождения, по сравнению с численностью неучтенных лиц и лиц, зарегистрированных в качестве иностранных жителей, можно использовать в качестве «учебных наборов данных». Однако использование одного и того же алгоритма для определения точной численности населения в течение всего периода между переписями нецелесообразно, поскольку данные переписи быстро устаревают из-за мобильности населения.

32. Была поставлена задача – определить, как можно обеспечить непрерывное обновление списка постоянных жителей государства, с тем чтобы в любой момент (на практике каждый год) можно было составить поименный список постоянных жителей страны и лиц, которые покинули страну (и которые потенциально могут вернуться). Оба списка образуют совокупность, которая далее именуется «расширенным населением»; это – лица, о которых имеются или раньше имелись сведения в эстонских регистрах (в том числе в регистре населения).

33. Для определения постоянных жителей был построен индекс резидентства $R_i(k)$, рассчитываемый по следующей формуле:

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

где (1) – значение индекса резидентства, присваиваемого лицу i в k году (индекс может принимать значения от нуля до единицы).

34. В формуле $R(k-1)$ – это значение индекса резидентства лица в предыдущем году, $X(k-1)$ – взвешенная сумма признаков жизни [20], собранных по каждому лицу i в предыдущем году:

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

35. Признак жизни b_j , соответствующий регистру j , является двоичной переменной, значение которой равно 1, если лицо i было активным в регистре j в предыдущем году, и 0, если лицо не было активным в этом регистре в предыдущем году ($j = 1, 2, \dots, m$), где число m (количество членов формулы (2)) – количество используемых регистров и подрегистров. В настоящее время в Эстонии их насчитывается более 20.

36. Таким образом, человек приобретает соответствующий признак жизни, если он, например, посещает семейного врача, учится в эстонской школе, получает пособия по социальному обеспечению от органов местного самоуправления или покупает лекарства по рецепту.

37. Первым членом формулы (1) является так называемый стабилизирующий член, который увязывает статус резидента с прошлым, и его вес определяется параметром d ; вторым членом является так называемый член признаков жизни, и его вес характеризует активность человека в предыдущем году [2].

38. Веса a_j признаков жизни b_j определяются на основе данных за предыдущие годы как соотношение условных вероятностей возникновения данного признака жизни в случае конкретных резидентов и конкретных нерезидентов.

39. Основой для определения параметров d и g ($d + g = 1$) в формуле (1) служит политико-демографическая экспертная оценка: как быстро иммигрант может получить статус постоянного резидента и как долго человек будет находиться в списке постоянных жителей, если он не проявляет никакой активности [2]. В данном случае используются значения $d = 0,8$, $g = 0,2$. Используется пороговое значение 0,7, которое, согласно полученным до сих пор данным, сводит к минимуму и уравнивает ошибки включения и исключения.

40. Согласно международным правилам, период стабильности в обоих направлениях составляет два года (он включает 12-месячный лимит, даже если признаки жизни проявляются в конце года), однако особенно активные лица могут стать резидентами в течение одного года, хотя дополнительным условием является наличие у них места жительства (зарегистрированного) в Эстонии.

41. Индекс резидентства рассчитывается ежегодно для всех лиц, принадлежащих к расширенному населению, и в k году резидентами считаются лица, индекс резидентства которых превышает пороговое значение s , которое определяется эмпирически.

VI. Преимущества и недостатки индексной оценки численности населения по сравнению с традиционной переписью населения

42. Численность населения Эстонии с 2012 года рассчитывается с помощью индекса, а с 2016 года вся демографическая статистика основывается на индексной концепции резидентства. Метод был проверен выборочными исследованиями, в ходе которых страна проживания людей определялась путем опроса.

43. Было установлено, что в сравнении с вопросниками при использовании индексного метода ошибки включения и исключения не превышают 1% [2].

44. По сравнению с традиционной переписью методом опроса определение места жительства с помощью индекса имеет следующие преимущества:

a) охват: в принципе, информация собирается о подлежащих переписи лицах независимо от того, где они находятся в данный момент;

b) единообразие: источниками данных обо всех лицах являются одни и те же регистры и одни и те же вычислительные алгоритмы. Таким образом, информация не зависит от того, отвечает респондент на вопросы сам или же данные предоставляют члены его домохозяйства;

c) объективность: результат не зависит ни от заявленных намерений человека и от его желания раскрывать свои намерения, ни от состояния его здоровья, психики или памяти;

d) вероятность случайных ошибок, в том числе человеческих, снижается. В случае проведения переписей методом опроса такие ошибки всегда в той или иной мере неизбежны.

45. Наибольшим недостатком индексного метода является зависимость от качества регистров, в том числе от того, насколько строго в них выдерживаются определения резидентства (например, выдают ли те или иные местные органы самоуправления пособия лицам, фактически выехавшим за границу).

46. В связи с большим количеством признаков жизни ошибки в одном регистре не оказывают очень сильного влияния на определение резидентства.

47. Однако в случае использования индексного метода абсолютно пассивный постоянный житель также может быть исключен из состава населения, если он активно не проявляет себя ни в одном из регистров в течение нескольких лет, т. е. если он не работает, не получает никаких пособий, не посещает медицинские учреждения и не нарушает никаких законов.

48. От преждевременного исключения из состава постоянного населения таких людей защищает стабилизирующий член индекса резидентства (см. формулу (1)).

49. Человек, проживающий за рубежом, также может оказаться включенным в число постоянных жителей, если он, находясь в Эстонии, активно пользуется услугами: посещает врача, покупает лекарства, совершает коммерческие сделки и т. д. Для предотвращения слишком раннего включения таких людей в число постоянных жителей используется так называемый период задержки.

VII. Заключение

50. В заключение следует отметить, что в случае использования индексной оценки численности населения ошибки также возможны, но вероятность их возникновения снижается с повышением качества регистров и задействования новых источников данных. Точность индексных оценок определяется с помощью дополнительных обследований, а результаты представляются с указанием величины возможной погрешности оценки. Включение новой информации (дополнительных признаков) позволит добиться стабильного повышения точности индексных оценок.

Библиография

- 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 Kvartalikiri*, 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.
- 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 Kvartalikiri*, nr 3, lk 79–95.
- Tiit, Ene-Margit (2011). 2011.aasta rahva ja eluruumide loenduse alakaetuse hinnang. *Eesti Statistika kvartalikiri* 4,12.
-