

Distr.: General
24 октября 2017

English

Европейская экономическая комиссия

Организации Объединенных Наций

Конференция европейских статистиков

Рабочая сессия по статистике миграции

Женева, Швейцария

30-31 октября 2017 года

Пункт 3 предварительной повестки дня

Интеграция данных и административные данные

Перепись населения 2016 года в Канаде: Интеграция административных данных по иммиграции

Подготовлено Статистическим управлением Канады*

Резюме

Перепись населения, являющаяся важным источником данных по демографическим и социально-экономическим характеристикам иммигрантов в Канаде, включает такие вопросы как год иммиграции, место рождения и гражданство. После сбора этих переменных от респондентов Статистическое управление Канады рассматривает данные по недостающим и непоследовательным ответам. В ходе переписи населения 2016 года была осуществлена увязка записей с административными данными по иммиграции в целях повышения качества традиционных переменных и добавления новых переменных, связанных с категорией приема иммигрантов (например, экономических иммигрантов, беженцев). В данном докладе представлен обзор того, как увязанные административные данные использовались для совершенствования методов обработки (редактирования и вменения) для ответов на вопросы по иммиграции. Кроме того, доклад посвящен методам и оценке качества, связанным с новыми переменными по категориям приема иммигрантов.

*Подготовлено Скоттом МакЛейхом, руководителем секции социальной статистики и статистики коренного населения, и Кэтрин Спес, аналитиком секции социальной статистики и статистики коренного населения



I. Предыстория: Вопросы иммиграции в переписи населения Канады

1. Перепись населения Канады является одним из основных источников данных по социально-экономическим характеристикам иммигрантов в Канаде. Проводимая каждые пять лет перепись включает несколько вопросов, связанных с иммиграцией, включая статус иммигранта, год иммиграции, место рождения, место рождения родителей и гражданство. Пользователи могут увязывать эти переменные с широким спектром тем, включая язык, образование, труд, доход и жилье, для удовлетворения своих потребностей в анализе. В последний раз иммиграционные переменные собирались в ходе переписи населения 2016 года, результаты которой планируется выпустить 25 октября 2017 года. Хотя вопросы по иммиграции не были включены в перепись 2011 года, они задавались в рамках Национального обследования домохозяйств (НОДХ) 2011 года, добровольного крупномасштабного обследования в рамках программы переписи.

A. Методы обработки данных для переменных по иммиграции

2. Для каждой переписи до сбора сведений вопросники подвергаются тщательной проверке и тестированию с целью обеспечения хорошего понимания вопросов и снижения до минимума нагрузки на респондентов (Статистическое управление Канады, 2017). Несмотря на эти меры, как и в случае с любым обследованием, респонденты не всегда отвечают на поставленные вопросы, и предоставляемые ответы не всегда точны. Для повышения согласованности между ответами применяются корректировки в целях обеспечения непротиворечивости, а для решения проблемы отсутствия ответов на те или иные вопросы применяются методы донорского вменения. Есть несколько возможных причин отсутствия ответов и представления неточных или непоследовательных ответов на вопросы по иммиграции. Можно привести следующие примеры:

- i. Респондент не понимает вопрос (например, год иммиграции в сравнении с годом приезда)
- ii. Респондент не помнит точный год иммиграции
- iii. Ошибки при представлении ответов по доверенности (человек, заполняющий вопросник не владеет всеми данными по каждому члену домохозяйства)

- iv. Ошибка ввода (например, неточное сканирование года иммиграции)
- v. Неоднозначные ответы, которые могут относиться к нескольким ответам (например, «Ирландия»)
- vi. Изменение географических структур (например, «СССР» в 1990 году в сравнении с «Россией» в 2011 году).

Методы обработки данных, используемые для переписи населения 2016 года, не новы; большинство из них уже применялись для нескольких переписей. Донорское вменение по ближайшему соседу выполнялось, используя Систему редактирования и вменения данных переписи населения Канады (CANCEIS) для расчета корректных ответов. В таблице 1 приведены доли вменения, примененные для НОДХ 2011 года для переменных по иммиграции. В целом доли составляли от 2,0% (место рождения) до 12,5% (год иммиграции).

Таблица 1: Проценты вменения, примененные для Национального обследования домохозяйств 2011 года по иммиграционным переменным

Переменная	Процент применения вменения (%)
Гражданство	2,3
Место рождения	2,0
Место рождения матери	5,7
Место рождения отца	6,0
Статус иммигранта	1,3
Год иммиграции	12,5

Источник Статистическое управление Канады, Справочное руководство по месту рождения, статусу поколения, гражданству и иммиграции, Национальное обследование домохозяйств 2011 года

3. В ходе всего процесса обработки данных корректировки по обеспечению непротиворечивости и результаты донорского вменения тщательно контролируются для обеспечения их надлежащего применения. Затем окончательные обработанные результаты проверяются, оцениваются и перепроверяются, используя другие источники, такие как административные данные, предыдущие переписи и прогнозы.

В. Корректировки по обеспечению непротиворечивости

4. Для вопросов по иммиграции в течение нескольких циклов переписи были разработаны и обновлены корректировки по обеспечению непротиворечивости. Они в значительной степени основываются на политике по иммиграции и гражданству в Канаде. Например, в вопросе о статусе иммигранта спрашивается: «Является ли этот человек сейчас или являлся ли он когда-либо прибывшим иммигрантом?» Если респондент отвечает «Нет», но при этом в вопросе о гражданстве отвечает, что является гражданином Канады **посредством натурализации**, а не **по рождению**, это будет отражать противоречие; для того, чтобы получить гражданство путем натурализации, человек сначала должен быть прибывшим иммигрантом. Однако исходя только из этих двух вопросов неясно на какой вопрос был представлен неправильный ответ. Для того чтобы определить какой ответ необходимо скорректировать, используются ответы на другие вопросы, такие как место рождения.

5. Еще одним примером противоречия является случай, когда респонденты сообщают о том, что они иммигрировали в Канаду за год до своего рождения, или сообщают, что они являются гражданами Канады по рождению и при этом дают положительный ответ на вопрос о статусе иммигранта. Для всех этих случаев выявленный ошибочный ответ исправляется на единственно возможный непротиворечивый вариант, либо остается незаполненным и отправляется на вменение.

6. Наиболее значительная корректировка по обеспечению непротиворечивости, применяемая к вопросам по иммиграции, касается респондентов, которые указывают себя гражданами Канады посредством

натурализации, родившихся за рубежом, но при этом неиммигрантами. В большинстве таких записей вносятся изменения, и респонденты классифицируются в качестве иммигрантов. Поскольку вопрос о годе иммиграции задается только тем респондентам, которые изначально дают положительный ответ на вопрос о статусе иммигранта, это означает, что допустимое значение года иммиграции необходимо будет вменить для всех случаев корректировки. Следствием этого является то, что процент вменения по вопросу о годе условного расчета гораздо выше процента вменения в отношении других вопросов по иммиграции, о чем свидетельствует Таблица 1.

С. Донорское вменение

7. Для записей с отсутствующими значениями (неудовлетворительных записей) в результате непредставления ответа или в результате редактирования, применяется донорское вменение по ближайшему соседу. В этом процессе используются доступные и коррелированные переменные (сопоставляемые поля) для выбора подходящих доноров для отсутствующих значений. В идеале выбранный донор должен соответствовать неудовлетворительной записи по всем полям. Однако зачастую такое невозможно, поэтому находится «ближайший сосед» путем максимально близкого соответствия по всем полям. Некоторым соответствующим полям придется большее значение, чем другим. Например, при вменении места рождения, поскольку оно больше связано с родным языком, более важно иметь соответствие по языку, нежели по возрасту.

8. Традиционно при вменении иммиграционных переменных сопоставляемые поля включали поля, указывающие возраст, пол, место проживания, родной язык, этнокультурные характеристики и какие-либо заполненные переменные по иммиграции. Вследствие важности семейных взаимоотношений при иммиграции вменение стратифицируется для учета таких внутрисемейных отношений. По мере возможности, вменение для пар осуществляется вместе, а вменение по детям осуществляется с учетом информации по их родным братьям/сестрам и родителям. Решения о том, какие переменные использовать в качестве сопоставляемых полей, во многом основываются на наблюдавшихся в прошлом связях и знании существа вопроса.

II. Административные данные по иммиграции в Канаде

9. Федеральным органом, отвечающим за вопросы иммиграции в Канаде, является Министерство по делам иммиграции, беженцев и гражданства Канады (IRCC). Оно собирает различные административные данные, связанные с иммигрантами, прибывающими в Канаду. В основном IRCC собирает информацию из заявки иммигрантов на получение визы на постоянное проживание в момент, когда иммигранты допускаются в Канаду. Эта информация включает год иммиграции, место рождения и категорию, в соответствии с которой иммигрантам был предоставлен вид на жительство (например, квалифицированные рабочие, члены семьи, беженцы и т.д.). Хотя данные административные данные IRCC предоставляют подробную

информацию об иммигрантах на момент их прибытия в Канаду, их нельзя использовать для оценки иммигрантов, проживающих в Канаде, в определенный момент времени, поскольку последующий сбор данных не производится. Однако эта информация может быть связана с другими наборами данных для связывания характеристик иммигрантов при приеме в страну с краткосрочными или долгосрочными результатами после их прибытия.

10. Статистическое управление Канады, в сотрудничестве с IRCC, ведет базу продольных данных по иммиграции (IMDB), которая связывает административную перепись иммигрантов, прибывших с 1980 года (используя данные IRCC) с ежегодными налоговыми отчетами с 1982 года (Evra, 2017). Эта база данных позволяет анализировать социально-экономические результаты иммигрантов в Канаду на длительной основе, с более чем 30-летним наблюдением за ранее прибывшим контингентом. Помимо обеспечения богатой информации, относящейся к доходам, в налоговом учете показывается проживают ли иммигранты в Канаде, а также модели внутренней миграции. С другой стороны, IMDB не обеспечивает прямые показатели по другим темам, включаемым в перепись населения, такие как образование в Канаде, труд, язык и жилье. Кроме того, поскольку результаты ограничиваются теми, которые можно найти в налоговом учете, дети и другие лица, которые не являются экономически активными лицами, в настоящий момент остаются неохваченными.

D. Связь записей с результатами переписи для оценки качества данных

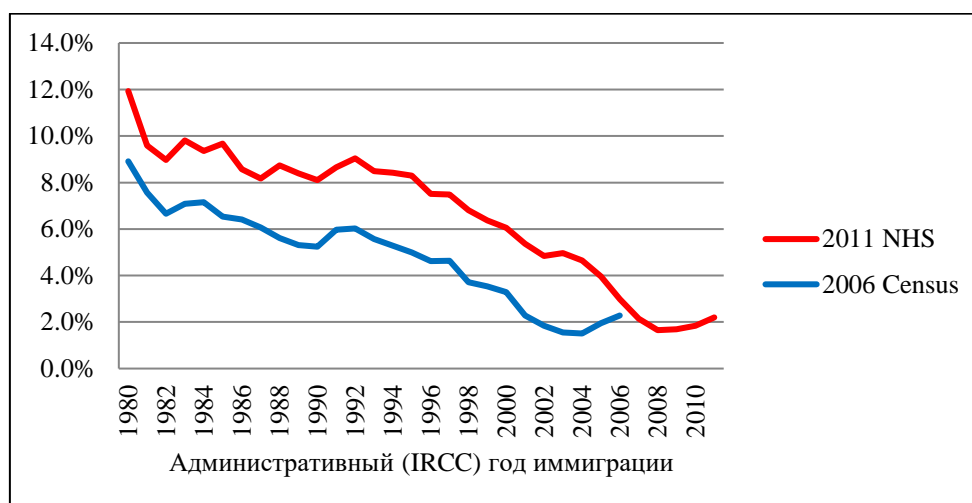
11. После проведения исследования осуществимости, включающего перепись 2006 года, НОДХ 2011 года было увязано с иммиграционными данными IRCC по тем лицам, кто прибыл в Канаду с 1980 года. Это было сделано для оценки качества данных НОДХ, где использовалось то же содержание по иммиграции, что и в предыдущих переписях (Brennan, 2011; Brennan, 2013; McLeish, 2012; McLeish, 2014). Хотя значительная часть проведенного анализа была посвящена воздействию непредставления ответов на добровольное НОДХ, такая увязка также обеспечила беспрецедентное изучение ошибок измерений, связанных с ответами на вопросы по иммиграции, и эффективности существующих методов обработки данных, включая донорское вменение.

12. По результатам анализа были внесены некоторые изменения в корректировки с целью обеспечения непротиворечивости и подходы к вменению, включая вменение неточных ответов, таких как «Ирландия». До этого для таких ответов обычно осуществлялось вменение с использованием донора с точными ответами в том же классе (например, потенциальные доноры для «Ирландии» были бы ограничены Республикой Ирландии и Соединенным Королевством). Однако проведенная связь показала, что при этом подходе игнорируется тот факт, что вероятность предоставления неточного ответа необязательно одинакова среди респондентов, родившихся в точных местах рождения, даже после проверки сопоставляемых полей. Например, респонденты, родившиеся в Республике Ирландия, с большей вероятностью ответят «Ирландия», чем те, кто родился в Соединенном

Королевстве. В результате данного анализа для этого вменения были использованы увязанные доли (Dobson, 2014).

13. Одним из важных обнаружений, сделанных в результате проведенного анализа, было то, что ответ по году иммиграции не отсутствовал случайным образом. Чем больше прошло времени с момента прибытия иммигрантов в Канаду, тем больше была вероятность того, что они дадут отрицательный ответ на вопрос о статусе иммигранта. На Графике 1 показана эта тенденция, определенная на основе связывания данных переписи 2006 года и НОДХ 2011 года. Как указывается в разделе В, несмотря на то, что эта ошибка в ответе корректируется детерминированно, исходя из ответов на другие вопросы, в отношении года иммиграции все же приходится делать вменение. При выборе донора для данного отсутствующего значения некоторые переменные, такие как возраст и место рождения, коррелируются с годом иммиграции. Однако пул подходящих доноров будет ограничен теми, кто ответил на этот вопрос. Это может привести к перевесу в пользу тех, кто прибыл относительно недавно.

График 1: Доля респондентов, связанных с административными данными по иммиграции, давшие отрицательный ответ на вопрос о статусе иммигранта, по связанному административному году иммиграции, переписи 2006 года и НОДХ 2011 года



Источник Статистическое управление Канады, Перепись 2006 года, связанная с административными данными IRCC по иммиграции, и НОДХ 2011 года, связанное с административными данными IRCC по иммиграции

14. В Таблице 2 показано соответствие между увязанными административными значениями и ответами НОДХ по году иммиграции до и после корректировок и вменения. Среди данных, увязанных с данными IRCC, корректировки в целях обеспечения непротиворечивости исправляют основную часть неправильных ответов (иммигрантов, которые отвечают, что они не являются иммигрантами). Однако если свыше 70% ответов, где не производилось вменение по году иммиграции, были равны увязанным административным данным, вмененные значения зачастую отличались более чем на 5 лет.

Таблица 2: Согласованность года иммиграции по данным НОДХ 2011 года с увязанными административными данными по году иммиграции до и после корректировок по обеспечению непротиворечивости и донорского вменения

Год иммиграции в НОДХ 2011 года		
Согласованность с увязанным административным значением	До	После
Тот же	70,9%	71,5%
Разница 1 год	8,8%	9,9%
Разница 2-5 лет	8,6%	11,4%
Разница более 5 лет	2,3%	5,8%
Неиммигрант	5,8%	1,3%
Отсутствует	3,7%	

Источник Статистическое управление Канады, НОДХ 2011 года, увязанное с административными данными IRCC по иммиграции

Е. Связь записей с результатами переписями в целях анализа

15. Помимо использования для оценки качества данных увязка записей из административных данных IRCC по иммиграции и НОДХ 2011 года также была использована для проведения анализа, связывающего характеристики приема иммигрантов (такие как категория приема) с результатами из НОДХ (McLeish, 2016). После обработки и взвешивания около 79% иммигрантов, прибывших после 1980 года в соответствии с НОДХ, были привязаны к административным записям. Получившаяся база данных была использована для выполнения специализированного табулирования без каких-либо корректировок на отсутствующие связи или несоответствия между административными значениями и ответами в рамках НОДХ.

III. Разработка переписи населения 2016 года

16. После оценки качества данных, выполненной для НОДХ 2011 года, было принято решение использовать увязанные административные данные непосредственно при обработке ответов на вопросы по иммиграции в рамках переписи 2016 года (Crowe, 2017). Несмотря на то, что какие-либо значения не будут детерминированно меняться в результате несовпадения с привязанным административным значением, (поскольку процесс привязки записей не без ошибок), привязка будет давать информацию для корректировок в целях обеспечения непротиворечия, и привязанные административные переменные будут использоваться в качестве сопоставляемых полей при выборе донора. Последнее является особенно важным, поскольку обеспечивает вспомогательные версии двух важных переменных иммиграции, которые могут потребовать вменения: место рождения и год иммиграции. Они будут более сильными ограничениями по

соответствиям, чем менее коррелированные переменные, такие как родной язык или место жительства.

Е. Разработка методов редактирования и вменения с использованием административных данных

17. В преддверии переписи населения 2016 года была создана тестовая среда с использованием данных НОДХ 2011 года, включая соответствующие увязанные административные переменные иммиграции. Поскольку процессы корректировки и вменения были пересмотрены и разработаны, они также были протестированы, используя эту среду. Это позволило лучше понять влияние введения административных переменных в процесс выборы донора или пересмотра корректировок в целях обеспечения непротиворечивости.

18. При оценке корректировок, направленных на обеспечение непротиворечивости, влияющих на то, следует ли классифицировать лиц в качестве иммигрантов, процентная доля привязанных записей дала косвенный показатель по количеству иммигрантов в определенной ситуации. Например, если связана небольшая доля респондентов в определенной непоследовательности, это может указывать на то, что в большинстве случаев эти лица не являются иммигрантами. Несмотря на то, что это было информативно и в некоторых случаях обеспечивало дополнительные детали, которые требовали изменений, корректировки с целью обеспечения непротиворечивости остались толкованием канадской политики по иммиграции и гражданству и не применялись только на основе результатов привязки записей.

19. Для вменения переменных по иммиграции увязанные переменные дали впечатляющие сопоставляемые поля, даже для тех респондентов, кто не ответил на несколько вопросов. В частности, использование этих значений повысило эффективность обращения с неигнорируемыми отсутствующими значениями по году иммиграции. С другой стороны, поскольку административные данные используются таким образом, их нельзя использовать в качестве независимого источника для перепроверки данных.

Г. Добавление переменных по категории приема иммигрантов

20. После увязки данных НОДХ 2011 года, IRCC предоставило финансирование Статистическому управлению Канады для включения новых переменных в перепись 2016 года, относящихся к категориям иммигрантов (McLeish, 2016). Эти переменные будут поступать непосредственно из административных данных IRCC, увязанных с переписью 2016 года по тем иммигрантам, которые прибыли после 1980 года (согласно их переписным записям). Включенными переменными были категория приема иммигранта и тип заявителя. Эти категории были определены для того, чтобы можно было провести детальный анализ конкретных программ иммиграции (т.е. иммиграция для специалистов, имеющих опыт работы в Канаде, провинциальные программы и программа для беженцев, направляемых в визовый центр), и при этом позволили классифицировать их по более обширным группам (т.е. экономические иммигранты, иммигранты,

спонсируемые членами семьи, беженцы и прочие иммигранты). Эти категории были разработаны в сотрудничестве с партнерами из IRCC.

1. Разработка корректировок с целью обеспечения непротиворечивости для новых переменных

21. Хотя новые увязанные переменные не подвержены тем же видам ошибок в ответах, как переменные, получаемые из вопросов переписи по иммиграции, имеют место ошибки в самом процессе увязки записей, что приводит к ложноотрицательным и ложноположительным результатам. Кроме того, увязанные переменные должны не противоречить традиционному содержанию переписи для завершения процесса их интеграции в качестве новых переменных переписи. В силу этого необходимо было разработать и применить корректировки, направленные на обеспечение непротиворечивости.

22. В отличие от традиционных переменных переписи по иммиграции, корректировки с целью обеспечения непротиворечивости для новой переменной по категории приема иммигрантов ранее не применялись. Поскольку эти переменные имеют политическую направленность, и, следуя подходу, применяемому к другим переменным, корректировки с целью обеспечения непротиворечивости были созданы на основе иммиграционной политики Канады. Под руководством партнеров из IRCC, правила непротиворечивости обеспечивали применение к увязанным данным сроков и требований программ. Например, респондент, сообщивший о своем прибытии в 1994 году (или это значение было вменено) не мог иметь значение увязанной категории приема по программе, которая не существовала в этом году. Эти корректировки, направленные на обеспечение непротиворечивости, были протестированы, используя увязанные результаты НОДХ 2011 года.

2. Разработка методов вменения для новых переменных

23. Частично в связи с ложноотрицательными результатами в процессе увязки записей и частично в связи с ошибками в ответах на вопросы переписи и ошибками вменения, не все иммигранты, прибывшие, согласно переписи, после 1980 года, увязаны с административными записями. Кроме того, для каждого случая, когда корректировка, направленная на обеспечение непротиворечивости, определяла, что увязанная категория приема иммигранта была неверной, требовалось новое (и верное) значение по категории приема. Эта задача была выполнена при помощи тех же методов донорского вменения, которые использовались для традиционных иммиграционных переменных. Учитывая, что для увязки НОДХ 2011 года требовалось вменение на уровне 21%, ожидалось, что этот процесс должен определить конечное качество данных по этим новым переменным (McLeish, 2016).

24. Как и в случае с корректировками с целью обеспечения непротиворечивости, ранее не проводилось определение стратегии вменения, сопоставляемых полей и их относительной важности. Используя подход,

схожий с тем, который использовался для традиционного содержания по иммиграции, вменение было стратифицировано для того, чтобы максимально учесть семейные сведения (вменение по лицам, проживающим в паре, производилось на основании информации по супругам, а по детям – на основе характеристик их родных братьев/сестер и родителей). Это объясняется результатами увязки данных НОДХ 2011 года, которая показала сильные внутрисемейные корреляции по категориям приема иммигрантов (Martin, 2017).

25. В каждой типической группе для определения того, какие сопоставляемые поля необходимо использовать и их относительной важности для процесса выбора донора, был использован алгоритм ReliefF (Kononenko, 1994; Kira and Rendell, 1992). ReliefF работает при допущениях схожих с теми, которые применялись в рамках методов вменения по донору, являющемуся ближайшим соседом, для переписи населения в Канаде. Алгоритм предполагает, что записи со схожими сопоставляемыми полями должны происходить из одного класса, а записи с разными сопоставляемыми полями должны происходить из разных классов. Сопоставляемые поля, которые оказываются сходными внутри классов и разными между классами, получают более высокий вес. В Таблице 3 приведены результаты тестовой среды для НОДХ 2011 года по лицам, живущим в паре, у которых указан тот же год иммиграции, что и у их супругов. Результаты алгоритма согласуются с предметными знаниями о том, что переменными, которые являются наиболее важными для прогнозирования категории приема иммигрантов для той или иной типической группы, являются категория приема супруга/супруги и тип заявителя, а также место рождения респондента и год иммиграции.

Таблица 3: Полученные веса ReliefF для сопоставляемых полей с наибольшим весом для вменения категории приема иммигранта для лиц, живущих в паре с тем же годом иммиграции, что и их супруга/супруги, тестовая среда НОДХ 2011 года

Сопоставляемые поля	Вес
Категория приема супруга/супруги	0,674
Тип заявителя супруга/супруги	0,329
Место рождения	0,284
Год иммиграции	0,18
Возраст при иммиграции	0,127
Уровень образования	0,0999
Род деятельности	0,0832
Провинция проживания	0,0713
Официальные языки	0,04
Пол	0,0387

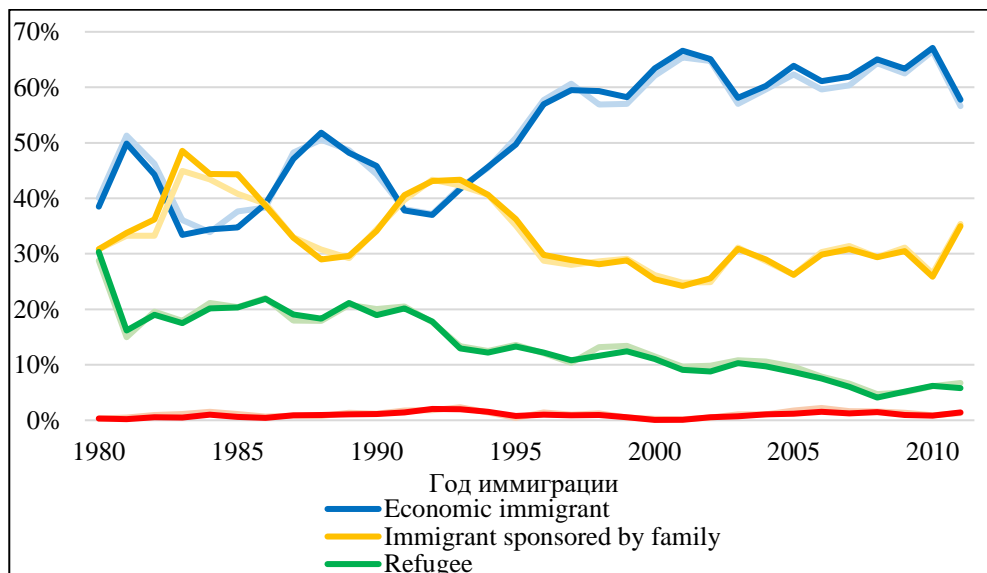
Источник Статистическое управление Канады, тестовые данные редактирования и вменения, увязка данных Национального обследования домохозяйств 2011 года и административных данных по иммиграции IRCC

Н. Проверка и подтверждение разработанных методов вменения

26. Для оценки качества методики вменения случайным образом выбранные подмножества полных записей из тестовой среды были обнулены и затем отправлены на вменение. Их вмененные значения затем можно было сравнить с известными ответами. Хотя это и свидетельствует о том, как вменение позволяет решать проблему с отсутствующими данными при разных допущениях, оно тем не менее основывается на ответах (или неотсутствующих значениях).

27. Для вновь разработанных методов донорского вменения для категорий приема иммигрантов этот анализ показал, что вменение не только поддерживало маргинальное распределение, но и важные кросс-табличные корреляции (например, по году иммиграции). На Графике 2 показано распределение по категориям приема иммигрантов по году иммиграции на основе исходных (более светлые тона) и на основе вмененных данных (более темные тона) среди взрослых.

График 2: Исходное и вмененное распределение категорий приема иммигрантов по одному году иммиграции, тестовая среда НОДХ 2011 года



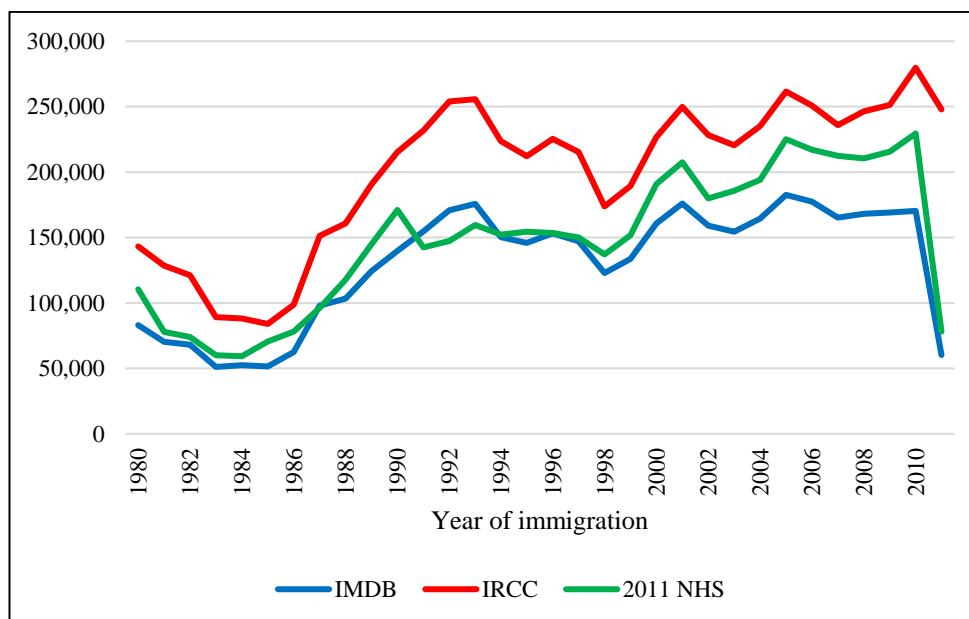
Источник Статистическое управление Канады, данные тестового редактирования и вменения, увязка данных Национального обследования домохозяйств 2011 года и административных данных по иммиграции IRCC

28. Одним из недостатков использования увязанных административных данных в процессе вменения является их утрата в качестве независимого источника для перепроверки результатов. Тем не менее, административные данные IRCC (без увязки записей) и IMDB могут использоваться для получения верхней и нижней границ окончательных оценочных значений переписи соответственно. Данные IRCC обеспечивают необходимую верхнюю границу, поскольку являются административной переписью всех иммигрантов, прибывающих в Канаду, но не учитывают случаи смерти или эмиграции. IMDB может быть использована для получения оценочного количества резидентов налогоплательщиков в последний налоговый год. Их

число должно быть меньше, чем численность иммигрантов по переписи, поскольку не учитываются дети и другие лица, не являющиеся налогоплательщиками. Эти два источника являются полезными, поскольку содержат сходные сведения, такие как год иммиграции, место рождения и сейчас – категорию приема иммигранта. Несмотря на то, что расхождения, возникающие в результате ошибок в ответах и при вменении, могут влиять на сопоставления по году иммиграции и месту рождения, данные по категории приема иммигрантов поступают из одного источника для всех трех источников данных, при этом расхождения могут быть вызваны ошибками увязки, расхождениями в охвате и вменением для данных переписи.

29. На Графике 3 показано как данные IRCC и IMDB могут быть использованы для проверки конечных результатов переписи. График показывает сравнение распределения иммигрантов за один год иммиграции в этих источниках данных. Как и ожидалось, данные IRCC обеспечивают верхнюю границу, а данные по налоговому учету 2010 года для IMDB обеспечивают нижнюю границу. В целом тенденции, связанные с активизацией и затуханием иммиграции в Канаду, последовательно отражаются в трех источниках. Этот анализ может быть повторен для того, чтобы сосредоточиться на определенных контингентах, местах рождения, возрастных группах, поле и категориях приема иммигрантов.

График 3: Оценочные данные по годам иммиграции для иммигрантов, прибывших в Канаду после 1980 года, административные записи IMDB, IRCC, Национальное обследование домохозяйств 2011 года



Источник Статистическое управление Канады, 2014 IMDB – иммигранты, подавшие налоговые декларации в 2010 году, Национальное обследование домохозяйств 2011 года; Министерство Канады по делам иммиграции, беженцев и гражданства, административные данные по иммиграции

IV. Заключение и следующие шаги

I. Использование административных данных позволило улучшить методы редактирования и вменения

30. На момент подготовки данного доклада окончательные результаты переписи 2016 года еще не были опубликованы (переменные по иммиграции будут опубликованы 25 октября 2017 года). Однако при тестировании и разработке было установлено, что интеграция административных данных позволила повысить качество корректировок с целью обеспечения непротиворечивости и донорского вменения по традиционным переменным по иммиграции. Корректировки с целью обеспечения непротиворечивости были более непосредственно сопоставлены с использованием увязки данных в качестве косвенного показателя по статусу иммигранта, а донорское вменение было улучшено за счет использования более непосредственно скоррелированных вспомогательных переменных, таких как увязанный административный год иммиграции и увязанное административное место рождения.

31. Увязка записей, произведенная для переписи 2016 года, была расширена для включения учетных записей по иммиграции с 1952 года, а также записей по просителям убежища и временным резидентам. Это позволит улучшить обработку данных за рамками того, чтобы было протестировано, используя результаты НОДХ 2011 года (Biernot, 2017).

32. Для будущих переписей можно будет использовать другое содержание из увязанных административных данных, придерживаясь схожих подходов, для того, чтобы обеспечить правильность корректировок с целью обеспечения непротиворечивости и учета наиболее актуальной имеющейся информации в процессе выбора донора для вменения.

J. Разработаны методы редактирования и вменения для переменных по категории приема иммигрантов

33. Сотрудничество между Статистическим управлением Канады и IRCC привело к большей обстоятельности корректировок с целью обеспечения непротиворечивости для новых переменных по категории приема иммигрантов. Такие корректировки были протестированы, используя результаты увязки данных НОДХ 2011 года, и затем протестированы повторно, используя результаты увязки данных переписи 2016 года.

34. Поскольку основную обеспокоенность, связанную с переписью 2016 года, представляла степень надежности донорского вменения для переменных по категории приема иммигрантов, тестирование показало, что разработанные методы вменения позволили получить надежные данные, соответствующие административным данным IRCC и IMDB. Использование ReliefF для определения сопоставляемых полей и их относительной важности в процессе выбора доноров для вменения обеспечило статистическое основание для данного процесса, вместо того, чтобы просто полагаться на профильные знания. Кроме того, показатель степени увязки данных переписи 2016 года был гораздо показателем, наблюдавшегося для НОДХ 2011 года, это

приведет к более низкой доле вменения, чем первоначально ожидалось для этих переменных (Biernot, 2017).

35. В целом, внедрение этих двух новых переменных открывает новые возможности для анализа социально-экономических результатов иммигрантов в Канаду, связывая огромный объем информации, которая традиционно собирается в рамках переписи населения, с информацией IRCC по программе приема иммигрантов. Были разработаны и протестированы методы для подготовки окончательных оценок для того, чтобы гарантировать хорошее качество данных, используемых в анализе.

V. Выражение признательности

36. Данный проект был бы невозможен без поддержки и сотрудничества со стороны многих специалистов, представляющих различные области знаний, в том числе профильных аналитиков, методологов и администраторов баз данных. В частности, Статистическое управление Канады выражает признательность за оказываемую поддержку Лорне Янцен и ее коллегам из IRCC, а также за их партнерство при включении переменных по категориям приема иммигрантов в программу переписи населения, а также за постоянную помощь по толкованию своих административных данных. В Статистическом управлении Канады авторы выражают особую благодарность Eric Mongrain, Laetitia Martin, H el ene Maheux, Mireille V ezina, Chantal Poirier, Hyunji Lee и Jarod Dobson из отдела социальной статистики и статистики коренного населения; Andrew Stelmack, Sean Crowe и Lyne Guertin из отдела методологии социальных обследований; Piotr Biernot, Caroline Pelletier, Paul Cascagnette и Colin Babyak из отдела методологии обследований домохозяйств, и многим другим лицам, внесшим вклад в разработку и производство иммиграционной составляющей переписи населения 2016 года.

VI. Библиография

- Biernot, P. 2017. "External linkage between the Immigration File (1952-2016) and the 2016 Census Response Database." Unpublished report. Ottawa: Statistics Canada.
- Brennan, J. 2011. "CIC Landing File to Census 2006 Linkage." Unpublished report. Ottawa: Statistics Canada.
- Brennan, J. 2013. "CIC Landing File to Census 2011/NHS Linkage." Unpublished report. Ottawa: Statistics Canada.
- Crowe, S. and Janes, D. 2017. "Edit and Imputation Report: Ethnocultural Process." Unpublished report. Ottawa: Statistics Canada.
- Dobson, J., Costa, R., and Martin, L. 2014. "Narrative of Editing and Imputation Procedures: EC Variables." Unpublished report. Ottawa: Statistics Canada.
- Evra, R. and Prokopenko, E. 2017. "Longitudinal Immigration Database (IMDB) Technical Report, 2014." *Analytical Studies: Methods and References*. Catalogue no. 11-633-x2017007. Ottawa: Statistics Canada.

- Guertin, L., Bureau, M., and Morel, J. 2014. "Editing the 2011 Census data with CANCEIS and options considered for 2016." *2014 Work Session on Statistical Data Editing*. Conference of European Statisticians. United Nations Economic Commission for Europe.
- Kira, K., Rendell, L. 1992. "The Feature Selection Problem: Traditional Methods and a New Algorithm." *AAAI-92 Proceedings*, 129-134.
- Kononenko, I. 1994. "Estimation Attributes: Analysis and Extensions of RELIEF." *Proceedings of the 1994 European Conference on Machine Learning*. Catania, 6-8 April 1994, 171-182.
- Martin, L. 2016. "Compte rendu des procédures de contrôle et imputation pour le processus catégorie d'admission (AC)." Unpublished report. Ottawa: Statistics Canada.
- Martin, L. 2017. "From the source to dissemination: processing, accessibility and possible uses of the Census admission category of immigrants" Presentation delivered to the 19th Canadian National Metropolis Conference. Ottawa: Statistics Canada.
- McLeish, S. 2012. "Feasibility Study: IMDB for Statistical Purposes." Unpublished report. Ottawa: Statistics Canada.
- McLeish, S. 2014. "Using administrative data to evaluate data quality: Immigrants in the 2011 National Household Survey", *Proceedings of Statistics Canada Symposium 2013*, Catalogue no. 11-522-X. Ottawa: Statistics Canada.
- McLeish, S. 2016 "Adding Immigrant Admission Category to the Canadian Census of Population" *2016 Work Session on Migration Statistics*. Conference of European Statisticians. United Nations Economic Commission for Europe.
- Statistics Canada. 2013. "Place of Birth, Generation Status, Citizenship and Immigration Reference Guide." 2011 National Household Survey. Catalogue no. 99-010-XIE2011008. Ottawa: Statistics Canada.
- Statistics Canada. 2014. "CANCEIS, version 5.2. Basic User Guide." Ottawa: Statistics Canada.
- Statistics Canada. 2016. "Dictionary, Census of Population, 2016" 2016 Census of Population. Ottawa: Statistics Canada.
- Statistics Canada. 2017. "Guide to the Census of Population, 2016" 2016 Census of Population. Catalogue no. 98-304-X. Ottawa: Statistics Canada.
-