

Использование административных данных для оценки численности населения

Предварительный анализ первого
полученного опыта

Глэдис Массе и Иван Уильямс

(gmasse@indec.gob.ar & iwilliams@indec.Gob.ar)

indec

Instituto Nacional de
Estadística y Censos
República Argentina

Краткое описание

- Общая информация
- Первый опыт
- Выводы
- Долгосрочный проект и дальнейшая работа

Общая информация

- Пандемия COVID-19 отложила завершение Национальной переписи населения, домохозяйств и жилищного фонда 2020 года.
- В этой ситуации ускорился поиск различных альтернатив:
 - Цифровая перепись;
 - Использование административных регистров.
- Национальный институт статистики и переписей (INDEC) начал работу над этой темой в 2016 году.
- Долгосрочный проект: Национальный регистр населения.

Но пока необходимо достигнуть некоторого существенного прогресса, и производство демографической статистики может быть в этом полезным.

Опыт (1)

Налоговая и социальная информационная система

- Многосекторальный регистр: предоставляет социально-экономическую информацию по государственным учреждениям.
- Банковская информация как наиболее полный источник данных о численности населения в возрасте от 18 лет и старше по сравнению с пробной переписью в провинции Мисьонес.
- Проверка на сходство адресов из выборки микроданных:
 - Пять разных адресов из 5 разных административных регистров.
 - У 90% людей было по **два** совпадающих адреса, указанных в 5 источниках данных.
 - У 60% людей были **все** совпадающие адреса, указанные в 5 источниках данных.

Опыт (2)

Федеральное управление государственных доходов

- Налоговый регистр ~ 52 млн записей.
- Разрозненная кодификация адресов и административных территорий.
- Пополнение данных регистров с помощью полевых данных переписи, полученных в ходе еще одной пробной переписи, проведенной в 2019 году в городе Умауака, Жужуй:
 - Переменные соответствия: Идентификационный номер, возраст, дата рождения, пол, ФИО.
 - 80% совпали по идентификационному номеру.
 - 70% были точными совпадениями со всеми переменными.



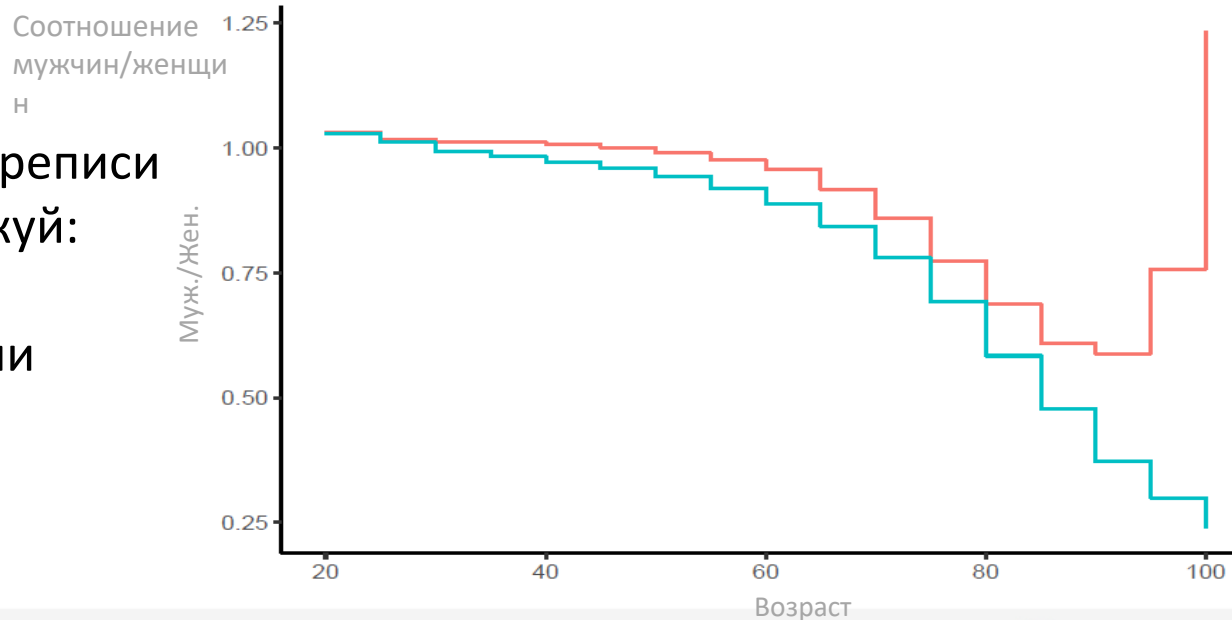
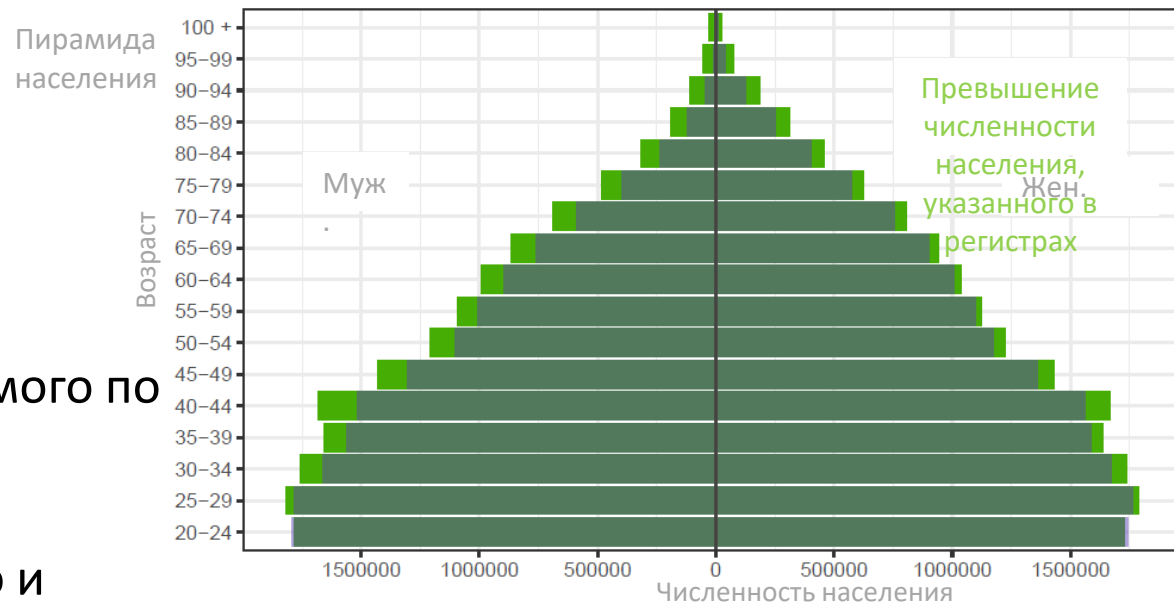
Частота появления «Умауака» (Humahuaca) в поле «название населенного пункта»

Город	n
humahuaca	15.325
ciudad humahuac	218
1032411 humahua	49
humauaca	6
humahuca	4
humahuaca0	3
32411n humahuac	2
dpto. humahuaca	2
humahuacaа	2
1032405 humahua	1
1032411-humahua	1
1032411 humauac	1
1032411 humauca	1
1032412 humahua	1
1032414 humahua	1
1032419 humahua	1
32411 humahuaca	1
el aguilar huma	1
huamahuaca	1
humahua	1
humahuac	1
humahuaca4630	1
humahuacva	1
humauhaca	1
jumahuaca	1
zo) humauaca	1

Опыт (3)

Регистр избирателей

- Не включено: возраст <16 и иностранцы.
- Соотношение мужчин и женщин выше ожидаемого по демографическому прогнозу.
- Превышение численности населения взрослого и пожилого возраста (особенно долгожителей).
- Пополнение данных регистров с помощью переписи
Пробная перепись в 2019 году в Умауаке, Жужуй:
 - 70% показали точное совпадение.
 - 80% имеющих точное совпадение приняли участие в переписи в той же провинции и населенном пункте, где они голосуют.

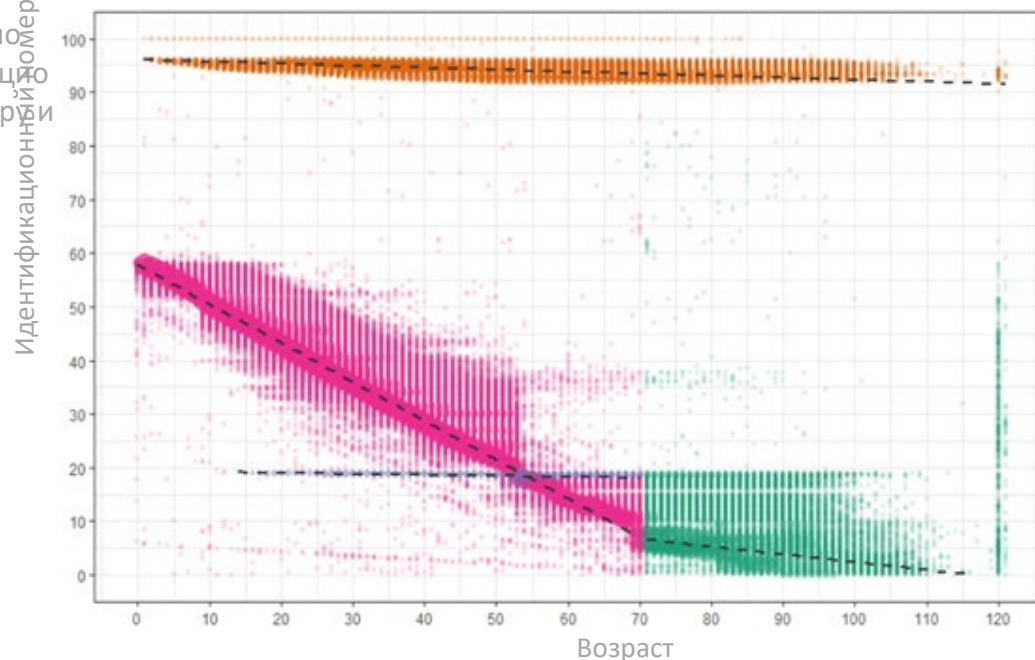


Опыт (4)

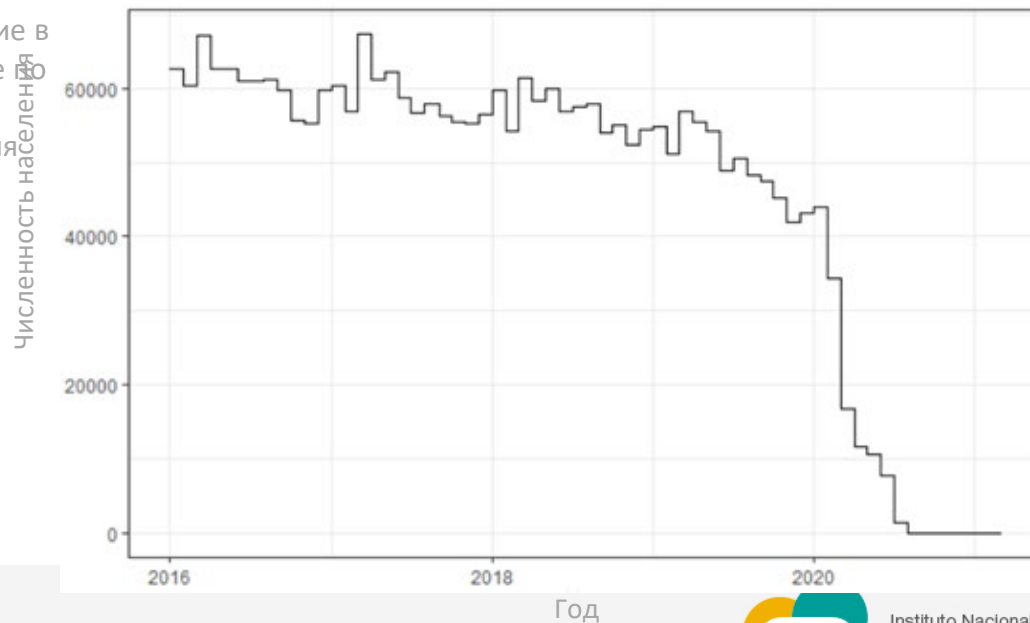
Национальный регистр физических лиц

- Ведомство, присваивающее идентификационный номер (ID) новорожденным и иммигрантам.
- 47,5 млн. (на апрель 2021 г.) по сравнению с прогнозом — 45 млн. (на июнь 2021 г.).
- Кластеры в идентификационной нумерации, связанные с возрастом:
 - Сильная корреляция до 70 лет.
 - Группа иностранцев: ID > 90 MM.
 - Обособленные группы.
- Статистика рождений и смертей отстает примерно на 1 год:
 - Проблемы с федеральной регистрационной системой.
 - Пандемическая ситуация.

Население по
идентификационному номеру
возрасту



Население в
разбивке по
годам
рождения



Первые выводы

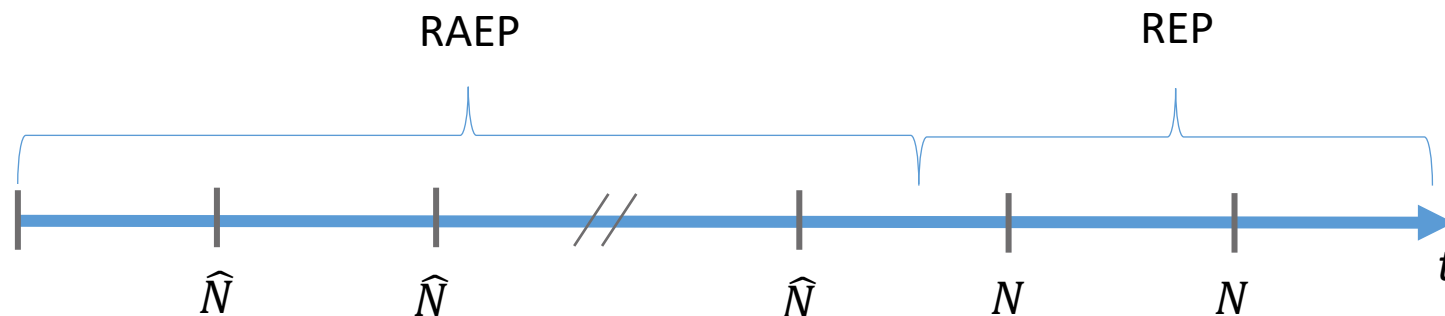


- Пересчет общей численности населения и лиц пожилого возраста.
- Недостаточный учет несовершеннолетних.
- Неопределенная задержка между событием и появлением записи в регистре.
- Больше мужчин, чем ожидалось (выбранная миграция?).
- Разрозненная кодификация адресов и административно-географических территорий.
- 70% точных совпадений с пробными переписями (в среднем).

Положительные аспекты:

- Мы провели краткую диагностику по основным административным регистрам Аргентины.
- Мы обнаружили интерес со стороны держателей административных регистров к совместной работе и к запуску совместного проекта.

Общий проект



- Административные записи для оценки численности населения (RAEP): связанный набор записей о населении с неполным охватом / качеством данных.
- Оценка численности населения (\hat{N}): оценка численности населения в разбивке по возрасту, полу и месту жительства на базе RAEP (Административные записи для оценки численности населения) и других источников.
- Статистический регистр населения (REP): полный, непрерывный и обновляемый список людей, проживающих в стране.
- Перепись на основе записей (N): подсчет численности людей в разбивке по возрасту, полу, месту жительства на базе REP (Статистический регистр населения).

Следующие этапы

- Получить полные характеристики каждого административного регистра.
- Диагностировать сильные и слабые стороны, чтобы получить информацию о населении, проживающем в стране.
- Сгенерировать первые экспериментальные национальные оценки численности населения.



Мы совершаем необходимое и суперинтересное путешествие, которое продолжается!

Спасибо!

indec