



ЭКОНОМИЧЕСКИЙ И СОЦИАЛЬНЫЙ СОВЕТ

Распространение: Общее
30 апреля 2012 г.
Оригинал: английский

Европейская экономическая комиссия ООН

Конференция европейских статистиков

60-е пленарное заседание

Париж, 6-8 июня 2012

2010 раунд переписи – инновации и уроки

Общий подход к использованию регистров при проведении переписи в Испании

Записка Национального института статистики Испании

Резюме

В данном документе описана Перепись 2011 г. в Испании, которая основывалась на комбинированном подходе, включавшем крупное выборочное обследование 12% населения (3 млн. жилищ и 5,8 млн. человек) и использование административных регистров.

В записке со статистической точки зрения объясняются сложности и преимущества такого подхода. Во-первых, необходимо провести глубокое исследование различных регистров и статистических источников и проанализировать их характеристики, такие как точность, охват и соответствие статистических и административных концепций. Необходимо определить разные стратегии в целях увязки всей этой информации. Кроме того, необходимо учесть другие аспекты, такие как несоответствия между разными источниками данных и то, насколько детальной является распространяемая информация.

В данном документе обобщены эти вопросы и подчеркнуты важные сильные и слабые стороны применяемой методологии, исходя из опыта Национального института статистики Испании.

I. Регистры в Испании

1. Два самых важных источника данных для Переписи 2011 в Испании это - административный регистр населения (PADRON) и Перепись населения и жилого фонда 2001 года.
2. PADRON - это административный регистр, управление которым осуществляется в соответствии со специальным законодательством. Операционная часть состоит из центральной базы данных, которая находится в Национальном институте статистики (INE), куда все муниципалитеты, отвечающие за регистрацию своих жителей, присылают изменения, произошедшие в их местных регистрах. INE, совместно с муниципалитетами и другими местными органами, работает над улучшением качества регистра, удаляя дублирующие записи, передавая информацию об изменении места жительства между

муниципалитетам, чтобы избежать дублирования, удаляя из регистра лиц, переехавших за границу и проч. Одним из главных недостатков регистра PADRON является отсутствие в нем демографических переменных. В регистре населения присутствует только такая базовая информация, как дата рождения, место рождения и гражданство¹.

3. Существуют некоторые проблемы в части использования регистра PADRON как прямого источника статистических данных. PADRON – это административный регистр, и регистрация в нем предоставляет людям основные права, такие как участие в голосовании, доступ к медицинским и образовательным услугам. Поэтому INE не может решать, включить кого-то в регистр или удалить из него только на основании статистических критериев. Это означает, например, что удаление из регистра человека, который переехал за рубеж, не уведомив муниципалитет, может занять несколько месяцев, так как нужно выполнить все шаги, предусмотренные законодательством. Для целей переписи, информация регистра нуждается в некоторой статистической обработке.

4. PADRON является важным источником получения демографической информации о населении. Хотя он содержит информацию о том, где живут люди, эта информация не организована с вычислительной точки зрения. Другими словами, это – регистр лиц, а не жилых помещений. Из-за этого основа выборки жилых помещений берется из Переписи 2001 г.

5. Начиная с 2001 г. эти списки жилых помещений обновляются с использованием информации из регистра PADRON обо всех людях, которые приезжают или переезжают в новые помещения. Таким образом, для целей Переписи 2011 у нас есть основа, которая содержит:

- (а) Информацию о жилых помещениях, которые существовали в Переписи населения и жилого фонда 2001;
- (б) Информацию о жилых помещениях, которые были заселены как минимум один раз между 2001 и 2011 годами.

6. Эта основа нуждается в обновлении для включения тех жилищ, которые были построены в период с 2001 и с тех пор не были заселены. По этой причине и для того, чтобы собрать информацию обо всех географических координатах зданий, мы провели почти исчерпывающую перепись зданий 2011 года².

7. Другими важными регистрами для переписи 2011 г. являются регистр национальных удостоверений личности и карт вида на жительство (они содержат важную информацию для верификации личности проживающего), бюллетени актов гражданского состояния (они позволяют проверить информацию о рождениях, смертях и заключении браков), регистры налоговых органов и органов социального обеспечения.

8. В табл.1 приведены регистры, которые использовались для Переписи 2011 в Испании.

Таблица 1

Регистры для Переписи 2011

<i>Название регистра</i>	<i>Ответственный за ведение регистра</i>
Регистр населения	INE
Перепись 2001	INE
Национальные удостоверения личности и карты постоянного проживания	Министерство внутренних дел
Кадастр: регистр жилых помещений	Казначейство
Налоговое управление	Казначейство
Социальное обеспечение	Министерство труда и социальных вопросов
Бюллетени актов гражданского состояния: рождения, смерти, браки	INE

¹ Также имеется пункт об уровне образования, но информация не очень точная.

² Для 20% счетных участков информация была получена непосредственно из обновленных списков жилых помещений, а координаты были получены из кадастра.

II. Методология в Испании: комбинированный подход

9. Переписи являются самой сложной задачей, стоящей перед статистической службой. Перепись – это не только мероприятие по сбору данных, но и логистическая операция. Для того чтобы показать это, скажем, что бюджет переписи 2001 г. в Испании составлял более 200 млн. евро (сегодня это было бы свыше 300 млн. евро), в проведении переписи было задействовано более 40 тыс. человек, включая переписчиков, руководителей групп и т.д. Более 70 тыс. контрактов было подписано в целях обеспечения замены в случае, если кто-то решал не продолжать работу.

10. Перепись 2011 впервые проводится в условиях, когда существует европейский регламент, в котором установлены общие рамки для проведения переписей населения и жилого фонда. Этот регламент содержит описание различных источников данных, которые могут быть использованы для целей переписи. Мы учитываем это во внимание, когда принимаем решения по поводу того, какие из этих источников наиболее подходят для нашей переписи.

11. Показатель численности населения является одним из важнейших результатов переписи. Кроме этого важны и другие виды информации в зависимости от интересов различных пользователей: демографические и социальные показатели, информация о жилых помещениях и зданиях и многое другое.

12. При определении численности населения следует принимать во внимание административный регистр населения в связи с его свойствами, которые были коротко описаны выше.

не должен использоваться без проверки содержащейся в нем информации. Данные административного регистра PADRON увязываются с данными из других служб социального обеспечения или налоговой службы, для проверки в Испании. Используя такой подход, мы можем быть уверены в том, что PADRON относится к лицам, действительно проживающим в Испании. Такая информация (пол, место рождения, местожительства и т.д.), в отношении которых не ясно, проживает ли в Испании, в возрасте 65 лет или старше, зарегистрированной в PADRON, не используется (например, в службе социального обеспечения или в налоговой службе).

Мы не сконцентрируем усилия на тех лицах, в отношении которых есть сомнения, а не на тех, в отношении которых такие сомнения отсутствуют. Это позволяет проводить полную перепись населения.

Административные регистры, которые могут обеспечить сведения о населении. Как отмечалось ранее, PADRON содержит информацию только о населении. Недоступна информация о жилье, которое было построено после 1980 года. Однако же имеет настолько отличную структуру данных, что наши усилия дали приемлемые результаты только в 20% переписных пунктов. Поэтому это решение провести исчерпывающую перепись зданий и

- (i) для оценки размера групп записей, относительно которых не ясно, проводить ли учет;
- (ii) для получения информации о переменных, которые не включены в предпереписной файл;
- (c) Исчерпывающая перепись зданий, которая поможет:
 - (i) сформировать полный список зданий (для использования в качестве основы выборки для будущих статистических обследований);
 - (ii) получить географические координаты зданий.

17. Информация поступала по нескольким каналам: через Интернет, по почте и в результате проведения персональных интервью при помощи компьютера (CAPI). Прежде всего, в заселенные жилища направлялось письмо с предоставлением возможности ответить через Интернет. Тем, кто не воспользовался этой возможностью, направлялось второе письмо, включающее анкету в бумажном виде, и предоставлялось два варианта: ответить через Интернет или направить анкету по почте. Наконец к тем, кто не предоставил ответы, были направлены переписчики для проведения интервью с использованием карманных компьютеров.

III. Увязка информации между регистрами

18. Высокое качество регистра PADRON позволяет ему играть главную роль, а информация из других регистров привязывается к нему для целей проверки. PADRON содержит идентификационные данные по каждому человеку, но эта информация хранится в трех разных полях: номер национального идентификационного номера (DNI), номер паспорта и номер карты вида на жительство. Личные данные для каждого человека, зарегистрированного в PADRON, будут содержаться в одном из этих разделов.

19. PADRON содержит информацию о личном идентификационном номере для каждого человека. Хотя и редко, но иногда можно обнаружить двух и более человек с одинаковыми или неправильными личными идентификационными номерами или даже без них. Для того чтобы решить эту проблему и сопоставить данные, имеющиеся в регистре PADRON, необходимо увязать информацию из регистра PADRON с информацией Министерства внутренних дел. Министерство внутренних дел является органом, ответственным за выдачу национальных удостоверений личности, паспортов и вида на жительство

20. Для того чтобы увязать два источника данных, необходимо полное совпадение в этих двух источниках индивидуального идентификационного номера, имени, фамилии, пола, даты и места рождения. По тем записям, где такое совпадение не является идеальным (возможно, информация в одном или обоих источниках является неверной), применяются вероятностные алгоритмы с внутренним порогом или контрольным параметром, используемые для сравнения сходных элементов между людьми. Для граждан Испании увязка личных идентификационных номеров в записях в регистре PADRON с базой данных Министерства внутренних дел составляет 99%: для иностранцев – 80%.

21. Такой же процесс применяется для других административных источников для того, чтобы убедиться в том, что информация об идентификационных номерах хорошего качества во всех используемых регистрах.

22. Информация из регистра PADRON и других регистров будет увязана, если существует совпадение идентификационных данных. Для тех записей, где этот метод не работает, используются вероятностные методы для повышения процента увязки.

IV. Как получить оценку численности населения

23. Как отмечалось ранее, предпереписной файл содержит записи из PADRON, увязанные с другими административными регистрами. Для получения оценки численности населения мы присваиваем каждой записи счетный фактор – вес каждой записи при расчете общей численности населения.

24. Прежде всего, должны засчитываться те записи, которые имеются в регистре PADRON и также были найдены в других регистрах. В этом случае мы присваиваем таким записям счетный фактор, равный 1.

25. Во-вторых, если запись имеет привязку к какому-либо регистру смертей, предоставленному регистром актов гражданского состояния, то мы присваиваем такой записи счетный фактор, равный 0.

26. И, наконец, если запись не была увязана ни с одним административным регистром, необходимо проверить, отмечена ли она каким-либо «предостерегающим» знаком³. Если запись отмечена «предостерегающим» знаком, то она будет иметь условный счетный фактор «X», где X имеет значение больше 0, которое должно рассчитываться с применением результатов обследования.
27. Самый простой способ расчета «X» заключался бы в оценке записей с «предостерегающим знаком» на основе данных выборки. Но это невозможно сделать, поскольку эти записи определяются административной ситуацией и эта информация, безусловно, не собирается в ходе переписи. Поэтому используется другой подход.
28. Стратегия состоит в определении кластеров, исходя из социально-демографических показателей (например, граждане стран ЕС старше 65 лет, проживающие в средиземноморском регионе). В таких кластерах будут записи с «предостережениями» и записи со счетным фактором, равным 1. Каждая запись с «предостережением» должна иметься только в одном кластере. Количество населения, включенного в кластер, не должно быть большим (около 5% всего населения). В идеале должно быть столько кластеров, сколько необходимо, но в каждом из них должно быть небольшое количество населения.
29. Далее можно оценить размер кластеров. Счетный фактор для каждой записи с «предостережением», включенной в кластер, будет равен оцениваемому размеру кластера за минусом записей, включенных в кластер со счетным фактором 1, деленному на количество первоначальных «предостерегающих» записей. Эти значения могут быть больше 1, если кластер недостаточно представлен в регистре PADRON, и меньше 1, если он представлен избыточно.
30. Агрегирование счетных факторов позволит получить оценку численности населения для географической области, и результаты выборочного обследования будут откалиброваны по отношению к этой взвешенной основе.

V. Заключение

31. Такая методология обеспечивает возможности для использования большей части информации, содержащейся в регистрах, и распространения детальной информации, основанной на относительно большой выборке.
32. Кроме того, она позволяет сократить затраты и количество задействованного персонала и повышает качество проводимой полевой работы. Она также обеспечивает подход к трактовке различий между административными понятиями и статистическими переменными.
33. В связи с тем, что в ходе выборочного обследования собирают информацию только по демографическим переменным, географическое распределение сокращается по сравнению с предыдущими переписями.
34. И наконец, эта методология является абсолютно новой для Испании, поэтому приходится иметь дело с некоторыми неожиданными трудностями.

³ Это - упрощенное приближение различных правил по присвоению «предостережения» записи в регистре:

- - Если запись относится к гражданину Испании старше 13 лет, не имеющему соответствующего идентификационного номера

- - Если запись имеет идентификационный номер, но отсутствует в базах данных полиции

- - Если запись, которая относится к иностранцу, не меняется в течение длительного времени в PADRON.