



**ЭКОНОМИЧЕСКИЙ  
И СОЦИАЛЬНЫЙ СОВЕТ**

Distr.  
GENERAL

ECE/CES/AC.6/2008/8  
4 March 2008

RUSSIAN  
Original: FRENCH

---

**ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ**

**КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ**

Совместное совещание ЕЭК/Евростата  
по переписям населения и жилищного фонда

Одиннадцатое совещание  
Женева, 13-15 мая 2008 года  
Пункт 4 предварительной повестки дня

**РЕДАКТИРОВАНИЕ И ПРОВЕРКА ДАННЫХ**

**Проверка данных переписи во Франции**

Записка Национального института статистики и экономических исследований Франции

**Резюме**

В настоящем документе описываются процедуры, разработанные Национальным институтом статистики и экономических исследований для проверки данных переписи во Франции. Сначала описывается метод расчета численности населения в коммунах. Затем приводится описание операций проверки в отношении всех источников данных, которые используются для расчета численности населения в коммунах.

## **I. ОПИСАНИЕ МЕТОДА РАСЧЕТА ЧИСЛЕННОСТИ НАСЕЛЕНИЯ И СТАТИСТИЧЕСКИХ РЕЗУЛЬТАТОВ**

1. Национальный институт статистики и экономических исследований (НИСЭИ) должен обеспечивать соблюдение двух условий при публикации данных о численности населения:

а) ежегодная публикация данных о численности населения всех коммун;

б) исходя из принципа равного подхода ко всем коммунам, а также по соображениям, связанным с качеством данных о численности населения на территориях ниже коммунального уровня (непротиворечивость суммируемых цифр), данные о численности населения должны относиться к тому же году, что и все другие цифры. Невозможно использовать в отношении одной коммуны цифру за 2004 год, а в отношении другой - цифру за 2008 год.

2. НИСЭИ производит расчет этих показателей численности населения следующим образом:

### **A. Численность домашних хозяйств**

3. В случае коммун с числом жителей 10 000 человек и более цифра опирается на среднюю скользящую, определяемую по выборкам за пять последовательных лет. Путем агрегирования пяти выборок, относящихся к годам A-4 - A, производится расчет среднего числа жильцов в жилище, репрезентативного для середины периода (год A-2). Затем это число умножается на число жилищ, зарегистрированных в Регистре локализованных зданий (РЛЗ) в начале года A-2, что дает в результате общую численность населения в домохозяйствах коммуны.

4. Параллельно в отношении коммун с числом жителей менее 10 000 необходимо использовать цифру численности населения в медианный год пятилетнего цикла для обеспечения непротиворечивости с коммунами с числом жителей 10 000 и более.

5. Переписные обследования проводятся по следующей схеме.

A-4	A-3	A-2	A-1	A
Перепись		→		
	Перепись	→		
		Перепись		
			← Перепись	
				← Перепись

6. Что касается населенных пунктов, перепись которых проводится в год A-2, то используется результат переписного обследования.

7. Что касается коммун, переписанных в год A-1 и A, то численность населения в году A-2 получают путем интерполяции между результатами переписного обследования и последними опубликованными результатами. В случае тех коммун, перепись которых проводится в годы A-4 и A-3, то производится экстраполяция между результатами переписного обследования и данными за год A-2; такая экстраполяция опирается на данные о жилищном налоге (местный налог на жилище), позволяющие отслеживать изменения в числе жилищ в коммуне. Ее результаты уточняются с учетом разницы между темпами роста числа жилищ и темпами роста числа лиц. Сопоставляя эту разницу, измеренную в промежутках между последними переписями, с изменениями данных по жилищному налогу, можно определить, насколько изменилась численность населения.

## **В. Население, проживающее вне домохозяйств**

8. Население институциональных общин рассчитывается по состоянию на 1 января 2006 года либо на ту же дату, что и численность домохозяйств. Если институциональные общины коммун были переписаны в 2006 году, то использовалась цифра, рассчитанная на основе сбора данных. Если они подвергались переписи в 2004 или 2005 году, то численность населения институциональных общин обновлялась путем прибавления населения новых общин и вычета населения исчезнувших общин. Эти изменения рассчитываются на основе регистра институциональных общин. Если они подвергались переписи в 2007 или 2008 году, то численность населения институциональных общин обновляется путем интерполяции между последней опубликованной цифрой и цифрой, полученной в рамках переписи общин.

9. Численность бездомных, проживающих в подвижных жилищах или же главным местом проживания которых является гостиница, не обновляется. На протяжении

четырёх лет после года сбора используется одна и та же цифра. Цифра, рассчитанная на основе нового сбора данных, заменяет предыдущую после ее подготовки.

## **II. ПРОВЕРКА: ЭТАП ЗА ЭТАПОМ**

10. Для расчета численности населения коммун используются, таким образом, следующие элементы:

- a) базы данных, составленные в рамках различных операций по сбору;
- b) совокупность жилищ, определенная на основе регистра локализованных зданий (РЛЗ);
- c) регистр институциональных общин;
- d) данные о налоге на жилище.

11. Объектом проверки является каждый из этих элементов. Она проводится по мере их разработки или их учета, поскольку погрешности намного легче выявляются на самых ранних этапах.

12. Проверка включает в себя также изучение двух аспектов, способных влиять на качество или толкование данных: изменения, вызванные различиями в концепциях одной переписи по сравнению с другой, и влияние корректировок. В крупных коммунах вследствие этапа взвешивания и экстраполяции результатов опроса проверка будет также касаться влияния "жилищ с большим числом жильцов", которые могут создавать "эффект грозди".

13. Проверка преследует две цели: с одной стороны, удостовериться в статистическом качестве подлежащих публикации данных, а с другой стороны, понять изменения, оценить их надежность и определить ограничения их использования. Этот аспект является особенно важным, если различия в концепциях оказывают значительное влияние на какую-то определенную коммуну.

14. Нижеследующая диаграмма иллюстрирует последовательность различных этапов переписи.

Диаграмма. Этапы переписи



#### А. Совокупность жилищ, определенная на основе РЛЗ

15. Основой для выборочных обследований в каждой крупной коммуне служит его регистр локализованных зданий (РЛЗ). Этот регистр представляет собой перечень зданий (жилых, административных, промышленных или торговых), идентифицированных и локализованных по их адресу благодаря географической информационной системе.

16. РЛЗ был создан на основе данных всеобщей переписи 1999 года и обновляется с тех пор с использованием административных (разрешения на строительство, файлы местных налоговых органов) или почтовых (файлы адресов почтовой службы) массивов данных. Он ежегодно проверяется коммунами и в конечном счете утверждается НИСЭИ.

17. Следовательно, существует важный этап проверки данного РЛЗ, реализуемой непосредственно перед его использованием в качестве основы выборки.

18. Региональные управления готовят затем "доклад о редактировании" по каждой коммуне, который отражает результаты анализа качества регистра. Он содержит оценки в отношении участия коммуны в работе по анализу, используемых источниках и т.д.

19. После завершения проверки РЛЗ центр передает дополнительные элементы информации региональным управлениям (ряды данных о числе жилищ, информация о крупных адресах, расхождения между результатами сбора данных и данными РЛЗ) в течение года, который следует после предоставления данных, для того чтобы позволить им лучше оценить статистическое влияние изменений в РЛЗ.

20. На национальном, и даже региональном, уровнях измерение качества регистра локализованных зданий (РЛЗ) осуществляется в рамках ежегодного обследования, проводимого на местах счетчиками НИСЭИ.

21. Его цель заключается в измерении полноты РЛЗ с точки зрения отсутствующих или избыточных адресов в категории домов и связанных с ними жилищ. Она призвана обеспечить, чтобы все зарегистрированные на местах адреса были отражены в РЛЗ и, наоборот, чтобы все адреса в регистре существовали на практике.

**Таблица 1. Недоучет и завышение числа адресов домов и строений в РЛЗ**

		РЛЗ 2003	РЛЗ 2004	РЛЗ 2005	РЛЗ 2006	РЛЗ 2007
<b>Адреса</b>	Недоучет	2,3%	1,9%	1,3%	1,3%	1,4%
	Завышение	2,0%	1,9%	1,8%	1,7%	1,4%
	Баланс недоучета/ завышения	-0,3%	0%	0,5%	0,4%	0%
<b>Жилища</b>	Недоучет	2,3%	1,9%	1,1%	0,9%	1%
	Завышение	2,1%	2,3%	1,8%	1,4%	1%
	Баланс недоучета/ завышения	-0,2%	0,4%	0,7%	0,5%	0%

Источник: НИСЭИ, обследование качества РЛЗ в 2003-2007 годах.

## **В. Файлы, опирающиеся на сбор данных**

22. Качество этих файлов зависит от двух факторов: качества сбора и качества ввода. Кроме того, вносимые корректировки направлены на повышение качества путем досчета или устранения противоречивых ответов. Этап кодирования также является объектом контроля качества.

23. Сбор данных зачастую является самым деликатным этапом статистического процесса. Его качество зависит от точного и четкого соблюдения процедур, а также от сотрудничества респондентов (которое обуславливает объективность их ответов) и мотивации тех, кто собирает информацию, т.е. счетчиков. Тщательная подготовка (отбор, подготовка) и проверки позволяют добиться высокого качества. В ходе сбора данных проверки проводят коммуны и НИСЭИ. После сбора эти функции полностью переходят к НИСЭИ, что позволяет проверить собранные данные и при необходимости исправить их.

24. При невозможности получения данных по главным жилищам (невозможность установления контакта, длительное отсутствие или отказ отвечать) счетчик заполняет формуляр необследованного жилья (ФНОЖ) с указанием предполагаемого числа лиц, проживающих в этом жилище, что позволит в будущем скорректировать численность населения. Это позволяет уменьшить риск пропусков.

### После сбора: проверки, осуществляемые НИСЭИ

25. После получения документов из коммун НИСЭИ проводит их проверку сначала в институте, а затем на местах. Протоколы проверок адаптированы к новой конфигурации переписных обследований. По сравнению со всеобщими переписями новая конфигурация позволяет проводить более тщательную проверку по каждой коммуне благодаря уменьшению объема ежегодно собираемых данных и числа обследуемых коммун. Материалы налоговых органов играют важную роль в проведении проверок в институте и позволяют проверять полноту сбора данных по коммунам или определенным участкам сбора. Эти проверки - в институте и на местах - проводятся как можно ближе к дате сбора данных. Выявленные в ходе их погрешности подвергаются корректировке.

26. Эти проверки носят выборочный характер, поскольку они являются результатом последовательного "отсева":

а) первая проверка, проводимая в момент получения и регистрации вопросников, позволяет составить суждение о качестве сбора данных по каждой из коммун; она предусматривает также подсчет полученных документов путем считывания штриховых

кодов, присутствующих на каждом из документов. Результаты подсчета сопоставляются, с одной стороны, с результатами подсчета, осуществленного коммунами в момент завершения сбора, а, с другой стороны, с ожидаемым числом документов (с учетом итогов предыдущей переписи и наблюдаемых тенденций). Результаты подсчета имеются в наличии на внутрикоммунальном уровне - уровне переписного участка счетчика, что позволяет в случае необходимости проводить целенаправленные проверки на непротиворечивость и качество;

b) исходя из результатов этой проверки, примерно по 10% вопросников проводится углубленная проверка в институте, охватывающая всю коммуну или ее часть;

c) и наконец, для решения проблем, которые не удалось урегулировать в институте, НИСЭИ проводит проверки на местах (в 2007 году на местах было проведено примерно 28 000 проверок). Одно из 200 обследованных жилищ становится объектом проверки на местах. Данные проверки подтверждают результаты сбора данных в большинстве случаев. Они позволяют также выявлять погрешности. Эти погрешности корректируются, и содержащие ошибки документы заменяются документами, не содержащими ошибок, которые могут быть также интегрированы в цепочку ввода.

27. Особое внимание уделяется численности и качеству ФНОЖ. Необходимо снизить удельный вес ФНОЖ до нескольких процентов для того, чтобы не возникало сомнений в надежности оценок. В настоящее время их удельный вес составляет около 3% благодаря значительным усилиям, предпринимаемым в ходе сбора данных у домашних хозяйств.

28. Следует также удостовериться, что ФНОЖ составлен надлежащим образом в отношении главного жилья и что указанное число лиц является надлежащей оценкой размера домохозяйств. Проконтролировать эту информацию позволяют проверки на местах.

#### Проверка ввода данных

29. Ввод данных осуществляется внешним субподрядчиком. Производится оптическое считывание вопросников, которые сканируются и обрабатываются при помощи программы распознавания образов. Процесс завершается вводом текстовой информации. Качество, требуемое от субподрядчика, определено точными спецификациями. Они касаются защиты данных при их передаче и хранении субподрядчиком (физическая и информационная защищенность, обязательство каждого лица не разглашать обрабатываемую информацию) и качества данных (число документов, вводимых по каждой зоне сбора, допустимый коэффициент ошибок по каждой переменной).



В заключение вопросники пересчитываются, и в НИСЭИ во время сканирования штриховых кодов проверяются отклонения от результатов первого подсчета (см. выше).

30. В ходе этого процесса документы (индивидуальные формуляры и формуляры по жилищам), следовательно, подсчитываются три раза: первый раз - в коммуне во время проверки качества, второй раз - в НИСЭИ во время этапа контроля и третий раз - субподрядчиком. Сопоставление результатов этих трех подсчетов позволяет гарантировать очень высокий уровень качества в отношении количества формуляров.

31. И наконец, качество ввода обеспечивается повторным вводом (см. вставку).

Повторный ввод для проверки качества введенных переменных.

Ошибки ввода оцениваются с помощью повторного ввода данных выборки формуляров, который производится вторым субподрядчиком, который использует отсканированные изображения, применяя те же самые правила ввода. Расхождения анализируются в НИСЭИ с целью определения процента ошибок при кодировании каждой переменной. Повторный ввод данных осуществляется на протяжении всего процесса ввода данных, что позволяет в конечном итоге скорректировать протоколы ввода в отношении следующих пакетов.

С субподрядчиком регулярно ведется обмен информацией и организуются совещания с целью профилактики проблем или скорейшего их решения. Исходя из результатов повторного ввода, субподрядчику рекомендуется внести корректировки или усовершенствования в ходе самой кампании или же в период между кампаниями сбора данных, если речь идет о более масштабных усовершенствованиях. Так, например, оптимизация инструментов распознавания образов позволила существенно снизить коэффициент ошибок при вводе дат рождения. В этой связи следует отметить, что даже при вводе таких простых данных, как дата рождения, невозможно избежать ошибок: поскольку исходная информация является рукописной, ошибки толкования неизбежны.

32. Корректировка. Цель заключается в повышении качества данных, полученных в ходе сбора. Речь идет об анализе случаев полного или частичного непредоставления данных с целью их условного расчета, а также вопросников с противоречивыми данными (например, 95-летний человек считает себя трудоспособным). Для этого часто используется процедура подбора, которая заключается в восполнении недостающей информации при помощи ответов "донора", похожего на респондента. Данный подбор с каждым годом становится все более точным благодаря предыдущим результатам, позволяющим эффективно повышать его качество.

33. В частности, корректировка ФНОЖ позволяет использовать для расчета численности населения в коммунах вмененные данные о числе жильцов основных жилищ, по которым счетчикам не удалось получить данные (по причине длительного отсутствия, невозможности связаться или отказа предоставлять ответы). Поскольку в случае основных жилищ число жителей обычно известно (в более чем 8 случаях из 10), вменяется такое количество человек, которое насчитывает домохозяйство согласно формуляру необследованного жилья (ФНОЖ). В случае других ФНОЖ также производится условный расчет по методике, предусматривающей использование исходя из размера домохозяйства некой усредненной структуры, наблюдаемой в национальном масштабе в случае повторно опрошенных ФНОЖ. С учетом того факта, что небольшое число ФНОЖ заполняются ошибочно в отношении неосновных жилищ, эти расчетные данные не должны превышать долю основного жилья в населенном пункте.

### **С. Регистр институциональных общин**

34. Как и РЛЖ, этот регистр был создан по результатам переписи 1999 года и обновляется с использованием административных источников (файлы санитарных и социальных служб, файлы интернатов, файлы пенитенциарных учреждений или учреждений судебной защиты молодежи и т.д.). Как и в случае РЛЖ, этот регистр передается коммуна для проверки перед началом переписного обследования. Коммуны получают перечни своих институциональных общин, и им предлагается указать общины, которые прекратили свое существование, общины, о которых НИСЭИ неизвестно, и общины, адрес которых изменился. НИСЭИ затем проверяет предложения о создании или исключении общин из регистра, сформулированные коммунами.

35. В начале сбора данных счетчику, которому поручен сбор данных по институциональным общинам, предлагается провести окончательную проверку перечня таких общин.

36. И наконец, НИСЭИ ежегодно сопоставляет РЛЖ и регистр институциональных общин. Речь идет, с одной стороны, о надлежащей географической локализации каждой из институциональных общин, включенных в регистр; в данном случае РЛЖ используется в качестве проверочного перечня адресов и географических характеристик (координаты X, Y, принадлежность к тому или иному кварталу, кантону, внутрикоммунальной зоне). С другой стороны, речь идет о выявлении случаев дублирования или пропуска между двумя регистрами (в частности, когда институциональная община преобразуется в комплекс обычных жилищ или наоборот).

#### **D. Файл налога на жилье**

37. НИСЭИ ежегодно получает от Главного налогового управления (ГНУ) файл с данными по всем зданиям, облагаемым налогом на жилье. Первый этап анализа состоит в выявлении зданий, которые не соответствуют жилым зданиям (в частности гаражи, паркинги).

38. Затем этот файл используется для решения следующих двух задач:

- а) проверка и контроль сбора налогов по соответствующим коммунам;
- б) создание рядов данных об общей численности жилищ, основных жилищ, вторичных жилищ. Эти ряды используются для расчета численности населения в коммунах с числом жителей менее 10 000.

39. С другой стороны, коммуны получают от ГНУ информацию о налогоплательщиках и стоимости объектов обложения. Тот факт, что любой пропуск приведет к уменьшению налоговых поступлений, позволяет надеяться, что коммуна сообщит о пропусках ГНУ. Кроме того, любое избыточное включение (неоправданное включение в перечень), приводящее к дополнительному обложению налогоплательщика, также приведет к тому, что о таких случаях будет сообщено ГНУ.

40. НИСЭИ систематически анализирует профиль составленных таким образом рядов данных о жилищах и в случае атипичной динамики (например, значительное увеличение после значительного снижения) обращается к ГНУ с целью выявления возможных погрешностей в управлении. Подтвержденные ГНУ ошибки исправляются.

#### **E. Экстраполяция в крупных коммунах**

41. Численность населения в крупных коммунах рассчитывается путем экстраполяции суммарной выборки за пять лет последовательных наблюдений. Проверка полученной цифры опирается на показания надежности, рассчитываемые путем анализа каждого из ежегодных обследований и каждой из "суммарных" цифр. Наблюдаемая динамика является эмпирическим показателем надежности общей оценки. Для проверки также используются и другие критерии: непротиворечивость оценки результатам предыдущей переписи. Экзогенные показатели динамики также служат показателями качества полученных цифр: коммуны, данные по которым сочтены наиболее атипичными (как центром, так и региональными управлениями), становятся объектом углубленного анализа, призванного выявить возможные проблемы с качеством РЛЖ или неучтенный

эффект "грозди". Коммуны, в которых наблюдается такой тип проблем, могут становиться объектом специальных оценок, призванных устранить данные недостатки.

42. Прилагаемый формуляр используется для проверки предварительных оценок, опирающихся на четыре обследования. Он дополняется внутрикоммунальным анализом, призванным выявить кварталы, динамика в которых в силу своих особенностей оказывает существенное влияние на окончательный результат.

#### **Ф. Различия в концепциях между двумя переписями**

43. Данные различия могут оказывать влияние на анализ динамики численности населения в коммуне. Следовательно, необходимо составить перечень коммун, априори подверженных значительному влиянию данных расхождений, с целью проверки данных по этим коммунам и статистического анализа, опирающегося на эти коммуны (или входящие в их состав зоны). Так, например, регистрация в рамках новой переписи совершеннолетних учеников интерната по коммуне расположения их интерната, а не по коммуне проживания их родителей, окажет значительное влияние на постоянное население коммуны, в которой расположен интернат. Об этом необходимо сообщать контролерам и пользователям данных.

#### **Г. Окончательная проверка**

44. Поскольку каждая из составляющих подвергается проверке по мере ее поступления, то невозможно говорить об окончательном этапе проверки как таковом. Речь скорее может идти об обобщении всех качественных и количественных элементов, позволяющих пояснить показатели численности населения. Это имеет чрезвычайно важное значение для поиска ответов на возможные вопросы, которые могут возникнуть у коммун, а также у других пользователей статистики населения.

### **III. ПРОВЕРКА ПОДРОБНЫХ ДАННЫХ**

45. Помимо показателей численности населения объектом проверки также являются все подробные данные. Несомненно, что проверка качества общей цифры гарантирует выявление большинства проблем с качеством (охватом, погрешностями выборки и т.д.) и их учет. Однако проводится еще две дополнительных проверки.

#### **А. Проверка кодирования вопросников**

46. Кодирование включает в себя, во-первых, автоматизированный этап кодирования, в частности на основе классификации видов деятельности и занятий, и, во-вторых, этап поиска операторами региональных управлений незакодированных данных. В настоящее время для этого процесса разрабатывается специальный механизм контроля качества. Его принцип заключается в проведении повторного выборочного кодирования вопросников, репрезентативного с точки зрения различных видов кодирования (автоматического и повторного), с последующим анализом расхождений.

47. Цель заключается в измерении качества автоматического кодирования и качества повторного ручного ввода (расчет сначала общего процента случаев надлежащего кодирования, а затем в долгосрочной перспективе процентной доли важных переменных, например, процента правильно закодированных данных о сотрудниках руководящего звена). Вторая цель механизма контроля качества кодирования заключается в измерении коэффициента не поддающихся кодированию индивидуальных формуляров (например, в силу нечеткого названия должности на государственной службе в отсутствие иных указаний), что позволяет поставить реалистичные задачи с точки зрения качества и избежать погони за недостижимым качеством.

48. Такой механизм измерения и контроля позволяет также обогатить базу знаний, используемую в процессе автоматического кодирования, повышая ее эффективность как инструмента подготовки кадров, осуществляющих повторное кодирование, и в конечном итоге обеспечить надлежащее управление качеством. Результат этого этапа проявляется уже в ходе следующей кампании.

#### **В. Контроль правдоподобия подробных данных**

49. Подробные данные подвергаются анализу на основе "плоской выборки" различных переменных (возрастная пирамида, семейное положение, структура семей, занятость, жилище и т.д.). Речь идет о том, чтобы, с одной стороны, сопоставить эти данные с данными из других источников, а, с другой стороны, оценить их надежность путем анализа их динамики от одного обследования к другому. Анализ уровней может позволить оценить качество процедур корректировки. В совокупности с оценками точности анализ надежности призван позволить определить уровень детализации распространяемой информации (например, идет ли речь о возрастных группах с пятилетним или с десятилетним шагом). Данная проверка также позволяет выявлять и количественно оценивать возможные последствия изменений в вопросниках. Это

является важной информацией для пользователей, которые желают анализировать изменения по сравнению с предыдущей переписью.

50. Проведенный в первые годы анализ позволил выявить определенные нежелательные эффекты процедуры корректировки (которые мы ограничили в ходе следующих обследований) или слишком неточные протоколы ввода (эта погрешность была устранена в 2005 году). Он также позволил изучить последствия пересмотра вопросника (в частности, вопросов о занятости и структуре семей).

51. В настоящее время контрольная группа проводит анализ динамики основных статистических переменных и предлагает региональным управлениям проанализировать конкретные выявленные проблемы.

-----