



**ЭКОНОМИЧЕСКИЙ
И СОЦИАЛЬНЫЙ СОВЕТ**

Distr.
GENERAL

ECE/CES/GE.41/2009/10
18 August 2009

RUSSIAN
Original: ENGLISH

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Группа экспертов по переписям населения и жилищного фонда

Двенадцатое совещание

Женева, 28-30 октября 2009 года

Пункт 4 предварительной повестки дня

КАЧЕСТВО ПЕРЕПИСИ И ПРЕДУПРЕЖДЕНИЕ ИДЕНТИФИКАЦИИ

**Оценка точности гиперкубов уровня 2 КТЕС, разработанных
с использованием стратегии выборки итальянской
переписи населения 2011 года**

Записка Национального института статистики Италии

I. ВВЕДЕНИЕ

1. Итальянский национальный институт статистики (ИСТАТ) занимается изучением возможности использования выборочных методов в рамках разработки стратегии длинного/короткого переписного листа для итальянской переписи населения 2011 года. Данное решение опирается на метод простого случайного отбора для определения выборочных совокупностей частных домохозяйств из регистров населения и калиброванные оценки.

2. Поскольку используемая стратегия выборки вводит погрешности выборки, в настоящее время проводятся тесты и исследования для оценки эффективности выборочных оценок и точности гиперкубов данных для распространения.
3. Основным ограничительным фактором для определения стратегии выборки является точность оценок на различных территориальных уровнях: чем большим является территориальный уровень, тем более высокой должна быть точность оценок по переменным длинного переписного листа и их перекрестной классификации с другими переменными, относящимися либо к одному и тому же классу, либо с географическими переменными (к которым не применяются выборочные методы).
4. В частности, применительно к заданной территориальной единице и конкретному гиперкубу можно определить одновременно процент ячеек, по которым может быть оценена абсолютная частотность с низким уровнем точности и процент лиц, классифицируемых по этим критическим ячейкам. Этот последний показатель описывает информацию, оцениваемую с низким уровнем точности; низкие величины процента лиц, классифицированных в критических ячейках, свидетельствуют о хорошем качестве данных, включенных в этот гиперкуб.
5. Оценки влияния погрешности выборки на качество распространяемых данных касаются некоторых гиперкубов Евростата с различными признаками, относящимися к уровню 2 КТЕС.
6. Планирование итальянской переписи населения 2011 года проводилось с учетом как критических проблем предыдущей переписи, так и возможности внедрения методологических нововведений в соответствии с международными рекомендациями¹. С целью повышения эффективности переписных мероприятий, снижения нагрузки на муниципалитеты и сведения к минимуму статистической нагрузки на граждан, участвующих в переписи, были рассмотрены многочисленные решения. Наиболее важные из них касаются использования регистров населения, рассылки по почте переписных листов, смешанной системы сбора данных, главным образом по почте и через Интернет.

¹ ЕЭК ООН (2006 год). *"Рекомендации по проведению переписей населения и жилищного фонда 2010 года. Европейская экономическая комиссия Организации Объединенных Наций и Статистическое управление Европейских сообществ"*. Конференция европейских статистиков, ECE/CES/STAT/NONE/2006/4.

7. Поскольку успех будет гарантирован только в случае высоких коэффициентов предоставления спонтанных ответов, с этой целью в рамках следующей переписи будет использоваться стратегия выборки, опирающаяся на применение коротких и длинных переписных листов. Данная стратегия предусматривает одновременно использование двух различных вариантов вопросников. Короткий вопросник содействует снижению нагрузки, в то время как выборка длинных вопросников сохраняет всеобъемлющий характер переписной информации. Таким образом выборка может рассматриваться в качестве логического развития нововведений, предусмотренных в ходе планирования.

8. Использование стратегии выборки в рамках следующего раунда переписи предполагает не только экономию средств, но и меньший объем данных для обработки, а также обеспечивает возможность для повышения общего качества данных. Можно будет разработать и осуществить больше проверок переписных листов и контрольных мероприятий с целью снижения не связанных с выборкой погрешностей. Еще одним преимуществом стратегии выборки является сокращение сроков проведения переписи, которые являются ограничительным фактором, поскольку данные переписи 2011 года должны быть представлены в Евростат к 1 апреля 2014 года.

9. Одним из важных недостатков является введение погрешности выборки, в связи с чем важное значение приобретает оценка эффективности выборочных оценок и точности гиперкубов данных для распространения (статистические таблицы, приводимые в перекрестной классификации по переменным переписи).

II. СТРАТЕГИЯ ВЫБОРКИ

10. Данный подход заключается в одновременном использовании коротких и длинных переписных листов: короткий переписной лист предназначен для сбора данных только по демографическим и жилищным переменным, в то время как целью длинного переписного листа является сбор данных по всему набору переменных переписи, включая образовательный уровень, занятие и поездки на работу и с работы. Таким образом, демографические данные будут собираться в отношении всего населения, в то время как информация, касающаяся других переменных, будет собираться по выборке домохозяйств (только *частных домохозяйств*).

11. Стратегия выборки будет охватывать муниципалитеты с числом жителей более 20 000; в муниципалитетах с числом жителей менее 5 000 планируется использовать

традиционный подход, заключающийся в применении длинного переписного листа в отношении всего населения. В муниципалитетах, численность жителей которых составляет от 5 000 до 20 000 человек, использование данной стратегии будет зависеть от эффективности соответствующих оценок.

12. С учетом необходимости принятия довольно простой стратегии были проведены тесты и исследования для оценки эффективности выборочных оценок, касающихся различных моделей выборки и методов оценки (косвенной или прямой).

13. Результаты эмпирических исследований говорят в пользу следующей стратегии: простая случайная выборка домохозяйств (ПСВД) из регистров населения и калиброванные оценки², которые используют окончательные веса, модифицированные надлежащим образом для обеспечения большей репрезентативности выборки. Еще одно предложение касается планирования, по мере возможности, такой выборки, которая обеспечила бы хорошую точность оценок по переписным участкам, которые представляют собой территории субмуниципального уровня с числом жителей примерно 15 000 человек.

14. Еще одним важным вопросом, связанным с определением коэффициента выборки, является то, что увеличение объема выборки ведет к повышению точности оценок. В настоящее время планируется использовать коэффициент выборки в 33% с целью сохранения всего богатства переписной информации в максимально возможной степени. В случае жестких бюджетных ограничений коэффициент выборки может быть снижен путем нахождения компромисса между требованием экономии финансовых средств и точностью, необходимой в отношении различных территориальных уровней.

15. Что касается выбора метода оценки, то калиброванные оценки обеспечивают согласованность между оценками и демографическими данными, собираемыми в отношении всего населения. Косвенные методы, опирающиеся на методы оценки по малым территориям, могут использоваться для обработки более точных оценок как в отношении наиболее подробных территориальных уровней, так и редких групп населения.

² Deville J.C., Särndal, C.E. (1992) *"Calibration Estimators in Survey Sampling"*. Journal of the American Statistical Association, vol. 87, pp. 367-382.

Предварительные результаты текущих экспериментов³ выглядят обнадеживающими, поскольку при абсолютной частоте менее 150 единиц достигается сокращение коэффициента вариации (КВ) от 40% до 80%.

III. ТОЧНОСТЬ ОЦЕНОК

16. На основе данных переписи населения 2001 года было проведено модельное исследование⁴. С целью учета значительных различий между итальянскими муниципалитетами использовался набор из 40 муниципалитетов с различной численностью населения и по различным итальянским территориям уровня 2 КТЕС. В частности, в охват этого исследования было включено чуть менее 10% домохозяйств и чуть менее 10% лиц, охваченных последней переписью Италии.

17. Около 90 величин количества клеток в многомерных таблицах переписи были оценены с использованием методов калибровки. Количества клеток были рассчитаны по каждому из 497 подвергнутых тестированию субмуниципальных участков с численностью населения от 5 000 до 15 000 человек. Свойство выборочных оценок анализировалось с использованием средних статистических КВ, рассчитанных по 1 000 выборочных репликаций в отношении каждого из протестированных коэффициентов выборки. Моделирование "Монте-Карло" выборочных пространств производилось методом простой случайной выборки домохозяйств с различными коэффициентами и методом территориальной выборки.

³ Borrelli F., Carbonetti G., De Felici L., Solari F. (2008) *"Metodologie di stima per piccole aree applicabili a variabili di censimento rilevabili tramite questionario long form"*. XXIX Italian Conference on Regional Sciences, Bari (Italy), September 2008.

⁴ Карбонетти Дж., Фортини М. (2008 год) *"Ожидаемая точность результатов выборочных обследований при проведении переписи населения и жилищного фонда в Италии"*. Совместное совещание ЕЭК ООН/Евростата по переписям населения и жилищного фонда, ООН, Женева, (Швейцария), май 2008 года. ECE/CES/AC.6/2008/4.

18. В таблице 1 приводятся результаты моделирования⁵, касающиеся ПСВД. В ней по каждому классу числа ячеек и по каждому коэффициенту и по каждому из протестированных коэффициентов выборки приводится средний и максимальный коэффициент вариации, рассчитанный по субмуниципальным территориям.

Таблица 1

Распределение средних и максимальных коэффициентов вариаций в процентах по классам количества ячеек в отношении трех протестированных коэффициентов выборки (ПСВД)

Классы числа ячеек	коэффициент выборки = 10%		коэффициент выборки = 20%		коэффициент выборки = 33%	
	Средний КВ в %	Максимальный КВ в %	Средний КВ в %	Максимальный КВ в %	Средний КВ в %	Максимальный КВ в %
<10	143,3	191,8	101,4	123,7	66,5	95,8
10 30	75,9	85,1	48,4	54,6	33,8	38,5
30 50	51,8	57,1	31,8	37,1	23,4	25,6
50 100	38,6	41,3	22,3	28,4	17,4	19,1
100 250	25,4	28,5	15,7	19,6	11,4	12,8
250 500	16,1	18,3	10,4	12,5	7,5	8,1
500 1,000	11,8	12,8	7,5	8,2	5,3	5,9
1,000 2,500	7,5	8,9	4,7	5,9	3,3	3,9
2,500 5,000	4,9	5,4	3,0	3,6	2,0	2,5
5,000 10,000	3,2	3,8	2,0	2,5	1,3	1,9

19. Так, например, в случае 10-процентного коэффициента выборки средний КВ составляет менее 10%, когда количество ячеек превышает 1 000; когда коэффициент выборки возрастает до 33%, порог снижается до 250. Ожидается, что меньшая частотность соответствует более высоким величинам КВ. Однако более высокие величины КВ в случае малых количеств соответствуют меньшим различиям в абсолютном выражении. Как и ожидалось, наиболее точные оценки получают при использовании

⁵ Carbonetti G., Fortini M., Solari F. (2008) *"Innovations on methods and survey process for the 2011 Italian population census"*. European Conference on Quality in Official Statistics, Roma (Italy), 2008.

наибольших коэффициентов выборки. В целом выигрыш эффективности, измеряемый в качестве относительной разницы КВ, составляет 33-38%, когда коэффициенты выборки составляют от 10 до 20% и около 53-58%, когда коэффициенты выборки возрастают с 10 до 33%.

IV. ВЛИЯНИЕ ПОГРЕШНОСТЕЙ ВЫБОРКИ НА ГИПЕРКУБЫ ДАННЫХ ДЛЯ РАСПРОСТРАНЕНИЯ

20. Одним из ограничительных факторов применительно к определению стратегии выборки является точность оценок в отношении различных территориальных уровней: чем выше территориальный уровень, тем большей должна быть точность оценок переменных *длинного переписного листа и их перекрестная классификация с другими переменными*, либо относящимися к одному и тому же классу, либо к демографическим переменным (не охватываемых выборкой). В данном разделе анализируется потенциальное влияние стратегии выборки на качество гиперкубов данных для распространения.

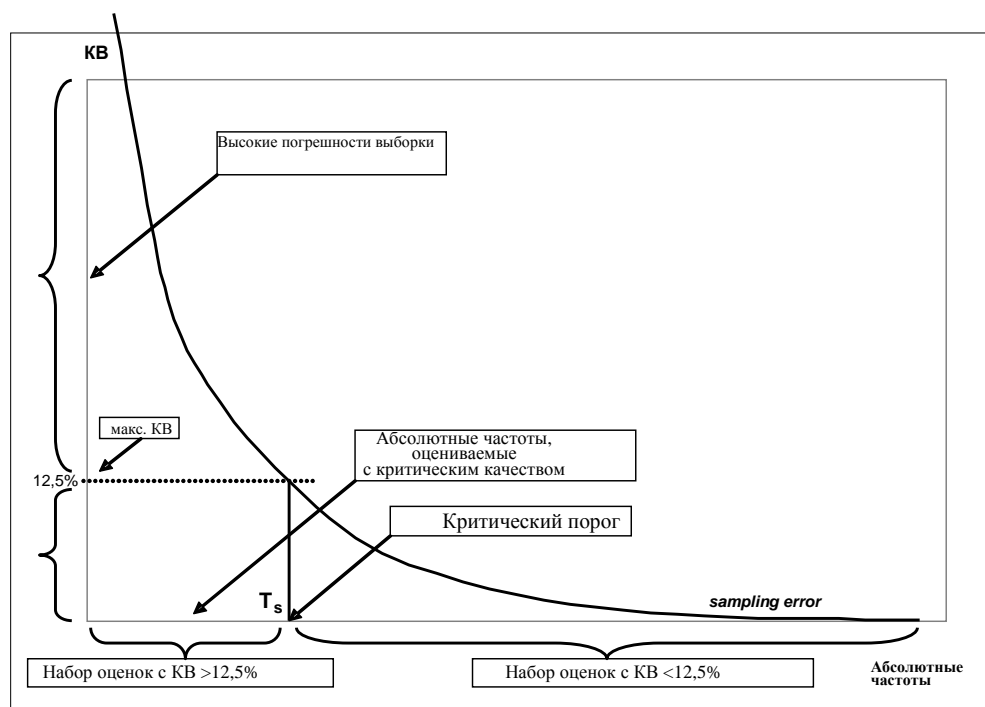
21. В соответствии с приемлемой величиной КВ кривая погрешностей выборки, построенная с помощью результатов моделирования⁶, позволяет установить абсолютную пороговую частоту, с тем чтобы каждая большая ее частота оценивалась с помощью КВ, большего фиксированной величины.

22. Рисунок 1 иллюстрирует, каким образом можно определять наборы оценок с критической точностью. В отношении фиксированного КВ, равного 12,5%, с помощью кривой погрешностей выборки, можно определить критический порог, ниже которого все абсолютные частоты оцениваются с большим, чем фиксированная величина КВ. С одной стороны, они могут рассматриваться в качестве критических случаев, обусловленных высокой погрешностью выборки. С другой стороны, более высокие по сравнению с пороговой частоты будут иметь меньшую погрешность выборки.

⁶ Carbonetti G., Dardanelli S., Fiorello E., Mastroluca S., Verrascina M., (2008) *“Ipotesi di innovazione per il censimento della popolazione del 2011: una valutazione degli effetti su un possibile piano di diffusione”*. XXIX Italian Conference on Regional Sciences, Bari (Italy), September 2008.

Рисунок 1

Выявление наборов оценок с "критической точностью" по кривой погрешностей выборки и критического порога T_s в отношении фиксированного CV



23. Если оценки касаются территорий, являющихся объектом частичного наблюдения, т.е. в которых не все население включается в выборку, тогда кривая погрешностей снижается, обеспечивая рост эффективности и снижение критического порога.

24. Таким образом, в отношении фиксированной территории и определенного гиперкуба данных для распространения может быть рассчитан одновременно процент ячеек, в которых расчетная абсолютная частота является меньшей наблюдаемого критического уровня, и процент лиц, классифицированных в этих критических ячейках. В частности, последний показатель количественно измеряет число данных, оцененных с низким уровнем точности; низкий процент лиц, классифицированных по критическим ячейкам, свидетельствует о хорошем качестве информации данного гиперкуба. Так, например, в качестве приемлемой величины показателя могут быть установлены 10%.

V. ПРОГНОЗИРУЕМОЕ КАЧЕСТВО ГИПЕРКУБОВ УРОВНЯ 2 КТЕС

25. Оценки⁷ влияния стратегии выборки на качество данных для распространения касаются некоторых гиперкубов с различной детализацией информации и относятся к уровню 2 КТЕС. В данном разделе после обсуждения выбора гиперкубов, включенных в программу распространения данных Евростата, приводятся некоторые результаты относительно прогнозируемого качества.

A. Рассматриваемые в исследовании гиперкубы

26. Из проекта программы распространения данных, которая будет одобрена на 2011 год, было отобрано 8 гиперкубов (Таблица 2), содержащих признаки итальянской переписи 2001 года, но также и, в частности, признаки в разбивке, близкой к нашим прошлым данным для распространения, с точки зрения количества и содержания информации, а также в разбивке, предусмотренной в отношении следующего раунда переписи с определением признаков, близких к тем, которые рекомендованы на международном уровне. Так, например, мы исключили гиперкубы, содержащие признак "проживание за границей", данные по которому Италия будет собирать впервые в рамках переписи 2011 года, а также признаки и разбивки, касающиеся статуса в домохозяйстве и семье, которые в некоторых случаях подвергаются в Италии иной разбивке по сравнению с разбивкой международного уровня для следующего цикла переписи.

27. Использование гиперкубов с признаками, касающимися домохозяйств, семей и населения в разбивке по статусу в домохозяйстве и семье, предполагает пересмотр определения и новый учет соответствующих разбивок, поскольку новые определения и разбивки в некоторых случаях отличаются от используемых в рамках итальянской программы распространения, что привело бы к увеличению длительности наших исследований. Мы отобрали гиперкубы, которые описывают только характеристики населения.

⁷ Carbonetti G. (2009) *"Use of sampling strategy in the Italian population census and accuracy of estimates for different territorial domains"*. ITACOSM09 - First Italian Conference on Survey Methodology, Siena (Italy), June 2009.

28. Гиперкубы уровня МАЕ2 не учитывались, поскольку они содержат "демографические" признаки, которые мы планируем включить в короткий переписной лист с целью получения информации по всему населению. Исходя из этого, мы сосредоточили внимание на признаках, требуемых только в отношении уровня 2 КТЕС, признаках, которые мы планируем включить в длинный переписной лист, а затем в гиперкубы общенационального уровня и уровней 1 и 2 КТЕС. Мы попытались отобрать гиперкубы, которые отличаются друг от друга и охватывают различные признаки, предлагаемые для наблюдения. Набор признаков является общим для всех с точки зрения пола и возраста, даже если в некоторых гиперкубах возраст разбивается на возрастные категории, а в других - подвергается более агрегированной классификации (М или S).

Таблица 2

Гиперкубы Евростата, использовавшиеся в исследовании влияния стратегии выборки на таблицы данных для распространения

Номер гиперкуба	Код	Название	База регистрации	Признаки (разбивка)
Первый гиперкуб	H.B1.E0.R1	Отдельные возрастные группы - "Текущий статус активности"	Все население	<ul style="list-style-type: none"> • Пол (М) • Возраст (L) • Текущий статус активности (М)
Второй гиперкуб	H.B1.E0.R2	Отдельные возрастные группы - "Занятие"	Все население	<ul style="list-style-type: none"> • Пол (М) • Возраст (L) • Занятие (М)
Третий гиперкуб	H.B1.E0.R3	Отдельные возрастные группы - "Отрасль"	Все население	<ul style="list-style-type: none"> • Пол (М) • Возраст (L) • Отрасль (отрасль экономической деятельности) (М)
Четвертый гиперкуб	H.B1.E0.R4	Отдельные возрастные группы - "Статус в занятости"	Занятые	<ul style="list-style-type: none"> • Пол (М) • Возраст (L) • Статус в занятости (М)
Пятый гиперкуб	H.B1.E0.R5	Отдельные возрастные группы - "Образовательный уровень"	Все население	<ul style="list-style-type: none"> • Пол (М) • Возраст (L) • Образовательный уровень (наивысший заверченный уровень образования) (М)

Четырнадцатый гиперкуб	H.B1.E1.R2	Занятые (лица в месте проживания) - "Занятие"	Все население в месте своего обычного проживания	<ul style="list-style-type: none"> • Пол (M) • Возраст (M) • Текущий статус активности (L) • Занятие (M) • Образовательный уровень (наивысший завершённый уровень образования) (M)
Пятнадцатый гиперкуб	H.B1.E1.R3	Занятые (лица в месте проживания) - "Отрасль"	Все население в месте своего обычного проживания	<ul style="list-style-type: none"> • Пол (M) • Возраст (M) • Текущий статус активности (M) • Отрасль (отрасль экономической деятельности) (M) • Образовательный уровень (наивысший завершённый уровень образования) (M)
Шестнадцатый гиперкуб	H.B1.E1.R4	Занятые (лица в месте проживания) - "Занятие" в разбивке по "Отрасли"	Все население в месте своего обычного проживания	<ul style="list-style-type: none"> • Пол (M) • Возраст (S) • Текущий статус активности (M) • Отрасль (отрасль экономической деятельности) (M) • Образовательный уровень (наивысший завершённый уровень образования) (M)

29. Число ячеек для каждого гиперкуба исходно рассчитывалось в качестве произведения числа категорий по каждому признаку, включенному в новое измерение А гиперкубов, с последующим введением нового измерения гиперкубов - "соответствующий размер", - рассчитываемого путем исключения из первого числа категорий, соответствующего итогам, промежуточным итогам и категориям "не указано".

30. В отношении отобранных гиперкубов мы рассчитали (таблица 3) "соответствующий размер" путем умножения числа категорий, используемых в программе распространения итальянской переписи 2001 года, и получили число потенциальных ячеек. Затем мы исключили "структурные нули", т.е. число ячеек, которые, несомненно, имеют нулевое значение в силу их соответствия невозможным результатам. В итоге мы получили число

приемлемых ячеек путем исключения несовпадаемых признаков (например, признак "возраст = 15 лет" и "образовательный уровень = категория 5 а)" МСКО).

Таблица 3

Число потенциальных ячеек и приемлемых ячеек по каждому анализируемому гиперкубу

Код	Число потенциальных ячеек	Число приемлемых ячеек
H.B1.E0.R1	1,212 (2x101x6)	1,062
H.B1.E0.R2	2,020 (2x101x10)	1,922
H.B1.E0.R3	3,434 (2x101x17)	3,126
H.B1.E0.R4	1,212 (2x101x6)	1,032
H.B1.E0.R5	1,414 (2x101x7)	1,342
H.B1.E1.R2	23,520 (2x21x8x10x7)	3,810
H.B1.E1.R3	29,988 (2x21x6x17x7)	5,574
H.B1.E1.R4	30,940 (2x13x10x17x7)	26,350

31. Как можно увидеть из пяти первых гиперкубов, перекрестная классификация демографических признаков только с одним социально-экономическим признаком отличается высокой степенью вероятности в ограниченном числе потенциальных ячеек. Это число снижается, если мы учитываем только приемлемые ячейки, т.е. по которым мы можем ожидать определенную частотность. Что касается второго блока гиперкубов, то даже в случае агрегированных категорий возраста (не отдельные годы, а пяти- или десятигодовые интервалы), число перекрестных признаков возрастает при одновременном увеличении соответствующего числа социально-экономических признаков при соответствующем потенциальном росте размера гиперкубов, даже если их размер существенно снижается в случае приемлемых ячеек. Гиперкуб наибольшего размера является последним; в нем осуществляется перекрестная классификация пола и возраста (десятигодовые категории), занятия, отрасли и образовательного уровня.

В. Некоторые результаты

32. Ниже приводятся некоторые результаты качественного анализа наиболее сложных гиперкубов из рассматривавшихся в исследовании и заполненных данными, полученными в рамках итальянской переписи населения 2001 года.

33. Первый пример касается гиперкуба Н.В1.Е1.Р.3, в котором осуществляется перекрестная классификация пола, возраста, текущего статуса активности, отрасли и образовательного уровня. Число *приемлемых ячеек* составляет 5 574 (без учета структурных нулей). Что касается этого гиперкуба, то в таблице 4 приводятся по каждому подвергнутому тестированию коэффициенту выборки (случайная выборка домохозяйств из регистров населения) и по трем итальянским территориям уровня 2 КТЕС различного размера (*Молизе* представляет малый размер, *Марке* - средний размер и *Сицилия* - крупный) критические пороги, проценты критических ячеек и проценты лиц в критических ячейках. Из этой таблицы легко увидеть прирост качества по мере увеличения выборочных коэффициентов и более высокое качество в случае крупных районов.

34. Так, например, в случае *Сицилии* при использовании 33-процентного коэффициента выборки и соответствующего критического порога в размере 100 59,4% клеток обладают абсолютной частотой ниже 100, но по этим ячейкам классифицирован только 1% соответствующего населения.

Таблица 4

Гиперкуб Н.В1.Е1.Р.3. Показатели качества (КВ = 12,5%), относящиеся к территориям Италии уровня 2 КТЕС: Молизе, Марке и Сицилии

Коэффициент выборки	Молизе			Марке			Сицилия		
	Критический порог	% критических ячеек	% лиц в критических ячейках	Критический порог	% критических ячеек	% лиц в критических ячейках	Критический порог	% критических ячеек	% лиц в критических ячейках
10%	100	79,2	10,7	250	78,8	6,9	500	75,9	4,2
20%	50	71,0	5,8	100	68,4	3,0	250	68,7	2,1
33%	30	63,6	3,4	50	59,8	1,5	100	59,4	1,0

35. В таблице 5 приводятся результаты, касающиеся более сложного гиперкуба (Н.В1.Е1.Р.4), в котором используются признаки пола, возраста, занятия, отрасли и образовательного уровня. Число *приемлемых ячеек* составляет 26 350.

Таблица 5

Гиперкуб Н.В1.Е1.Р4. Показатели качества (КВ = 12,5%), касающиеся трех территорий Италии, соответствующих уровню 2 КТЕС: Молизе, Марке и Сицилия

Коэффициент выборки	Молизе			Марке			Сицилия		
	Критический порог	% критических ячеек	% лиц в критических ячейках	Критический порог	% критических ячеек	% лиц в критических ячейках	Критический порог	% критических ячеек	% лиц в критических ячейках
10%	100	91,9	14,9	250	91,1	11,2	500	91,8	7,3
20%	50	86,5	9,3	100	84,4	6,0	250	87,6	4,5
33%	30	81,3	6,5	50	77,1	3,4	100	79,4	2,2

36. Выводы являются такими же, что и в предыдущем случае. Разница заключается в более низком качестве, поскольку данный гиперкуб обладает большим числом ячеек. Однако отмечается легкое снижение точности при том, что общее качество статистической таблицы остается приемлемым.

37. В таблице 6 обобщены результаты, касающиеся процентов лиц в индивидуальных ячейках в отношении всех 20 итальянских территорий, соответствующих уровню 2 КТЕС, в отношении всех гиперкубов, учитывавшихся в исследовании, а также в отношении трех тестируемых коэффициентов выборки.

Таблица 6

Распределение итальянских территорий уровня 2 КТЕС по проценту лиц, классифицированных в критических ячейках (КВ = 12,5%), в отношении некоторых гиперкубов Евростата

Число территорий уровня 2 КТЕС	s.r. = 33%		s.r. = 20%			s.r. = 10%				
	Процент лиц в критических ячейках									
Гиперкубы Евростата (приемлемые ячейки)	<5%	5-10%	<5%	5-10%	10-15%	<5%	5-10%	10-15%	15-20%	>20%
Н.В1.Е0.Р1 (1,062)	20	0	20	0	0	19	1	0	0	0
Н.В1.Е0.Р2 (1,922)	20	0	20	0	0	15	4	1	0	0
Н.В1.Е0.Р3 (3,126)	20	0	17	3	0	11	4	4	1	0
Н.В1.Е0.Р4 (1,032)	20	0	16	4	0	7	8	5	0	0
Н.В1.Е0.Р5 (1,342)	20	0	20	0	0	15	5	0	0	0
Н.В1.Е1.Р2 (3,810)	20	0	18	2	0	12	6	2	0	0
Н.В1.Е1.Р3 (5,574)	20	0	17	3	0	10	5	5	0	0
Н.В1.Е1.Р4 (26,350)	15	5	8	10	2	1	9	6	3	1

38. Можно отметить, что при использовании 33-процентного коэффициента выборки в семи из восьми рассматривавшихся гиперкубов процент индивидуальных лиц, классифицированных в критических ячейках, составляет менее 5% в случае всех территорий, в то время как в случае самого крупного гиперкуба H.B1.E1.R4 показатель менее 5% отмечается в отношении 15 территорий, а величина от 5 до 10% в остающихся пяти.

39. Выборка с коэффициентом в 20% предполагает некоторую утрату качества. По сути, процент лиц в критических ячейках в отношении различных территорий составляет менее 10%, за исключением последнего гиперкуба, в котором только в случае двух (наименьших) территорий этот показатель составляет от 10 до 15%.

40. Качество может быть приемлемым и при использовании стратегии выборки в 10%. В данном случае по большинству территорий могут наблюдаться показатели качества, составляющие менее 10% по различным гиперкубам; некоторые проблемы могут возникнуть в случае более сложных гиперкубов и наименьших территорий.

VI. ВЫВОДЫ

41. Полученные результаты свидетельствуют о целесообразности использования методов выборки в целях проведения итальянской переписи населения 2011 года. Стратегия простой случайной выборки домохозяйств из регистров населения и использование калиброванных оценок позволяют получать точные оценки и снизить не связанные с выборкой погрешности. Оценки по малым территориям могут повысить точность данных в отношении территорий малого размера или редких групп населения.

42. Определения и классификация итальянской переписи полностью соответствуют требованиям Евростата, что позволяет расчет всех искомых гиперкубов. Тем не менее использование стратегии выборки внесет некоторую вариацию в итоги переписи. Если говорить более конкретно, то все гиперкубы, требуемые на уровне 2 МАЕ, будут заполняться данными, получаемыми с помощью коротких переписных листов и, следовательно, будут свободны от выборочной вариации, в то время как гиперкубы общенационального уровня и уровней 1 и 2 КТЭС, которые содержат признаки, данные по которым будут собираться выборочным методом с использованием длинных переписных листов, будут подвержены выборочной вариации и, следовательно, давать оценки с погрешностью выборки.

43. Эксперименты по оценке влияния стратегии выборки на качество некоторых гиперкубов, предусмотренных в отношении уровня 2 КТЭС, подчеркнули эффективность нового подхода к проведению итальянской переписи населения. Главный вывод заключается в том, что гиперкубы с более чем 20 000 ячеек могут оцениваться с низким процентом единиц в критических ячейках даже при небольшом коэффициенте выборки. Некоторые проблемы наблюдаются только в случае малых территорий (например, Валь д'Аоста).

44. Поскольку существует зависимость между размером выборки и затратами на сбор данных, принятие окончательного решения будет определяться компромиссом между объемом финансирования и минимальным уровнем точности, требуемым в отношении различных территорий.

45. Еще одним смежным вопросом является возможность получения данных большей информативности на более низком территориальном уровне. Различные подходы свидетельствуют о целесообразности использования *разбивок с сокращенным числом* категорий или использования альтернативных косвенных оценок. На практике методы малых территорий, как представляется, содействуют повышению точности оценок в отношении малых территориальных единиц, в случае которых выборка может быть нерепрезентативной, и в отношении весьма малых чисел ячеек в случае крупных территорий.
