



Дата: август 2020 года (исправленная версия №2)

Ключевые слова: перепись населения и жилищного фонда, моделирование численности населения, недостаточный охват переписи

Значение моделирования численности населения для планирования и подготовки переписи

Перепись населения и жилищного фонда относится к наиболее сложным и масштабным мероприятиям, которые та или иная страна проводит в мирное время. В большинстве стран с низким или средним уровнем дохода перепись населения и жилищного фонда является основным источником данных о численности населения, его пространственном расселении, возрастной и половой структуре, жилищно-бытовых условиях и о других ключевых социально-экономических показателях. Перепись дает всеобъемлющее представление о жилье и его состоянии в стране. Данные переписи имеют принципиальное значение для качественного государственного управления, развития, снижения рисков и кризисного реагирования, реализации программ социального обеспечения и анализа рынков. При этом перепись остается дорогим и сложным проектом с логистической точки зрения и в идеальном случае проводится раз в десять лет.

Национальные статистические службы (НСС) часто сталкиваются с препятствиями, которые могут повлиять на организацию и проведение национальной переписи населения в целом, а также на качество данных. В настоящее время разрабатываются и анализируются инновационные статистические методы, которые могли бы послужить хорошим подспорьем и потенциально улучшить планирование и проведение переписей. Потенциалом с точки зрения содействия планированию и повышению качества переписей обладает методика моделирования численности населения.

Подходы к оценке численности населения, основанные на статистическом моделировании, разрабатываются с учетом корреляции между плотностью населения и уровнями геопространственных ковариантных данных. С учетом такой корреляционной связи можно построить статистическую модель для расчета плотности населения с определенным уровнем неопределенности в зависимости от изначальных вводных данных. Преимущество таких моделей — их относительная дешевизна. Кроме того, они могут стать альтернативным способом получения свежих оценок численности населения для планирования и проведения переписей. Тем не менее, в силу их вероятностного характера оценки численности населения, основанные на принципе моделирования, не следует воспринимать так же, как реальный подсчет численности населения. Их скорее следует расценивать как возможность совершенствования планирования в условиях, когда последняя перепись проводилась очень давно, и ее данные в целом устарели. Если говорить конкретно, оценки зависят от имеющихся или специально проведенных выборочных обследований домохозяйств, в



Методические рекомендации

частности, проведенных в рамках подсчета домохозяйств для этих обследований. Результаты таких выборочных обследований затем комбинируются с уровнями геопространственных ковариантных данных, и в результате получается основа статистической модели. Создав модель, можно использовать и другие независимые источники данных, например, данные обследования домохозяйств, для проверки вновь смоделированных оценок численности населения.

Численность населения моделируется как функция от уровней ковариантных данных, полученных со спутниковых снимков или благодаря геопространственной информации, например, расстояний до дорог, интенсивности освещения в ночное время или классификации почвенного покрова. Лучшими предикторами плотности населения часто оказываются наличие, плотность и принципы размещения построек и сооружений. Предположив, что население в основном сосредоточено в сложившихся, оседлых районах, мы можем рассчитать население страны, скомбинировав показатели плотности населения с картами этих сложившихся районов. Для этого требуется составление карт населенных пунктов, чаще всего с использованием новых спутниковых снимков очень большого разрешения. Далее с помощью указанной модели можно рассчитывать показатели плотности населения или числа людей на единицу площади по любым территориальным единицам, например, по административно-территориальному делению или по счетным участкам или даже по координатным сеткам 100 x 100 метров.

Что касается надежности таких данных, смоделированные оценки численности населения содержат в себе элемент неопределенности, из-за которого может возникнуть необходимость применения приближенных доверительных интервалов в отношении расчетной плотности населения. На сегодняшний день не существует механизма определения допустимого уровня неопределенности и его учета при принятии решений. В отношении общих показателей, в основе которых лежат данные обследований, рекомендуется учитывать, что тот или иной показатель не является надежным, если коэффициент вариации (стандартное отклонение от среднего) меньше 33%.

Тем не менее, несмотря на то, что оценки численности населения могут содержать неточности, их можно использовать:

- в качестве базовой информации для планирования переписей и переписной картографии;
- вместо подсчета в недоступных районах;
- в качестве базовой информации для актуализации основного выборочного фрейма;
- для оценки охвата переписью;
- для анонимизации результатов переписи;
- для совмещения переписей населения и жилищного фонда с сельскохозяйственными переписями.



Базовая информация для планирования переписей и переписной картографии

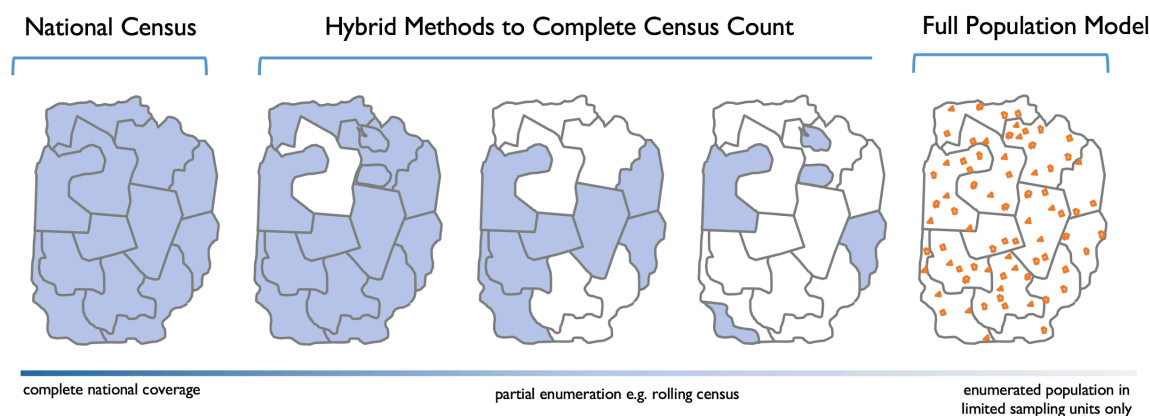
Процесс планирования переписи и картографических работ на местах фундаментальным образом зависит от проекций (прогнозов) численности населения по регионам, районным муниципальным образованиям, сельским населенным пунктам и счетным участкам (СУ). От этих расчетов численности населения зависит распределение ресурсов на проведение переписи, в том числе денежных средств, выделения времени на отдельные этапы проведения переписи, на персонал и оперативную деятельность. Тем не менее, официальные демографические прогнозы, составляемые сразу после очередной переписи, со временем могут устаревать, если их постоянно не актуализировать свежими данными обследований, полученными в период между переписями. Кроме того, хотя демографические прогнозы могут быть точными на национальном и региональном уровнях, этой точности может не быть на более низких уровнях, особенно на уровне СУ. Поскольку модели численности населения основаны на самых свежих геопространственных ковариантных данных, они часто дают более точную оценку на уровне района, сельского населенного пункта или СУ.

Модели численности населения в сочетании с геопространственными данными по дорожной/речной сети можно использовать для планирования максимально эффективных маршрутов для картографов. Более эффективные работы в поле могут снизить общие расходы и количество времени, необходимого для картографирования. В связи с огромными финансовыми затратами на проведение переписи принципиально важно учитывать потенциальные возможности для экономии. Снижать затраты можно, в частности, за счет уменьшения времени, необходимого картографам или счетчикам для работы на местах, за счет прокладки более эффективных маршрутов, более точного планирования числа картографов, необходимых на каждый СУ и т. д. Кроме того, картографирование населенных местностей с помощью спутниковых снимков улучшает планирование и актуализацию карт, так как позволяет выявить небольшие населенные пункты, ранее на картах не отмеченные. Моделирование численности населения и данные о речной и дорожной сетях можно также использовать для формирования новых счетных участков.

Замена подсчета в недоступных районах

При проведении переписи некоторые страны сталкиваются с трудностями с точки зрения безопасности или доступности некоторых своих районов, в которые невозможно отправить счетчиков. В этом случае можно моделировать численность и плотность населения недоступных районов с помощью переписных данных, собранных ранее, или специальным образом скомпонованных данных обследований (рис. 1). Точность оценок будет зависеть от качества подсчета в доступных районах, истинной эффективности модели и вероятности допущений, принимаемых при моделировании. Разработанная модель численности

населения всегда адаптируется к ситуации на местах (т. е. в зависимости от доступности данных или демографического контекста) для формирования оптимальных оценок численности с учетом имеющихся исходных данных и времени.



Набор общих оперативных баз данных по демографической статистике

Набор общих оперативных баз данных (ООБД) — это оперативные референсные массивы данных, необходимые для поддержки оперативной деятельности и принятия решений для всех участников гуманитарного реагирования. Набор общих оперативных баз данных по демографической статистике (COD-PS) — это актуальный массив демографических данных самого нижнего из имеющихся географических уровней, дезагрегированных по половозрастной структуре. Их используют в гуманитарных ситуациях только для удовлетворения потребности в данных в рамках кризисного реагирования и вмешательства. По сути их базовый принцип — более гибкие и прагматичные критерии «лучшего из имеющихся вариантов», а не официальные статистические стандарты. В этом случае можно оценить COD-PS и показатели плотности населения недоступных районов с использованием данных переписей, собранных частично для построения модели. По сути, если модели данных о численности населения являются лучшим из имеющихся источников актуальных данных, дезагрегированных по половозрастной структуре и географическому расположению, их можно использовать в качестве COD-PS для содействия оперативной деятельности и принятию решений в гуманитарных ситуациях.

Основные рамки выборки и оценки численности населения

После валидации данных переписи обычно формируется основная выборка СУ с использованием информации из базы данных переписи. Эта выборка используется в период между переписями для проведения крупных социально-экономических и демографических



Методические рекомендации

обследований в стране. Параметры основной выборки со временем обычно отклоняются от рабочей выборки из-за роста численности населения и изменений в его пространственном расселении. В зависимости от вида переписных картографических материалов, на базе которых была сформирована первоначальная основная выборка, для актуализации основного выборочного фрейма можно использовать оценки численности населения с координатной привязкой. Для этого есть два способа. В зависимости от вида переписных картографических материалов, на базе которых была сформирована первоначальная основная выборка, для актуализации основного выборочного фрейма можно использовать оценки численности населения с координатной привязкой. Для этого есть два способа. Если карты нецифровые, обновление численности населения в СУ оценками с координатной привязкой невозможно, поскольку границы СУ не оцифрованы, и на слои СУ невозможно наложить слои с численностью населения и координатной привязкой. В этой ситуации мы бы рекомендовали сформировать новую основную выборку на базе оценок численности населения с координатной привязкой, рассматривая каждую сетку или группу смежных групп сеток в качестве первичных единиц выборки. Эта вторая стратегия является усовершенствованным и хорошо известным методом районированной выборки, так как в этом случае мы имеем более точную оценку численности населения данного «района» (который в данном случае представляет собой сетку или группу смежных сеток). Как и в других случаях, это возможно, только если смоделированные оценки достаточно надежны, чтобы отражать изменения в демографических показателях.

Оценка охвата переписи

Обычно оценка охвата переписи проводится после окончания этапа подсчета с использованием различных методов и источников данных (демографических прогнозов, данных регистрации актов гражданского состояния, медицинских данных, образовательных данных и т. д.) и может включать также проведение обследования по итогам подсчета.

Дополнительный источник для сопоставлений — геопространственные оценки численности населения. Чтобы геопространственные оценки численности населения могли служить источником информации для оценки охвата переписи, расчет должен проводиться с использованием данных из счетных участков с идеальным охватом и хорошим качеством данных. Еще раз: геопространственные оценки следует рассматривать, только если эффективность модели оказалась высокой, а соответствующие допущения — реалистичными.

Анонимизация результатов переписи

Для анонимизации массива данных переписи можно использовать статистические методы (рис. 1, левая панель). Отдельные наблюдения можно агрегировать с помощью ГИС-программ в суммарные показатели высокого разрешения по населению с привязкой к координатной сетке, а также с помощью статистических методов. С помощью таких методов эти наблюдения (например, суммарные показатели по административным единицам) можно



Методические рекомендации

дезагрегировать до разрешения 100x100 м и, таким образом, создать уровень анонимных данных, которые могут использовать государственные органы для принятия решений и планирования обследований, а также общественность, если эти данные будут обнародованы.

Совмещение с сельскохозяйственными переписями

Совмещение переписи населения и жилищного фонда с сельскохозяйственной переписью в рамках национальной статистической системы соответствует Глобальной стратегии по совершенствованию сельскохозяйственной и сельской статистики¹. На оперативном уровне такого совмещения можно добиться за счет 1) создания интегрированной системы обследований, 2) разработки основного выборочного фрейма для сельского хозяйства и 3) интеграции систем управления данными. Опыт показывает, что сбор ограниченного объема строго определенных сельскохозяйственных данных во время проведения переписи населения и жилищного фонда может внести существенный вклад в построение эффективной основной фрейма для сельскохозяйственных переписей и обследований во многих развивающихся странах. Кроме того, модели численности населения дают возможность уточнять выборочный фрейм для сельскохозяйственных переписей, поскольку показатели плотности населения отражают площади занятой земли, что дает информацию о потенциальных сельскохозяйственных районах с низкой плотностью населения и потенциальных несельскохозяйственных районах с высокой плотностью населения. Эту информацию можно использовать для дальнейшего уточнения сельскохозяйственного выборочного фрейма за счет:

- совмещения смоделированных оценок с географической информацией сельскохозяйственных переписей;
- классификации урбанизированных или населенных районов, которые должны быть исключены при построении выборочного фрейма для сельскохозяйственных обследований, так как это поможет избежать задействования ресурсов на оперативную деятельность в районах, не используемых в сельском хозяйстве.

Заключение

Модели численности населения могут принести пользу на различных этапах переписи: от планирования до проведения, от оценки качества до применения данных. Тем не менее, важно отметить, что такие оценки следует использовать с осторожностью: разные модели дадут разные оценки численности населения. К счастью, существуют механизмы для объективного «выбора» самой статистически точной модели и ее валидации. Несмотря на возможность выбора объективно лучшей модели, необходимо иметь в виду, что

¹ Продовольственная и сельскохозяйственная организация ООН. Глобальная стратегия по совершенствованию сельскохозяйственной и сельской статистики. 2010
http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/ICAS5/Ag_Statistics_Strategy_Final.pdf



Методические рекомендации

геопространственная модель населения не может корректно отображать реальность. Как говорил Джордж Бокс, следует «помнить, что все модели неправильные; с практической точки зрения вопрос должен звучать так: до какой степени неправильными они должны быть, чтобы их нельзя было использовать».

В подготовке настоящих рекомендаций (версия 2) участвовали сотрудники Фонда ООН по народонаселению Сабрина Джуран, Матиас Кюпи, Морин Джоунс, Каталина Валле, Катрин Вени, Матиас Натан, а также участники проекта WorldPop при Саутгемптонском университете Энди Тейтем, Аттила Лазар, Дуглас Лейжер и Хизер Чемберлен.