

Связанные микроданные переписи населения 2002 г.: двадцатипроцентная выборка по частным домохозяйствам и индивидам

Лицензия: для обезличенных датасетов без возможности скачивания

Доступ: Продвинутый

Дата обновления: 17.02.2022

Теги:

Росстат, панельные данные, перепись населения

Описание

Двадцатипроцентная выборка с микроданными Всероссийской переписи населения 2002 года по частным домохозяйствам и индивидам, проживающим в этих домохозяйствах, которая частично дополнена связанными с ней микроданными Всероссийской переписи населения 2010 года.

При формировании выборки по 2002 году использовался метод систематического сэмпинга (Two-Stage Systematic Sampling), а также исключались наблюдения, которые соответствуют временно пребывающим на территории России, лицам из коллективных домохозяйств и лицам, проживающим в населенных пунктах, которые входят в состав ЗАТО. При формировании выборки не учитывались домохозяйства, в которых проживает одиннадцать и более лиц.

Единица наблюдения в датасете — член частного домохозяйства. Выборка разделена на восемь частей по федеральным округам. Минимальный уровень географической детализации — муниципальное образование второй ступени (муниципальные районы, городские округа и внутригородские районы). Также отражены крупные города.

Выборка частично была дополнена микроданными Всероссийской переписи населения 2010 года по частным домохозяйствам и всем индивидам, проживающим в этих домохозяйствах. В том случае, если была найдена связь только между некоторыми индивидами в домохозяйстве в 2002 и 2010 годах, сохранялась информация о всех членах домохозяйства в обе волны переписи.

В выборке представлены основные показатели из переписных листов: характеристики индивидов, характеристики домохозяйства и информация о характеристиках жилища.

Выборка содержит 28 020 429 наблюдений в 2002 году, 84 атрибута в 2002 году и 100 атрибутов в 2010 году.

Датасет доступен в продвинутом режиме доступа в формате CSV (кодировка: «UTF-8», разделитель: «;»).

Характеристики датасета

Использование

Лицензия: [для обезличенных датасетов без возможности скачивания](#)

Доступ: Продвинутый

Предоставление данных

Формат:

CSV

Характеристики набора

Временной период: 2002 г.

Количество наблюдений: 28020429

Обновления

Размещено: 17.02.2022

Обновлено: 17.02.2022

Частота обновления: не указано

Дата следующего обновления: не указано

Происхождение

Источники данных: Федеральная служба государственной статистики

Ответственные за набор данных: Копыток Витовт Константинович, Кузьмина Юлия

Викторовна, Тихонов Сергей Владимирович

Структура набора данных

Информация об атрибутах, их содержании, кодах, с помощью которых закодированы значения атрибутов, а также расшифровки кодов представлены в [интерактивном описании \(Codebook\) этого набора](#). Также кодбук в формате html добавлен в архив с датасетом ([codebook_2002_linked.html](#)).

Внесенные в набор данных изменения по сравнению с источниками

Всего использовалось последовательно три способа связывания:

- детерминистическое связывание домохозяйств в пределах региона;
- детерминистическое связывание индивидов в пределах населенного пункта;
- вероятностное связывание индивидов в пределах региона.

Для формирования выборки используется систематический сэмплинг (Two-Stage Systematic Sampling).

- Сначала с помощью систематического сэмплинга отбираются частные домохозяйства, соответствующие выбранному критерию по числу лиц, проживающих в этих домохозяйствах (для двадцатипроцентной выборки – не более десяти лиц в одном домохозяйстве);
- На втором этапе отбираются все индивиды, проживающие в этих домохозяйствах.

Отбор домохозяйств происходит по следующему алгоритму.

1. Исключаются наблюдения, которые соответствуют временно пребывающим на территории России и лицам из коллективных домохозяйств (наблюдения, соответствующие домохозяйствам бездомных, сохраняются), а также лицам, проживающим в населенных пунктах, которые входят в состав закрытых административно-территориальных образований (ЗАТО).
2. К микроданным, исходя из кода ТЕРСОН-МО населенного пункта, в котором проживает домохозяйство, добавляются географические атрибуты: код федерального округа, код региона, код муниципального района/городского округа/внутригородского района.
3. Уникальная выборка домохозяйств последовательно упорядочивается по следующим атрибутам:
 1. федеральный округ;
 2. регион;
 3. муниципальный район, городской округ, внутригородской район;
 4. тип населенного пункта (город/село);
 5. количество лиц, проживающих в домохозяйстве;
 6. идентификатор домохозяйства.
4. Выбирается случайное число от 1 до 5. С шагом 5 отбираются номера домохозяйств, включаемых в выборку.

На втором этапе в выборку включаются все индивиды, которые проживают в отобранных домохозяйствах.