

Выборка 10% микроданных переписи населения 2002 г

Версия для печати больше не поддерживается и может содержать ошибки обработки. Обновите закладки браузера и используйте вместо этого функцию печати браузера по умолчанию.



Содержание

- 1 История изменений
- 2 Общая информация о датасете
- 3 Общее описание набора данных
- 4 Структура набора данных
- 5 Известные ограничения данных
 - 5.1 Источники данных
 - 5.2 Описание проведенных проверок и их результатов
- 6 Внесенные в набор данных изменения по сравнению с источниками
- 7 Источники данных
- 8 Ответственные за набор данных

История изменений

Дата	Версия	Автор	Описание изменений
16.12.2021	1.0	Копыток Витовт Константинович Кузьмина Юлия Викторовна Тихонов Сергей Владимирович	Документ создан

Общая информация о датасете

Десятипроцентная выборка с микроданными Всероссийской переписи населения 2002 года по частным домохозяйствам и индивидам, проживающим в этих домохозяйствах.

При формировании выборки 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 **восемь и более лиц**.

В каталоге в продвинутом режиме доступа размещена расширенная двадцатипроцентная выборка.

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).

В выборке представлены основные показатели из переписных листов: характеристики индивидов (возраст, пол, уровень образования, владение языками, экономическая активность, источники дохода и др.), характеристики домохозяйства (размер, родственные связи) и информация об основных характеристиках жилища.

Всего набор содержит 13 647 647 наблюдений по 57 атрибутам.

Датасет доступен для работы в формате CSV (кодировка: «UTF-8», разделитель: «;»).

Общее описание набора данных

Атрибут	Значение
Общие сведения	
Полное наименование набора данных	Микроданные переписи населения 2002 г.: десятипроцентная выборка по частным домохозяйствам и индивидам
Сокращенное наименование набора данных	Выборка 10% микроданных переписи населения 2002 г.
Краткое описание набора данных	<p>Десятипроцентная выборка с микроданными Всероссийской переписи населения 2002 года по частным домохозяйствам и индивидам, проживающим в этих домохозяйствах. При формировании выборки 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 восемь и более лиц. В каталоге в продвинутом режиме доступа размещена расширенная двадцатипроцентная выборка.</p>
Краткое описание источника набора данных	<p>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).</p> <p>В выборке представлены основные показатели из переписных листов: характеристики индивидов (возраст, пол, уровень образования, владение языками, экономическая активность, источники дохода и др.), характеристики домохозяйства (размер, родственные связи) и информация об основных характеристиках жилища.</p> <p>Всего набор содержит 13 647 647 наблюдений по 57 атрибутам.</p> <p>Датасет доступен для работы в формате CSV (кодировка: «UTF-8», разделитель: «;»).</p>
Краткое описание источника набора данных	Выборка построена на микроданных Всероссийской переписи населения 2002 года, полученных Росстатом после обработки индивидуальных переписных листов
Покрываемый временной период	2002 год
Доступные форматы	CSV (кодировка: «UTF-8», разделитель: «;»)
Размер набора данных	2 ГБ
Тематика набора данных	Статистика
Режим доступа	Открытый (у Вас есть возможность работать с данными на своем личном ПК, в том числе скачивать данные после верификации и подтверждения номера телефона)
Периодичность публикации и обновления	
Дата размещения в каталоге	16.12.2021
Периодичность обновления в каталоге	Не обновляется
Дата последнего обновления набора данных в каталоге	Отсутствует
Характер последнего обновления	Не обновляется
Дата следующего обновления набора данных	Не обновляется
Дополнительные сведения	
Цитирование набора данных на русском языке	<p>Микроданные переписи населения 2002 г.: десятипроцентная выборка по частным домохозяйствам и индивидам // Росстат; обработка: Копыток В.К., Кузьмина Ю.В., Тихонов С.В. Инфраструктура научно-исследовательских данных, АНО «ЦПУР», 2021. Доступ: открытая лицензия в исследовательских целях. Размещено: 16.12.2021. URL: https://data.rcsi.science/data-catalog/datasets/189/</p>
For references (English)	<p>2002 Russian Census Microdata: 10% Sample of Private Households and Individuals // Rosstat; data-processing: Kopytok V., Kuzmina Y., Tikhonov S., The Research Development Infrastructure (RDI), CAG, 2021. Access: open for research purposes only. Posted: 16.12.2021. URL: http://data.rcsi.science/data-catalog/datasets/189</p>
Ссылка на открытый репозиторий	Отсутствует
Геоданные (пространственная привязка набора данных)	Нет
Лицензия, под которой публикуется набор данных	Простая (неисключительная) лицензия для обезличенных наборов данных (с возможностью скачивания).
Лицензия, под которой публикуется набор данных	<p>Вы можете скачать этот набор данных, а также выполнять с ним вычислительные и аналитические операции только в исследовательских и образовательных целях. Вы также можете в этих же целях передавать другим лицам результат выполненных операций, не допуская при этом деобезличивания исходных данных и приведя ссылку на платформу ИНИД. Вам запрещается каким-либо образом деобезличивать исходные данные, раскрывать их другим лицам. Подробные условия Лицензии приведены на странице.</p>
Стандарт публикации	http://opendata.gosmonitor.ru/standard/3.0
Параметры запросов API	Интеграция через API не предусмотрена

Структура набора данных

Набор данных разделен на восемь частей по федеральным округам:

- [census_2002_30000.csv](#) — Центральный федеральный округ;
- [census_2002_31000.csv](#) — Северо-Западный федеральный округ;
- [census_2002_33000.csv](#) — Приволжский федеральный округ;
- [census_2002_34000.csv](#) — Уральский федеральный округ;
- [census_2002_35000.csv](#) — Сибирский федеральный округ;
- [census_2002_36000.csv](#) — Дальневосточный федеральный округ;
- [census_2002_37000.csv](#) — Южный федеральный округ;
- [census_2002_38000.csv](#) — Северо-Кавказский федеральный округ.

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_open.html](#)).

Известные ограничения данных

Источники данных

При формировании выборки для обеспечения конфиденциальности исключаются наблюдения, которые соответствуют:

- временно пребывающим на территории России (239 018 наблюдений);
- лицам из коллективных домохозяйств (2 283 762 наблюдений);
- лицам, проживающим в населенных пунктах, которые входят в состав закрытых административно-территориальных образований (ЗАТО) (1 291 663 наблюдений).

Десятипроцентная выборка строится с учетом ограничения максимального числа лиц, проживающих в одном домохозяйстве, — **не более семи человек**. В частных домохозяйствах с числом лиц не более семи проживало 136 477 181 человек (96,3% от общей численности населения частных домохозяйств без учета домохозяйств, проживающих в ЗАТО).

Описание проведенных проверок и их результатов

При подготовке выборки с микроданными исходные значения атрибутов корректировались только в том случае, если это требовалось для снижения детализации в целях сохранения конфиденциальности. В остальных случаях сохранены как названия атрибутов, так и их значения.

Переписные листы заполняются со слов респондентов. Сведения, которые респондент сообщил переписчику, дополнительно не проверяются, поэтому в некоторых случаях в выборке могут

встречаться неправдоподобные или несогласованные между собой значения атрибутов.

Сопоставление распределения наблюдений в разрезе регионов, пола и возраста в выборке и исходных микроданных (по которым строилась выборка) приведено в файле shares_2002.xlsx, который размещается вместе с выборкой.

Внесенные в набор данных изменения по сравнению с источниками

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 десятипроцентной выборки — не более семи 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 10. С шагом 10 отбираются номера домохозяйств, включаемых в выборку.

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

Источники данных

Атрибут	Значение
Полное наименование источника данных	Обработанные микроданные Всероссийской переписи населения 2002 года
Сокращенное наименование источника данных	Микроданные переписи населения 2002 года
Владелец источника данных	Федеральная служба государственной статистики
Краткое описание источника данных	Выборка сформирована на основе микроданных Всероссийской переписи населения 2002 года. Эти данные получены Росстатом после обработки индивидуальных переписных листов. Выборка была передана Росстатом в АНО «ЦПУР» в соответствии с Дополнительным соглашением от 10.12.2021 № 1 к Соглашению об информационном и экспертно-аналитическом взаимодействии между Росстатом и ЦПУР от 11.09.2020 № 8-С
Ссылка на источник данных	Прямая ссылка отсутствует, источник — информационная система ограниченного доступа
Перечень используемых классификаторов	ТЕРСОН-МО (территориальная единица разработки статистики о населении) версии 2010 года, адаптированный для населенных пунктов, существующих на момент проведения переписи в 2002 году.
Понятия, используемые в источнике	Частными домохозяйствами считаются домохозяйства, размещающиеся постоянно в обычных жилых помещениях — квартирах, индивидуальных (одноквартирных) домах, комнатах в общежитиях неквартирного типа, других жилых помещениях и помещениях, приспособленных для жилья. Выборка с микроданными построена по частным домохозяйствам. Домохозяйства бездомных и студенческие общежития также учитывались при формировании выборки. Коллективными домохозяйствами называются совместно проживающие группы людей, обычно объединенных общей целью, подчиненных общим правилам и совместно питающихся. К коллективным домохозяйствам относятся лица, постоянно живущие в институциональных учреждениях, то есть учреждениях социального и медицинского обслуживания, казармах, местах заключения, религиозных организациях. Наблюдения, соответствующие лицам, проживающим в коллективных домохозяйствах, не учитывались при формировании выборки
Методология и изменения методологии для источника данных	Подробная методология проведения переписи и подведения ее итогов отражена в Методологических пояснениях к микроданным Всероссийской переписи населения 2002 и 2010 гг.
Ссылки на методологию	Методологические пояснения к микроданным Всероссийской переписи населения 2002 и 2010 годов: http://www.gks.ru/free_doc/new_site/perepis2010/croc/Documents/Materials/metod_comments.docx

Ответственные за набор данных

№	ФИО ответственного лица	Электронная почта
1.	Копыток Витовт Константинович	
2.	Кузьмина Юлия Викторовна	ask@data.rcsi.science
3.	Тихонов Сергей Владимирович	

Источник —

https://wiki-data.rcsi.science/index.php?title=Выборка_10%25_микроданных_переписи_населения

_2002_r&oldid=1179