



РАБОТА С МИКРОДАННЫМИ ОБСЛЕДОВАНИЙ РИДМИЖ И ПЕРЕПИСЕЙ НАСЕЛЕНИЯ РОССИИ

**БИРЮКОВА СВЕТЛАНА
ТЫНДИК АЛЛА**

**Центр анализа доходов и уровня жизни
НИУ ВШЭ**

13 февраля 2014

РОДИТЕЛИ И ДЕТИ, МУЖЧИНЫ И ЖЕНЩИНЫ В СЕМЬЕ И ОБЩЕСТВЕ (РИДМИЖ)

- 2004, 2007, 2011. Более 11000 домохозяйств. Более 30 регионов РФ.
- Репрезентативное для всего населения страны в возрасте 18-79 лет.
- Выборка: принцип построения аналогичен принципу построения выборки РМЭЗ (до изменений).

Принцип многоступенчатого вероятностного отбора жилищ, в которых затем выбирали домохозяйства, а в домохозяйстве случайным образом (с использованием процедуры Киша) отбирали одного респондента.

- 2007 и 2011 год – «добор» респондентов (1) в возрасте 18-20 лет (старение панели) + новая независимая выборка



РАЗДЕЛЫ ОБСЛЕДОВАНИЯ

- Домохозяйство;
- Дети;
- Брак(и)/союз(ы);
- Распределение домашних обязанностей;
- Родители и родительский дом;
- Беременность, бесплодие и планы иметь детей;
- Здоровье и благополучие;
- Деятельность и доходы респондента;
- Деятельность и доходы партнера;
- Имущество домохозяйства, доходы и трансферты;
- Ценности и установки;
- Пенсионное обеспечение и пенсионная реформа

Сохранена структура опросника во всех трех волнах.

По возможности сохранены набор и формулировки вопросов для обеспечения возможности сопоставлений.

НО

внутри разделов между волнами могут быть различия.



КАРТОЧКА ДОМОХОЗЯЙСТВА

КАРТОЧКА ДОМОХОЗЯЙСТВА

1	2	3	4	5	6	7	8	9	10	11
	№	Временно отсутствует	Имя, Отчество	Степень родства по отношению к Респонденту	Пол М-1 Ж-2	Возраст	Месяц и год рождения	Занятия	Месяц и год начала жизни с Р (только для неродных детей)	Нуждается в уходе
Р	1	1		Р	1...2	□□	М □□□ Г □□□	□□□		1
П	2	1		□□	1...2	□□	М □□□ Г □□□	□□□		1
Дети респондента	3	1		□□	1...2	□□	М □□□ Г □□□	□□□	М □□□ Г □□□	1
	4	1		□□	1...2	□□	М □□□ Г □□□	□□□	М □□□ Г □□□	1
	5	1		□□	1...2	□□	М □□□ Г □□□	□□□	М □□□ Г □□□	1
	6	1		□□	1...2	□□	М □□□ Г □□□	□□□	М □□□ Г □□□	1
	7	1		□□	1...2	□□	М □□□ Г □□□	□□□	М □□□ Г □□□	1
	8	1		□□						
другие члены	9	1		□□						
	10	1		□□						
	11	1		□□						
	12	1		□□						
	13	1		□□						
	14	1		□□						

Карточка 1: Степень родства членов домохозяйства по отношению к респонденту

- 1 – партнер или супруг(а)
- 2 – родной ребенок от нынешнего партнера или супруга(и)
- 3 – родной ребенок от бывшего партнера или супруга(и)
- 4 – пасынок/падчерица – не усыновленный мной ребенок моего нынешнего партнёра/супруга (и)
- 5 – усыновленный ребенок – ребёнок, официально усыновленный мной
- 6 – приемный ребенок - не родной ни мне, ни партнёру/супругу(е) и официально не усыновленный
- 7 – мама/ папа - родные или официально меня усыновившие
- 8 – мачеха или отчим - официально меня не усыновившие - или опекун

- 9 – мама/ папа нынешнего партнера или супруга(и) - родные или официально его/её усыновившие
- 10 – мачеха/отчим нынешнего партнера или супруга(и), официально его/её не усыновившие или опекун
- 11 – внук/внучка или правнук/правнучка (мой/моя или партнера/супруга(и))
- 12 – бабушка/дедушка или прабабушка/прадедушка - мой или партнера/супруга(и)
- 13 – мой брат или сестра
- 14 – брат или сестра партнера или супруга(и)
- 15 – другой мой родственник
- 16 – другой родственник партнера
- 17 – не родственник

Обследование проходит один человек

Он отвечает про всех членов д/х

Связующее звено - карточка

КАРТОЧКА ДОМОХОЗЯЙСТВА

- В массивах данных сведения из карточки домохозяйства идут в начале, до переменных первого раздела
- Во второй и третьей волнах в вопроснике есть специальный раздел, в котором собираются сведения о вышедших из состава д/х членах и появившихся в нем (см. вопросники для повторно участвующих)
- По карточке и этому разделу в массив вносятся переменные
 - а, b, с в начале – идентификатор волны обследования
 - «р» = partnter, «d» = deti, «o» = other

9	bp3	Numeric	1	0	Партнер, проживавший с Вами в 2006 году, сейчас является членом домохозяйства?	{1, да}...	-6
10	bp5	Numeric	2	0	Кем Вам приходится партнер, проживавший с Вами в 2006 году и живущий с Вами т...	{1, партнер ...	-6
11	bp6	Numeric	1	0	Каков пол партнера, проживавшего с Вами в 2006 году и живущего с Вами теперь?	{1, мужской...	-6
12	bp7m	Numeric	8	2	Месяц рождения партнера, проживавшего с Вами в 2006 году и живущего с Вами те...	None	-6,00
13	bp7y	Numeric	8	2	Год рождения партнера, проживавшего с Вами в 2006 году и живущего с Вами тепе...	None	-6,00
14	bp8	Numeric	8	0	Сколько полных лет партнеру, проживавшего с Вами в 2006 году и живущего с Вам...	None	-6
15	bp9	Numeric	2	0	Какая из позиций лучше всего описывает то, чем партнер, проживавший с Вами в 20...	{1, учится и...	-6
16	bp11	Numeric	1	0	Партнер, проживавший с Вами в 2006 году и живущий с Вами теперь, нуждается в ...	{0, нет}...	-6
17	bd35	Numeric	2	0	Кем Вам приходится первый ребенок?	{1, партнер ...	-6
18	bd36	Numeric	1	0	Каков пол первого ребенка?	{1, мужской...	-6
19	bd37_m	Numeric	3	0	Месяц рождения первого ребенка?	None	-6
20	bd37_y	Numeric	3	0	Год рождения первого ребенка?	None	-6
21	bd38	Numeric	2	0	Сколько полных лет первому ребенку?	None	-6
22	bd39	Numeric	2	0	Какая из позиций лучше всего описывает то, чем первый ребенок занимается в наст...	{1, учится и...	-6
23	bd311	Numeric	1	0	Первый ребенок нуждается в уходе вследствие физического или умственного забол...	{1, да}...	-6
24	bd45	Numeric	2	0	Кем Вам приходится второй ребенок?	{1, партнер ...	-6
25	bd46	Numeric	1	0	Каков пол второго ребенка?	{1, мужской...	-6
26	bd47_m	Numeric	3	0	Месяц рождения второго ребенка?	None	-6
27	bd47_y	Numeric	3	0	Год рождения второго ребенка?	None	-6
28	bd48	Numeric	3	0	Сколько полных лет второму ребенку?	None	-6
29	bd49	Numeric	2	0	Какая из позиций лучше всего описывает то, чем второй ребенок занимается в насто...	{1, учится и...	-6
30	bd411	Numeric	1	0	Второй ребенок нуждается в уходе вследствие физического или умственного забол...	{1, да}...	-6

2 АНКЕТЫ В ВОЛНАХ 2007 И 2011 ГГ.

- Вопросники для вновь входящих в выборку и повторно участвующих отличаются друг от друга ⇒ в массиве одни и те же переменные могут встречаться дважды, иметь разные названия...

!ПРИ РАБОТЕ НУЖНО ВСЕ ВРЕМЯ СМОТРЕТЬ В ВОПРОСНИКИ!

2.51. Сколько их?

_____ ВНУКОВ

впервые участвующие

У РЕСПОНДЕНТА ОДИН ВНУК ИЛИ ВНУЧКА 1 ⇒ [ПЕРЕХОДИТЕ К 2.55.]

У РЕСПОНДЕНТА ДВОЕ И БОЛЕЕ ВНУКОВ ИЛИ ВНУЧЕК 2 ⇒ [ЗАДАВАЙТЕ 2.53.]

2.53. Когда родился(лась) Ваш(а) старший(ая) (внук/внучка)? Назовите месяц и год.

|_|_| МЕСЯЦ

|_|_| ГОД

2.54. А младший(ая)?

|_|_| МЕС

|_|_| ГОД

2.51. Сколько их?

[ИНТЕРВЬЮЕР! ПОСМОТРИТЕ НА ТАБЛИЦУ ИНФОРМАЦИИ 2007 ГОДА ПОД КАРТОЧКОЙ ДОМОХОЗЯЙСТВА: СКОЛЬКО ВНУКОВ БЫЛО У РЕСПОНДЕНТА В 2007 ГОДУ? ЕСЛИ СЕЙЧАС ИХ НАЗВАНО МЕНЬШЕ - УТОЧНИТЕ, В СВЯЗИ С ЧЕМ.]

_____ ВНУКОВ

2.52. У Вас появились внуки с лета 2007 года, когда проводился предыдущий опрос?

ДА 1

НЕТ 2 ⇒ [ПЕРЕХОДИТЕ К 2.56.]

2.53.1. С лета 2007 года сколько внуков у Вас появилось?

ОДИН ВНУК 1 ⇒ [ПЕРЕХОДИТЕ К 2.55.]

ДВА И БОЛЕЕ ВНУКОВ 2 ⇒ [ЗАДАВАЙТЕ 2.53.]

2.53. Из тех внуков, которые родились с лета 2007 года, когда родился(лась) Ваш(а) старший(ая) (внук/внучка)? Назовите месяц и год.

|_|_| МЕСЯЦ

|_|_| ГОД

повторно участвующие

РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРКИ

- Структура генеральной совокупности: данные макростатистики (Росстат). Есть отклонения.
- Выборка приводится в соответствие с ней путем перевзвешивания наблюдений по (1) полу, (2) возрасту и (3) доходу
- Используемые в расчетах веса:
[gender age income]
 - 2001 г.: whgt_gai04
 - 2007 г.: wght_gai
 - 2011 г.: wght_gaiK2

!Не относится к работе с панелью!



ВЫЧИЩЕНИЕ ОШИБОК В МАССИВАХ РИДМИЖ

- «Чистка» переменных:
 - Несоответствия/ошибки внутри одного массива
 - Несоответствия в панельной выборке
- Проведена работа по проверке всех детей респондента при переходе от 2007 к 2011 году (А. Тындик). Есть новые переменные и описание алгоритма чистки.
- Внимательно работать с массивом 2011 г.!



РидМиЖ: ЕАЭСД НИУ ВШЭ

- Массивы переданы в архив, но пока не размещены на сайте (<http://sophist.hse.ru/>)
- «Линейный» формат представления данных значительно сокращает возможности работы с массивом. Полностью теряется возможность работы с панелью.
- *«Вследствие технических ограничений ЕБСД размещенные в ней массивы обследования РидМиЖ представлены в линейном формате, который не позволяет проводить полноценную работу с панельной выборкой. Для получения возможности работы с оригинальными массивами данных с целью проведения научных исследований просьба обращаться в Независимый институт социальной политики (<http://www.socpol.ru/>) или в Центр анализа доходов и уровня жизни населения НИУ ВШЭ (<http://csils.hse.ru/>).»*



ПАНЕЛЬ РИДМИЖ

○ 2004 год = 11261 респондент



○ 2007 год = 7786 старых + 3331 новый
= 11117 респондентов



○ 2011 год = 7419 старых + 3765 новых = 11184 респ.



из них:

5622 респ. – 2004, 2007 и 2011

1778 респ. – только 2007 и 2011

19 респ. – только 2004 и 2011



ПАНЕЛЬНЫЕ ВЫБОРОЧНЫЕ ОБСЛЕДОВАНИЯ: МЕТОДИКА РАБОТЫ

- Два вида панельного массива данных:
 - Полная панель
 - респонденты одни и те же
 - используется для анализа динамики изменений
 - используется для «чистки» переменных
 - Pooled cross-sectional
 - часть респондентов одинакова, часть – разная
 - чаще используется для анализа, где каждое наблюдение трактуется как отдельный респондент (расширение объема выборки)



СКЛЕИВАНИЕ ПАНЕЛЬНОГО МАССИВА РИДМИЖ

- Идентификаторы респондента:

- anumber 2004-2007-2011
- bredid 2007-2011
- credid 2011

Сортировка!!

- Склеивая 2 волны, можно получить 4 разных массива

1 волна: 11261 респондент



2 волна: 11117 респондентов

- Полная панель: 7786

- Неполные панели: 7786 (2раза) +3475 или 7786 (2раза) +3331

- Pooled cross: 14592



СКЛЕИВАНИЕ ПАНЕЛЬНОГО МАССИВА РИДМИЖ

- Как получить 7786:
 - удалить из второй волны новых респондентов
 - к ней приклеить 1-ую волну методом “non-active dataset is keyed table”
- Как получить 14592:
 - склеить 1 и 2 методом “both files provide cases”
- Syntax: ***MATCH FILES***
- NB: имена переменных должны различаться

