



# ВСЕРОССИЙСКОЕ ОБСЛЕДОВАНИЕ ДОМОХОЗЯЙСТВ ПО ПОТРЕБИТЕЛЬСКИМ ФИНАНСАМ (5-Я ВОЛНА)

Методология

2023

Методология подготовлена ООО «Демоскоп».

Содержание настоящего доклада отражает личную позицию авторов. Содержание и результаты доклада не следует рассматривать, в том числе цитировать в каких-либо изданиях, как официальную позицию Банка России или указание на официальную политику или решения регулятора. Любые ошибки в данном материале являются исключительно авторскими.

Все права защищены. Воспроизведение представленных материалов допускается только с разрешения авторов

Фото на обложке: Shutterstock/FOTODOM

Адрес: 107016, Москва, ул. Неглинная, 12

Телефоны: +7 495 771-91-00, +7 495 621-64-65 (факс)

Официальный сайт Банка России: [www.cbr.ru](http://www.cbr.ru)

© Центральный банк Российской Федерации, 2023

## Методология проведения обследования финансового поведения домохозяйств на тему «Всероссийское обследование домохозяйств по потребительским финансам» (5-я волна)

### 1. Дизайн обследования

Обследование финансового поведения домохозяйств относится к категории лонгитюдных. Термин «лонгитюдное» применяется к крупномасштабным, протяженным во времени обследованиям, предполагающим, что основная часть выборки в каждой волне обследования представляет собой панель, то есть единицы наблюдения в этой части выборки (в данном случае домохозяйства) остаются теми же самыми, что были отобраны для проведения первой, базовой волны.

Лонгитюдность обследования дает ему неоспоримые преимущества перед любыми повторяющимися с определенными промежутками обследованиями, в которых для каждой волны формируется новая независимая репрезентативная выборка. Это происходит потому, что появляется возможность исследовать изменения во времени, происходящие на индивидуальном уровне (в данном случае на уровне домохозяйства или индивида), что позволяет существенно расширить круг решаемых задач. Например, в случае изучения проблемы бедности данные, полученные с помощью повторяющихся обследований, позволяют оценить увеличение или уменьшение во времени числа домохозяйств с среднедушевым доходом ниже прожиточного уровня или изменение структуры типов домохозяйств, составляющих данную группу (агрегированные данные), но не позволяют ответить на вопрос как долго отдельное домохозяйство находится в группе бедных и, соответственно, оценить среднюю продолжительность нахождения домохозяйств в состоянии ниже уровня бедности. А это очень важно с точки зрения разработки государственных программ помощи малоимущим. И примеров такого рода задач множество.

В данном обследовании изначально ставилась задача, чтобы в дополнении к лонгитюдным оценкам была возможность делать и одномоментные оценки для каждой волны. Для решения этой задачи был использован специальный дизайн лонгитюдного обследования, известный как сплит-панель (*Split panel*). Сплит панель (расщепленная панель), представляющая собой некую комбинацию одномоментной и панельной выборок в каждой волне обследования, была предложена Лесли Кишем (L.Kish) в 1987 году. Используемый дизайн представляет собой серию одномоментных опросов, в которых в последующих волнах сохраняется максимально возможная доля элементов исходной выборки. Такой дизайн обычно определяется как обследование с пересекающимися

(перекрывающимися) выборками (overlapping survey) и может рассматриваться как вариант сплит-панели. В этом случае целью исследования является получение последовательности одномоментных оценок с сохранением возможности лонгитюдных оценок для большей части исходной выборочной совокупности.

## **2. Общий обзор исходной выборки домохозяйств по Российской Федерации**

Для проведения обследования с целью изучения финансового поведения населения используется **дизайн (модель) стратифицированной, многоступенчатой, вероятностной, территориальной адресной выборки.**

Для того, чтобы провести данное обследование необходимо сформировать выборку домохозяйств. Под "домохозяйством" понимаются люди, проживающие вместе по одному адресу и имеющие общие доходы и расходы. Любые временно проживающие лица (например, гости), имеющие другое постоянное место жительства, не являются членами домохозяйства. Респондентами являются все члены домохозяйств, начиная от 18 лет, находящиеся по месту жительства на время проведения обследования.

Хотя для целей исследования необходимо создать предварительную выборку домохозяйств, в общепринятой мировой практике в этом случае формируется выборка жилищ (адресов), в которых эти домохозяйства проживают. Это происходит потому, что только о жилищах (а точнее их адресах) имеется необходимая, привязанная к территории статистическая информация и, соответственно, существует возможность их переписать и обследовать. Прежде чем сформировать выборку домохозяйств, составляются пронумерованные списки жилищ таким образом, чтобы практически во всех случаях в каждом жилище находилось только одно домохозяйство. При этом условии полученная выборка жилищ фактически является выборкой домохозяйств.

Поскольку всероссийская выборка строится по территориальному признаку, на первом этапе необходимо выбрать первичную территориальную единицу отбора (ПЕО). В качестве основы для создания ПЕО выбираются административно-территориальные единицы, которые составляют основу административно-территориального деления субъектов РФ.<sup>1</sup> Административно-территориальные единицы (АТЕ) сгруппированы в 2029 преобразованных административно-территориальных единиц по территориальному признаку, которые и являются первичными единицами отбора. ПЕО группируются в 38 страт, главным образом на основании географических факторов и уровня урбанизации. В

---

<sup>1</sup> Основания, по которым в качестве основы для создания первичных единиц отбора выбраны административно-территориальные единицы и правила создания ПЕО, будут подробно изложены в разделе «Принципы выбора и формирования первичных единиц отбора (ПЕО)».

необходимых случаях в качестве стратообразующего фактора используется этническая составляющая.

В обследовании, как и в большинстве общенациональных выборочных обследований, предусматривающих личные (face-to-face) интервью по месту проживания респондентов, некоторые отдаленные и малонаселенные районы РФ исключаются из выборки по финансовым соображениям.

Из оставшихся районов, где проживает 95,7% населения России, как обязательные отбираются три крупнейших населенных конгломерата – г. Москва, Московская область и г. Санкт-Петербург, каждый из которых, вследствие своего объема, сам образует отдельную, так называемую саморепрезентирующую (СР) страту. Остальные, преобразованные административно-территориальные единицы группируются в 35 несаморепрезентирующих страт (НСР) примерно с одинаковой численностью населения. То есть всего формируются 38 страт. Затем, по методу отбора с «вероятностью пропорциональной численности» (PPS) из каждой несаморепрезентирующей страты выбирается по одному району. Это означает, что вероятность выбора данного района в данной страте является прямо пропорциональной доле численности населения этого района в общей численности населения всей страты.

Из общего целевого (планируемого) объема выборки 17,6% (соответственно доле в населении РФ) распределяются по 3-м саморепрезентирующим стратам. В соответствии с принципами отбора с «вероятностью пропорциональной численности» (PPS), остальные домохозяйства распределяются равномерно по преобразованным административно-территориальным единицам - первичным единицам отбора (ПЕО), взятым по одной в каждой из 35-ти несаморепрезентирующих страт примерно одинакового объема. Отбор преобразованной административно-территориальной единицы из страты осуществляется с «вероятностью пропорциональной численности».

Ввиду отсутствия сводного списка домохозяйств или жилищ ни по одной из 38 ПЕО, в соответствии с сложившейся практикой вводится промежуточная стадия отбора. Население каждой ПЕО стратифицируется на городское и сельское (подстраты), и объем целевой выборки в ПЕО делится пропорционально доле населения в каждой из этих подстрат. Например, если 40% населения данной первичной единицы отбора проживает в сельской местности, 40 из 100 адресов (жилищ) выбираются в сельской местности.

Как для городского, так и для сельского населения вторичными единицами отбора (ВЕО) являются населенные пункты. Городские населенные пункты стратифицируются по численности, уровню промышленности и удаленности от районных центров. Объем выборки распределяется пропорционально доле населения в каждой из этих страт для

городских населенных пунктов. Конкретные городские населенные пункты отбираются с «вероятностью пропорциональной численности» (PPS). Сельские населенные пункты (с.н.п.) стратифицируются только по двум параметрам: численности и удаленности от районных центров, в некоторых административно-территориальных районах к этим параметрам добавляется национальный состав сел. Отбор с.н.п. происходит по процедуре аналогичной отбору городских населенных пунктов. Следующей единицей отбора в с.н.п. являются жилища (адреса).

В городах, пгт появляется еще один, третий уровень отбора. Третичными единицами отбора (ТЕО) в городских населенных пунктах являются избирательные участки. В случаях, когда численность населения, проживающего на избирательных участках, сильно различается, для отбора участков используется метод отбора пропорционально численности. В отобранных избирательных участках список жилищ составляется с помощью специального натурного обхода территории участка.

В сельских населенных пунктах в качестве списка жилищ используются похозяйственные книги.

Затем производится систематический отбор необходимого количества жилищ, начиная с адреса, выбранного из списка случайным образом. Если во время натурного обхода обходчик определяет, что по какому-то адресу проживает более одного домохозяйства, он должен внести в список жилищ не одно, а столько жилищ, сколько домохозяйств проживает по данному адресу.

**И в городских, и в сельских поселениях от интервьюеров требуется посетить каждое жилище не менее трех раз для установления контакта домохозяйством и отдельными респондентами. Замены любого вида не допускаются.**

На протяжении всех волн проводится опрос домохозяйств именно из жилищ исходной выборки, независимо от того опрашивалось ли эти домохозяйства в предыдущих волнах. Если даже в какой-либо волне домохозяйство, проживающее по данному адресу, отказывается участвовать в опросе, то в последующих волнах интервьюер обязан повторять попытки контакта с данным домохозяйством вплоть до его категорического отказа. Если домохозяйство переехало, то опросу подлежит новое домохозяйство, проживающее в данном жилище на момент опроса. Если это домохозяйство тоже недоступно или категорически отказывается от опроса, данное исходное жилище заменяется другим жилищем (домохозяйством), отбираемым с использованием того же самого дизайна, в пределах того же опросного участка или населенного пункта, где была зафиксирована потеря. Такой подход, он называется «повторяющейся выборкой жилищ», дает

возможность репрезентации генеральных совокупностей на момент проведения каждой волны обследования.

Постепенное выбывание элементов исходной выборки в лонгитюдных обследованиях является естественным процессом и называется эффектом осыпания панели (attrition). Объем исходной части выборки постепенно уменьшается и для поддержания целевого объема выборки выбывшие домохозяйства должны замещаться. По сравнению с дизайном лонгитюдного обследования с «фиксированной панелью» используемый дизайн «сплит-панели» позволяет расширить возможности лонгитюдного анализа путем включения в него домохозяйств с более короткими траекториями участия.

Так как осыпание и старение исходной выборочной совокупности не являются случайными, то с течением времени наблюдается смещение такого демографического показателя как возраст, и необходимо через определенное количество волн проводить работу по «ремонту выборки», а именно, замещение части выборки новыми домохозяйствами, содержащими членов домохозяйств, принадлежащих к недопредставленным в ней возрастным группам, для компенсации последствий осыпания и естественного старения исходной выборки. Без такого замещения в 2022г ожидается продолжение тенденции превышения доли респондентов пенсионного возраста.

Выборка строится таким образом, чтобы, по возможности, исключить необходимость перевзвешивания.

### **3. Принципы выбора и формирования первичных единиц отбора (ПЕО)**

Составляющей большинства моделей территориальных выборок является их многоступенчатость, так как объектом территориальной выборки является население (или определённая часть населения), проживающее на обследуемой территории, а в России не существует списков людей, домохозяйств или жилищ по объектам административно-территориального деления Российской Федерации. Более того, использование таких списков для построения выборки по крупным территориальным объектам, таким как Российская Федерация, в целом, или субъекты РФ, представляется с практической точки зрения задачей, лишённой всякого смысла, так как в случае отбора респондентов из этого списка они оказались бы распылены по всей обследуемой территории и проведение обследования по такой выборке потребовало бы огромных и, что не менее важно, теоретически мало обоснованных расходов. В таких случаях и используется **многоступенчатый отбор**. При этом генеральная совокупность разбивается каким-либо

естественным образом на отдельные подсовокупности, «кластеры», которые и служат единицей отбора на первом шаге (ступени) выборки (первичная единица отбора - ПЕО), а последующий отбор единиц наблюдения осуществляется уже только в отобранных на первом шаге кластерах. В отличие от уже упомянутых ранее страт, кластеры по численности относительно невелики, но многочисленны. Первичные единицы отбора в многоступенчатой вероятностной выборке задают первый уровень кластеризации единиц наблюдения в генеральной совокупности. Теория выборки предъявляет основное требование к таким промежуточным единицам отбора (кластерам) - они должны быть максимально гетерогенными по изучаемым признакам.

На практике к этому основному теоретическому требованию к ПЕО добавляются ещё несколько, связанных с особенностями проведения массовых социологических обследований. При выборе ПЕО мы исходим из следующих требований к ПЕО при использовании территориальной выборки:

1) ПЕО должны иметь чётко зафиксированные географические (территориальные) границы. Для них должны существовать и быть доступны статистические материалы, необходимые для построения выборки.

2) ПЕО должно быть достаточно много, чтобы ошибка выборки на первой ступени не была слишком велика.

3) Население ПЕО должно быть достаточно велико, чтобы обеспечить возможность проведения исследований в течение нескольких лет в одних и тех же отобранных ПЕО. Разработка выборки, а тем более создание интервьюерской сети требует очень много затрат, чтобы пренебрегать данным условием, если исследуется население достаточно крупного территориального объекта.

4) Территория ПЕО должна обеспечивать приемлемые расстояния для возможностей поездки интервьюеров непосредственно в точки опроса.

Что конкретно означает 2-е требование к ПЕО?

Рассмотрим стандартную ошибку для многоступенчатой выборки. Пусть мы имеем  $K$ -ступенчатую выборку. При этом исследуемая совокупность состоит из  $N_1$  единиц первой ступени, каждая из которых содержит  $N_2$  единиц второй ступени и т.д. Пусть также при реализации выборки на каждой ступени было отобрано, соответственно,  $n_1, n_2, \dots, n_k$  единиц. Тогда, если на каждой ступени применялся простой случайный отбор, выборочное среднее есть несмещённая оценка среднего значения для генеральной совокупности с дисперсией:

$$V(y) = (1-f_1) \cdot S_1^{**2} / n_1 + (1-f_2) \cdot S_2^{**2} / (n_1 \cdot n_2) + \dots + f_1 \cdot f_2 \cdot f_3 \cdot \dots \cdot (1-f_k) \cdot S_k^{**2} / (n_1 \cdot n_2 \cdot \dots \cdot n_k) \quad (3.1)$$

где  $S_i$  - дисперсия среднего значения в единице отбора  $i$ -ой ступени,

$f_i = n_i / N_i$  - частота отбора на  $i$ -ой ступени,

$1-f_i$  - поправка на конечность совокупности на  $i$ -ой ступени

Несмещённая оценка  $V(y)$  по выборке есть:

$$v(y) = (1-f_1) \cdot s_1^2 / n_1 + f_1 \cdot (1-f_2) \cdot s_2^2 / (n_1 \cdot n_2) + \dots + f_1 \cdot f_2 \cdot f_3 \cdot \dots \cdot (1-f_k) \cdot s_k^2 / (n_1 \cdot n_2 \cdot \dots \cdot n_k) \quad (3.2)$$

где  $s_i$  - выборочные аналоги  $S_i$

Из формулы (3.1) видно, что, если размер выборки фиксирован, то каждая ступень добавляет свою долю в дисперсию, то есть чем меньше ступеней, тем меньше должна быть стандартная ошибка. А это в свою очередь означает, что лучше всего с теоретической точки зрения двухступенчатая или трехступенчатая выборка (одноступенчатая выборка невозможна, так как для этого нет основы выборки - организованного списка жилищ на территории РФ). Общепринятой нормой для стандартной ошибки является 10% среднего квадратичного отклонения, а из формулы (3.2) видно, что это условие достигается при  $n_1 > 100$

Российская Федерация подразделяется на 89 субъектов Федерации. Субъекты Федерации в свою очередь содержат 2775 базовых административно-территориальных единиц (в том числе 1868 районов, 579 города республиканского, краевого, областного, окружного подчинения, 328 внутригородских районов, городских округов).

Число субъектов РФ слишком мало (не удовлетворяют пункту 2 требований к ПЕО) и они слишком велики по территории (не удовлетворяют пункту 4 требований к ПЕО), чтобы служить в качестве первичных единиц отбора. Тогда как **административно-территориальные единицы являются почти идеальными первичными единицами отбора для построения репрезентативной выборки домохозяйств для Российской Федерации в целом.**

В связи с двумя особенностями определения административно-территориальных единиц в РФ возникает необходимость объединить некоторые из них, прежде чем начать формирование выборки. Во-первых, часть районов имеют в своих границах города федерального, республиканского или областного подчинения, выделенные в государственной статистике в самостоятельные административно-территориальные единицы. Поскольку в РФ таких городов много, такие самостоятельные города вносятся в состав районов, на территории которых они географически расположены. Это обеспечивает большую гетерогенность ПЕО и, соответственно, повышает качество выборки.

Далее, крупные российские города разделены на несколько районов. В соответствии со стандартными принципами формирования выборок такие города рассматриваются в выборке в качестве отдельных единиц. Таким образом, в результате внутреннего перераспределения исходных районов окончательный список первичных единиц отбора состоит из 2029 модифицированных административно-территориальных единиц (АТЕ).

**Намеренно исключенные территории.** Значительная часть территорий Российской Федерации находится на большом удалении, и плотность населения на этих территориях очень мала. Например, плотность населения в Эвенкийском автономном округе составляет в среднем всего один человек на 30 кв. км, в Камчатской области – один человек на 1 кв. км. Часть таких территорий заранее исключается из выборки. Из выборки исключаются территории, где проживает около 4,3% населения России – в связи с низкой плотностью населения, транспортными трудностями, сложной ситуацией для проведения опроса.

**Саморепрезентирующие территории.** Три субъекта Федерации: Москва, Московская область и Санкт-Петербург включаются в выборку автоматически. Вследствие высокой численности своего населения эти территориальные объекты на этапе стратификации образуют так называемые саморепрезентирующие страты.

**Стратификация.** Для повышения точности оценок проводится стратификация административно-территориальных единиц (ПЕО), не исключенных из выборки и не входящих в саморепрезентирующие территории.

Сначала формируется 10 модернизированных экономических регионов (см.: Таблицу 1). Регионы формируются таким образом, чтобы в каждом помещалось целое число страт заданной численности, и при этом минимально были бы изменены границы существующих экономических регионов. Затем эти регионы делятся на страты, в соответствии с показателем численности населения в каждом модернизированном регионе, и таким образом, чтобы объемы страт были примерно равны. Например, Уральский регион разделен на 6 страт, а Волго-Вятский регион на 3 страты. В Таблице 1 приведены 10 регионов и количество страт в каждом из них.

Таблица 1 - Десять модернизированных экономических регионов Российской Федерации (исключая 3 саморепрезентирующие и исключенные территории)

№	Регион	Кол-во Страт
1	Северный регион и Костромская область	2
2	Северо-западный регион	1
3	Центральный регион без Костромской области	4
4	Волго-Вятский регион	3
5	Центрально-Черноземный регион без юга Воронежской области	2

6	Поволжский регион без Астраханской и Пензенской областей и Калмыкии	4
7	Северо-Кавказский регион, Астраханская область, юг Воронежской области, Калмыкия	5
8	Уральский регион	6
9	Западно-Сибирский регион	4
10	Восточно-Сибирский и Дальневосточный регионы	4
<b>Итого</b>		<b>35</b>

Окончательное описание всех страт приведено в Таблицах 2 и 3. Первые 3 страты являются саморепрезентирующими территориальными объектами. Страты 4-38 представляют собой несаморепрезентирующие объекты. Следует обратить внимание на то, что, хотя страты с 4-й по 38-ю имеют примерно одинаковую численность населения (согласно дизайну выборки), количество ПЕО в стратах значительно варьируется (см. правую колонку).

Соответствующее количество страт формируется в каждом из 10 регионов на основе уровня урбанизации. Географические особенности и этническая структура также принимаются во внимание там, где эти факторы имеют важное значение.

Таблица 2 - Стратификация территории Российской Федерации – саморепрезентирующие страты

№	Саморепрезентирующие страты
1	г. Санкт-Петербург
2	г. Москва
3	Московская область

Таблица 3 - Стратификация территории Российской Федерации – несаморепрезентирующие страты

№	Несаморепрезентирующие страты
4	СЕВЕРНЫЙ РЕГИОН И КОСТРОМСКАЯ ОБЛ. Процент городского населения более 87%
5	Процент городского населения менее 87%
6	СЕВЕРО-ЗАПАДНЫЙ РЕГИОН Все районы региона

№	Несаморепрезентирующие страты
7 8 9 10	<p>ЦЕНТРАЛЬНЫЙ РЕГИОН (без Костромской обл.)</p> <p>Север: Владимирская, Ивановская, Тверская, Смоленская, Ярославская области</p> <p>Процент городского населения более 82%</p> <p>Процент городского населения менее 82%</p> <p>Юг: Брянская, Калужская, Орловская, Рязанская, Тульская области</p> <p>Процент городского населения более 79%</p> <p>Процент городского населения менее 79%</p>
11 12 13	<p>ВОЛГО-ВЯТСКИЙ РЕГИОН и Пензенская обл.</p> <p>Областные центры и столицы автономий с населением более 300 000</p> <p>Процент городского населения более 55%</p> <p>Процент городского населения менее 55%</p>
14 15	<p>ЦЕНТРАЛЬНО-ЧЕРНОЗЕМНЫЙ РЕГИОН (без юга Воронежской области)Процент</p> <p>городского населения более 75%</p> <p>Процент городского населения менее 75%</p>
16 17 18 19	<p>ПОВОЛЖСКИЙ РЕГИОН (без Астраханской, Пензенской областей и Республики Калмыкии)</p> <p>Казань Республика Татарстан</p> <p>Областные центры с населением более 900 000 (Волгоград, Самара, Саратов)</p> <p>Процент городского населения более 70%</p> <p>Процент городского населения менее 70%</p>
20 21 22 23 24	<p>СЕВЕРО-КАВКАЗСКИЙ РЕГИОН,</p> <p>Астраханская обл., юг Воронежской обл., Республика Калмыкия</p> <p>Авт. республики Северного Кавказа без Адыгеи</p> <p>Процент городского населения более 95%</p> <p>Процент городского населения 58-95%</p> <p>Процент городского населения 36,5-58%</p> <p>Процент городского населения менее 36,5%</p>
25 26 27 28	<p>УРАЛЬСКИЙ РЕГИОН</p> <p>Областные центры и столицы автономий (сдвоенная страта)</p> <p>Процент русского населения менее 45%</p> <p>Гор. населения более 93%; русского более 45%</p>

№	Несаморепрезентирующие страты
29	Гор. населения 67,5-93%; русского более 45%
30	Гор. населения менее 67,5%; русского более 45%
ЗАПАДНО-СИБИРСКИЙ РЕГИОН	
31	Новосибирск, Омск, Томск
32	Процент городского населения более 90%
33	Процент городского населения 57,5%-90%
34	Процент городского населения менее 57,5%
ВОСТОЧНО-СИБИРСКИЙ И ДАЛЬНЕВОСТОЧНЫЙ РЕГИОНЫ	
35	Восточная Сибирь – Гор. населения более 89%
36	Дальний Восток – Гор. населения более 84%
37	Гор. населения 64,7-89% (В.С.); 64,7-84% (Д.В.)
38	Городского населения менее 64,7%

#### 4. Выбор конкретных первичных единиц отбора (ПЕО)

Вероятностный отбор предполагает, что из каждой страты должна быть выбрана по крайней мере одна ПЕО.

В связи с этим в каждой несаморепрезентирующей страте выбирается одна ПЕО по методу: «отбора с вероятностью, пропорциональной численности» (ВПЧ). То есть, чем больше доля численности населения данной ПЕО в общей численности населения данной страты, тем больше вероятность ее выбора.

#### 5. Принципы выбора и формирования вторичных (ВЕО) и третичных (ТЕО) единиц отбора

В зависимости от типа ПЕО выборка может не иметь третьей ступени

<p><b>ПЕО – состоит из одного города</b></p>	<p><b>Внутри одного города, ВЕО являются избирательные участки.</b> Конкретные избирательные участки отбираются простым механическим отбором с фиксированным шагом из соответствующего списка избирательных участков города.</p> <p>На выбранных избирательных участках производится отбор домохозяйств для участия в опросе. Для этого используются списки жилищ, существующие на избирательных участках, которые проверяются путем натурного обхода. Отбор жилищ производится из скорректированного списка простым механическим отбором с фиксированным шагом.</p> <p>Если при натурном обходе выясняется, что в жилище проживает более одного домохозяйства, то обходящий случайным образом выбирает одно из них для включения в выборку домохозяйств.</p>
<p><b>ПЕО – состоит из городов, пгт и снп</b></p>	<p><b>При такой структуре ПЕО, вторичными единицами отбора (ВЕО) являются все три вида поселений.</b></p> <p>Первоначально осуществляется стратификация населения по численности на городское и сельское. Пропорционально долям распределяется выборка домохозяйств.</p> <p><b>Конкретные города и пгт (ВЕО) отбираются из списка городов и пгт с помощью метода механического отбора пропорционального численности.</b> Внутри каждого отобранного городского поселения, появляется третья ступень отбора - <b>избирательные участки, которые становятся третичной единицей отбора ТЕО.</b></p> <p>Конкретные избирательные участки отбираются простым механическим отбором с фиксированным шагом из соответствующего списка избирательных участков города.</p> <p>На выбранных избирательных участках производится отбор домохозяйств для участия в опросе. Для этого используются списки жилищ, существующие на избирательных участках, которые</p>

	<p>проверяются путем натурального обхода. Отбор жилищ производится из</p>
--	---

	<p>скорректированного списка простым механическим отбором с фиксированным шагом.</p> <p>Если при натурном обходе выясняется, что в жилище проживает более одного домохозяйства, то обходящий случайным образом выбирает одно из них для включения в выборку домохозяйств.</p> <p><b>Конкретные снп (ВЕО) отбираются из списка снп с помощью метода механического отбора пропорционального численности.</b></p> <p><b>В случае сельских поселения ТEO отсутствует, так как отбор домохозяйств производится по похозяйственной книге общей для всего снп.</b> Список в похозяйственной книге проверяется путем натурального обхода. Отбор жилищ производится из скорректированного списка простым механическим отбором с фиксированным шагом.</p>
--	---

## 6 Методики расчета оценок параметров обследования

Любые показатели, характеризующие генеральную совокупность, называются параметрами, в отличие от тех же показателей, рассчитанных для выборочной совокупности. В первую очередь, исследователей интересуют расчеты таких показателей генеральной совокупности или целевых групп этой совокупности, как суммы по совокупности, средние или их аналог для номинальных данных – доли. Эти показатели относятся к категории аддитивных. Процедура оценивания аддитивных параметров по значениям показателей в выборочной совокупности состоит в суммировании значений переменных, входящих в состав этих показателей в выборочной совокупности, умноженных на выборочные веса единиц наблюдения составляющих всю выборочную совокупность или определенные группы, интересующие исследователя. Критерии принадлежности единицы наблюдения могут быть либо количественными, например, определенный уровень дохода респондента или подушевого дохода домохозяйства либо качественными, например, принадлежность единиц наблюдения к определенному типу, например, дихотомия – имеющие депозиты в банках и не имеющие.

Приведенная ниже формула (6.1) используется для оценки суммарных показателей для изучаемых совокупностей по данным выборки:

$$Y_d = \sum_{k \in d} w_k y_k \quad (6.1)$$

В формуле (6.1) суммирование ведется по единицам наблюдения, относящимся к целевой группе ( $d$ ) генеральной совокупности (такой группой могут быть домохозяйства, имеющие в своем составе более двух детей);

$Y_d$  - оценка суммарного значения показателя исследования (например, суммарный доход домохозяйств по рассматриваемой группе);

$w_k$  - выборочный вес или фактор распространения, рассчитанный исходя из использованного дизайна выборки;

$y_k$  - значение признака  $k$ -ой единицы выборки.

Формула (6.2) используется для оценки средних показателей исследования по данным выборки:

$$\bar{Y}_d = \frac{\sum_{k \in d} w_k y_k}{\sum_{k \in d} w_k} \quad (6.2)$$

В формуле (6.2) суммирование ведется по единицам наблюдения в выборке, относящимся к целевой группе ( $d$ ) генеральной совокупности (такой группой могут быть домохозяйства, имеющие, например, в своем составе более двух детей);

$\bar{Y}_d$  - оценка среднего значения показателя исследования (например, средний доход домохозяйств по рассматриваемой целевой группе);

$w_k$  - выборочный вес или фактор распространения, рассчитанный исходя из использованного дизайна выборки;

$y_k$  - значение признака  $k$ -ой единицы выборки.

Для оценки по выборке доли в генеральной совокупности используется следующая формула:

$$P_d = \frac{\sum_{k \in d} w_k I_k}{\sum_{k \in d} w_k} \quad (6.3)$$

Суммирование в формуле (6.2), также, как и в (6.1), ведется по единицам выборки, относящимся к целевой группе ( $d$ ) генеральной совокупности (например, по пенсионерам);

$p_d$  - оценка доли, например, работающих пенсионеров;

$I_k$  - индикатор, который принимает значение 1 в случае наличия представляющего интерес свойства у  $k$ -ой единицы выборки или 0 иначе. Например, таким свойством может быть факт занятости пенсионера;

$w_k$  - выборочный вес или фактор распространения, рассчитанный исходя из использованного дизайна выборки;

В случае отсутствия специализированного программного обеспечения для получения распространенных выходных показателей и соответствующих ошибок выборки следует использовать расчетные формулы, которые будут описаны в следующем разделе.

## **7. Расчет выборочных ошибок с учетом дизайн-эффекта**

Уровень репрезентативности выборки оценивается качеством полученной информации, которая определяется степенью возможности распространения выборочных данных на генеральную совокупность и, в какой-то мере, на интересующие исследователя ее подсовокупности. Критерием достоверности распространенных показателей обычно выступает уровень коэффициента вариации оценки данного показателя (отношение величины стандартной ошибки к значению самой оценки).

Описываемая выборка создавалась как всероссийская и, соответственно, достоверность показателей должна рассчитываться именно на федеральном уровне. Оценки показателей, рассчитанные по всей генеральной совокупности в целом, или по целевым группам классификации, следует считать достоверными, если соответствующие значения коэффициента вариации не превосходят 2-3%. Это означает, что предельная ошибка выборки не превосходит 5%, полученного значения оценки представляющего интерес показателя.

Чтобы получить доверительный интервал оценки суммарного, среднего значений или доли нужно величину стандартной ошибки оценки скорректировать на множитель (квантиль), соответствующий выбранному уровню надежности. Обычно в социально-экономических исследованиях в качестве уровня надежности выбирается 95%-ая или 90%-ая доверительная вероятность. То есть значение показателя для генеральной совокупности с вероятностью 95% или 90% лежит в интервале: плюс или минус 5% от значения выборочного показателя. Соответствующие значения поправочного коэффициента в этом случае равны 1,96 или 1.64.

Для расчета стандартных ошибок результатов, получаемых по используемой стратифицированной, многоступенчатой, вероятностной, территориальной адресной выборки с отбором пропорционально численности населения, применяется формула, учитывающая дизайн-эффект (или эффект плана) данного сложного плана отбора.

Смысл дизайн эффекта в том, что он показывает, насколько данные, полученные с помощью используемой модели, при том же объеме выборки могут отличаться от данных, которые были бы получены при применении простого случайного отбора. Для стандартной ошибки среднего:

$$SE(\bar{y}_d) = \sqrt{\frac{\sigma_{y_o}^2}{n_d} \left( 1 + icc \left( \frac{n_d}{a_d^*} \right) \right)} \quad (7.1)$$

$SE(\bar{y}_d)$  - стандартная ошибка оценки среднего  $\bar{Y}_d$  (или доли  $p_d$ );

$\sigma_{y_o}^2$  - дисперсия признака ( $y$ );

$n_d$  - объем выборки в подгруппе ( $d$ ) генеральной совокупности;

$icc$  - предполагаемая величина внутригрупповой корреляции (intra-class correlation), которая точно не может быть рассчитана по данным выборки.

Внутригрупповая корреляция в целом измеряет степень различия кластеров - первичных единиц отбора (ПЕО). Если ее значение близко к нулю, то кластеры приблизительно одинаковые, что является необходимым условием для получения точных оценок показателей;

$a_d^*$  - эффективное число ПЕО, относящихся к целевой группе ( $d$ ) исследования.

Величина в правой части (7.1) в скобках под квадратным корнем  $\left( 1 + icc \left( \frac{n_d}{a_d^*} \right) \right)$  как

раз и является эффектом дизайна выборки (дизайн эффектом). Она является отношением дисперсии оценки показателя, рассчитанной с учетом плана выборки, к дисперсии оценки того же показателя, рассчитанной в предположении простого случайного отбора. Поэтому чем меньше дизайн эффект отличается от единицы, тем более точны распространенные результаты исследования.

Расчет эффекта дизайна по формуле (7.1) достаточно труден, поэтому он редко проводится в практике проведения массовых опросов в России, хотя многоступенчатые выборки для проведения массовых обследований используются довольно часто. Однако, отсутствие данных по величине дизайн-эффекта существенно уменьшает доверие к приводимым исследователями данным об оценке достоверности публикуемых показателей.

Проведенные в рамках обследования RLMS, использующего аналогичный дизайн формирования выборки, расчеты дизайн-эффекта показали, что дизайн-эффект практически для всех показателей, использованных в исследовании, близок к единице (или несущественно превосходит 1). Поэтому для расчета стандартной ошибки оценок в случае использования данного дизайна выборки можно воспользоваться упрощенными формулами простой случайной выборки.

Для расчета стандартной ошибки оценки среднего ( $\bar{Y}_d$ ):

$$SE(\bar{Y}_d) = \sqrt{\frac{s_y^2}{n_d} \left( 1 + \frac{n_d}{N_d} \right)} \quad (7.2)$$

$s_y^2$  - скорректированное среднее квадратическое отклонение признака (y) в выборке

Для расчета стандартной ошибки доли ( $p_d$ ):

$$SE(p_d) = \sqrt{\frac{p_d(1-p_d)}{n_d - 1} \left( 1 + \frac{n_d}{N_d} \right)} \quad (7.3)$$

Дизайн исходной выборки обеспечивает репрезентативность генеральных совокупностей домохозяйств и респондентов от 18 лет для отдельных групп населения Российской Федерации (в случае их достаточной численности относительно размера выборки), обычно интересующих исследователей, например, проживающих в городе или селе, или в городах с разной численностью населения и т.д.

Выборка не репрезентативна:

- относительно регионов РФ, так как отбор административно-территориальных единиц проводится для того, чтобы репрезентировать страту, а не отдельные регионы, входящие в эту страту, в которых конкретные (АТЕ) были выбраны;

- относительно людей и домохозяйств, находящихся в верхнем квинтиле по доходам, который практически не представлен в выборочной совокупности при массовых опросах. Для его учета и включения в выборку необходимо проводить специальное обследование.