Утверждена

Приказом Руководителя

Бюро национальной статистики Агентства по стратегическому планированию и реформам

Республики Казахстан

от «20» декабря 2023 года

№ 237

**Методика построения выборки домашних хозяйств**

**по обследованию уровня жизни**

### Общие положения

1. Методика построения выборки домашних хозяйств по обследованию уровня жизни (далее – Методика) относится к статистической методологии, формируемой в соответствии с международными стандартами и утверждаемой в соответствии с Законом Республики Казахстан «О государственной статистике».
2. Настоящая Методика предназначена для использования структурными подразделениями Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан при формировании выборки домашних хозяйств по обследованию уровня жизни.
3. Целью настоящей Методики является построение выборки для выборочного обследования домашних хозяйств по оценке уровня жизни (далее – Обследование), включая описание этапов формирования, стратификации и расчета выборочных весов.
4. Настоящая Методика соответствует принятым в мировой практике методам построения выборки и обеспечивает получение качественных первичных статистических данных.
5. Обследование по своему характеру является многоцелевым. Его основными задачами являются получение широкого спектра первичных статистических данных об уровне жизни населения, весовых показателей для расчета индекса потребительских цен и первичных статистических данных для составления счетов сектора домашних хозяйств в системе национальных счетов.
6. В настоящей Методике используются следующие понятия и сокращения:
7. домашнее хозяйство – экономический субъект, состоящий из одного или более физических лиц, проживающих совместно, объединяющих полностью или частично свои доходы и имущество, и совместно потребляющие товары и услуги;
8. интервьюер – лицо, осуществляющее опрос респондентов и домашних хозяйств при проведении общегосударственных статистических наблюдений и национальных переписей;
9. стандартная ошибка выборки (далее - СО)– стандартное отклонение значения параметра выборки от выборочного среднего значения этого параметра;
10. относительная стандартная ошибка (далее - ОСО)– отношение оцениваемой статистической величины к его среднему значению.

**2. Определение исследуемой совокупности (рамки выборки)**

7. Основным источником для формирования выборочной совокупности домашних хозяйств является информационная система «Статистический регистр жилищного фонда», компонент интегрированной информационной системы «е-Статистика» (далее – CРЖФ). СРЖФ содержит данные о жилых домах и жилых помещениях (квартирах), а также о численности людей, проживающих в них. Проживающие в них люди выступают в качестве домашних хозяйств при формировании выборки по обследованию уровня жизни.

8. Использование указанной базы данных обусловлено рядом преимуществ, к которым относятся следующие:

наличие готовой рамки выборки, что исключает одну из существенных статей затрат, связанных с составлением основы выборки;

наличие постоянно актуализируемой базы данных, что необходимо для осуществления ротации домашних хозяйств.

1. В генеральную совокупность включаются жилые дома и жилые помещения (квартиры) за исключением коммунальных квартир, общежитий, домов-интернат для престарелых и лиц с инвалидностью, детских домов, тюрьм, гостиниц, религиозных общин и других аналогичных жилых помещений.
2. Конечной единицей отбора определено домашнее хозяйство, являющееся также и единицей обследования.

**3. Размер выборочной совокупности**

1. Размер выборки определяется на основе принципа оптимального сочетания затрат и заданных критериев точности результатов.
2. В качестве показателей точности оценивания, используются СО и ОСО выборки.
3. СО определяется по формуле 1:



где,

SE – СО;

δ – дисперсия;

Deff – влияние дизайна выборки (для города и села составляют 1.0 и 2.0 соответственно);

n – размер выборки.

1. ОСО определяется по формуле 2:



где,

RSE – ОСО выборки;

SE – СО выборки;

 – среднее значение переменной, использованной для оценки величины ОСО.

Влияние дизайна выборки (Deff) для города и села составляют 1.0 и 2.0 соответственно. Исходя из этого предположения, рассчитываются ОСО. Предположение Deff = 1.0 для городской местности объясняется тем, что городские кластеры обследования приближены к случайной выборке, поскольку крупные города не подразделяются на более мелкие территориальные единицы. Предположение Deff = 2.0 в сельской местности основано исключительно на опыте других стран.

1. Размер выборки обеспечивает получение результатов с ошибкой не более 4% на республиканском уровне и не более 7% на региональном уровне.Таким образом, размер выборки устанавливается в 12 000 домашних хозяйств.

**4. Алгоритм формирования выборочной совокупности**

1. Выборочная совокупность домашних хозяйств формируется методом двухступенчатой вероятностной (случайной) выборки с использованием процедур стратификации и случайного отбора на каждой из ступеней формирования выборки. Процедура стратификации нацелена на формирование представительной выборочной совокупности домашних хозяйств, адекватно отражающей территориальные особенности расслоения населения.
2. Процесс формирования выборки проводится в два этапа.

На первом этапе генеральная совокупность стратифицируется по территориальному признаку, включая распределение на городскую и сельскую местность. Таким образом, формируется 37 страт - это отобранные городские и сельские территории в двадцати регионах страны (всего 37 страт с учетом, что в городах Астана, Алматы и Шымкент нет сельской местности).

1. В качестве первичных выборочных единиц (далее – ПВЕ) отбираются 400 территориальных единиц, которые представляют собой городскую и сельскую местности и определяются первыми шестью знаками кода по Классификатору административно-территориальных объектов.
2. Объем работы каждого из интервьюера составляет 30 домашних хозяйств в течение одного квартала. Каждая группа из 30 домашних хозяйств называется кластер и соответствует объему работы одного интервьюера в течение одного квартала.
3. Первый этап формирования выборки включает в себя две последовательные процедуры. Первоначальная процедура предусматривает определение ПВЕ внутри каждой страты с вероятностью пропорциональной размеру (далее – ВПР), то есть количество ПВЕ в страте формируется в зависимости от числа домашних хозяйств, присутствующих в страте.
4. Вероятность отбора (Phi) в выборе ПВЕ (hi) в страте (h) определяется по формуле 3:



где,

Phi– вероятность отбора;

sh – число ПВЕ, выбранных в страте h;

nhi – число домашних хозяйств в ПВЕ hi, по данным СРЖФ;

Nh – общее число домашних хозяйств в страте, по данным СРЖФ.

1. Следующая процедура нацелена на достижение оптимального значения ОСО по каждой из 37 страт (400 ПВЕ).
2. Для оценки величины ОСО, в качестве переменной используются показатели, которые считаются наиболее важными для обследования. По обследованию уровня жизни к таким показателям относят «среднедушевой доход, использованный на потребление», «среднедушевой денежный доход», «среднедушевой денежный расход».
3. Количество отобранных ПВЕ и число домашних хозяйств в каждой ПВЕ, а также их распределение по регионам приведены в таблице, согласно приложению к настоящей Методике.
4. При проведении фактического отбора ПВЕ в городах необходимо учесть, что в СРЖФ отсутствует деление крупных городов на мелкие территориальные единицы. Например: город Кокшетау с населением более 100 тысяч человек представлен в СРЖФ одной строкой, без дальнейшей дезагрегации. Учитывая относительно крупный размер таких ПВЕ, многие из них выбираются более одного раза в результате стандартной процедуры ВПР.
5. На втором этапе формирования выборки, в каждой ПВЕ случайным отбором выбираются 30 домашних хозяйств. Основой формирования выборки на втором этапе является перечень отдельных жилых помещений в ПВЕ. Кластер (или кластеры) жилищ, которые нужно посетить в ходе обследования, выбирается с равной вероятностью из числа подходящих жилых помещений в ПВЕ.
6. Вероятность (phij) выбора домашних хозяйств (hij) в ПВЕ (hi) страты (h) определяется по формуле 4:



где,

phi – вероятность выбора ПВЕ hi – заданная уравнением (1);

mhi – число необходимых жилищ в ПВЕ hi (нормативно всегда 30);

n’hi – общее число подходящих жилищ в ПВЕ.

**5. Сегментация выборки**

1. Для практического применения выборки важным фактором является территориальная расположенность отобранных единиц обследования друг от друга внутри одного кластера, так как отдаленность их друг от друга создает дополнительные затраты и неудобства при проведении обследования. Некоторые из отобранных ПВЕ оказываются слишком крупными по численности населения и единицы обследования в них находятся территориально разбросанными друг от друга.
2. При возникновении подобных случаев существует возможность корректировки выборки посредством процесса сегментации (дробления или разукрупнения).
3. Сначала в ПВЕ путем систематической равновероятностной выборки (далее – СРВВ) выбираются три населенных пункта с учетом числа заселенных в них домашних хозяйств по СРЖФ и затем в каждом из трех населенных пунктов, также путем СРВВ, отбираются 10 домашних хозяйств и 10 резервных домашних хозяйств (для замены).
4. Эта процедура является дополнительным этапом выборки и позволяет получить более кластерную выборку, вместе с тем она не приведет к изменению вероятностей отбора или к изменению весов отобранных домашних хозяйств. Следовательно, формулы 4 и 5 применяются без изменений в разукрупненных (сегментированных) ПВЕ.

**6. Смещение выборки**

1. В ходе обследования имеют место случаи, когда не удается опросить домашние хозяйства по причине того, что жилище не найдено, или не заселено, или домашнее хозяйство отказывается участвовать в обследовании. Отказ от участия является нежелательным явлением, поскольку сокращает размер выборки и в основном потому, что является источником потенциального смещения выборки и ведет к искаженным статистическим результатам.
2. Эффективной профилактикой преодоления проблемы неполучения первичных статистических данных является тщательное документирование каждого случая.
3. При возникновении сложности по соблюдению принципа «недопущения замены» следует предусмотреть список резервных домашних хозяйств в размере 30 домашних хозяйств в каждом кластере.

**7. Взвешивание**

1. В целях получения распространенных на генеральную совокупность первичных статистических данных, производится статистическое взвешивание итогов обследования. Реализация данного метода осуществляется посредством присвоения каждому обследованному домашнему хозяйству статистического веса, который характеризует общее число домашних хозяйств, представляемых частью, попавшей в выборку. Веса для показателей уровня жизни населения рассчитываются ежеквартально.
2. Для расчета весов используется первичные статистические данные СРЖФ о распределении обследуемых домашних хозяйств отдельно по городскому и сельскому населению в региональном разрезе.
3. Вес вероятности (whij) домашних хозяйств в ПВЕ (hi) страты hi(h) представляет собой обратное значение его вероятности отбора phij и определяется по формуле 5:



где,

Whij – вес домашниххозяйств;

Phij – обратное значение вероятности отбора;

Nh – общее число домашних хозяйств в страте, по данным СРЖФ;

n’hi – общее число подходящих жилищ в ПВЕ;

Sh – число ПВЕ, выбранных в страте h;

mhi – число необходимых жилищ в ПВЕ hi (нормативно всегда 30).

Поскольку n’hi ≈ n’hi и mhi ≈ 30 во всех ПВЕ, выборка будет примерно самовзвешенной в пределах каждой страты.

1. Сумма «базовых» весов обеспечивает оценку количества всех домашних хозяйств данного региона и страны в целом. Однако их использование позволяет сохранить соответствие выборочной совокупности изначальным этапам формирования выборки, при этом в практике имеют место случаи, нарушающие данное соответствие (отказы от участия в обследовании и другие случаи неполучения первичных статистических данных).
2. Недополучение ответов от отобранных домашних хозяйств является серьезным недостатком, искажающим результаты обследования. В этих условиях возникает необходимость компенсации недостающих первичных статистических данных. Данная процедура призвана осуществить досчет показателей, то есть компенсировать недополучение первичных статистических данных выборочного обследования.
3. Недополучение первичных статистических данных имеет место в случае, когда выборочная единица вообще не участвует в обследовании, либо принимает в нем частичное участие. Для компенсации случаев при отсутствий ответов применяется простая схема корректировки весов, путем задания больших весов для всех ответивших домашних хозяйств в данном населенном пункте. Веса всех ответивших на вопросы домашних хозяйств в данном населенном пункте увеличиваются на один и тот же коэффициент. Например: если 90 процентов домашних хозяйств в ПВЕ ответили на все вопросы, веса для всех ответивших увеличиваются на коэффициент 1/0,9=1,11.
4. В случае отсутствия первичных статистических данных только по определенным вопросам применяется метод вменения (импутации) пропущенных значений. Данный метод заключается в подстановке отсутствующих первичных статистических данных по отдельной позиции значением, в отношении которого делается прогноз с учетом первичных статистических данных по данному объекту или по другим объектам данного обследования.
5. Рассчитанные персональные веса в качестве дополнительных переменных записываются в базу данных по каждому домашнему хозяйству и используются при формировании сводных итогов по соответствующему кварталу на республиканском и региональном уровнях. Распространенные первичные статистические данные в среднем за год получаются путем объединения квартальных первичных статистических данных.

**8. Ротация выборки**

1. В целях недопущения эффекта усталости домашних хозяйств от участия в обследовании, выборка домашних хозяйств периодически подвергается ротации. Процесс систематической ротации выборки не связан с заменой домашних хозяйств, отказавшихся участвовать в обследовании.
2. В целях отслеживания влияния сезонных колебаний и других явлений на одни и те же домашние хозяйства в течение отчетного года, ротацию целесообразно проводить на начало года, то есть ежегодно в декабре необходимо обновлять выборку.
3. Ротацию домашних хозяйств (обновление выборки) необходимо проводить ежегодно в размере 1/4 (по 7-8 домашних хозяйств на одного интервьюера) от общего числа обследуемых домашних хозяйств. То есть, ежегодно 1/4 часть домашних хозяйств выводится из выборки и заменяется на другие. Таким образом, по истечении четырех лет произойдет полное обновление выборки.
4. При формировании выборки в списке отобранных домашних хозяйств предусмотрено дополнительно поле с признаком ротации. «1» в данном поле означает, что домашнее хозяйство наблюдается первый год подлежит ротации по истечению четырех лет, «2» – по истечению трех лет и «3» – по истечению двух лет, «4» – по истечению одного года. Таким образом, систематическая ротация осуществляется путем выведения из выборки групп с признаком –«4».
5. По субъективным причинам ежегодно из выборки выбывают домашние хозяйства, отказавшиеся от дальнейшего участия в обследовании, признак ротации которых меньше «4». По указанным домашним хозяйствам в период ежегодной ротации осуществляется дополнительная замена вне ротационного количества из резервного списка. При этом замена домашних хозяйств производится на те домашние хозяйства, которые относятся к тому же сегменту или кластеру. Замещающему домашнему хозяйству присваивается признак ротации выбывшего домашнего хозяйства(например, на конец отчетного периода у выбывшего домашнего хозяйства признак ротации был равен «2», то замещающему его домашнему хозяйству присвоится признак «3»).

Приложение

к Методике построения выборки домашних хозяйств по обследованию уровня жизни

**Таблица. Распределение первичных выборочных единиц по стратам (городская и сельская местность)**

|  |  |  |
| --- | --- | --- |
| Код | Наименование региона | Число первичных выборочных единиц |
| Город | Село | Всего |
| 10 | Абай | 8 | 9 | 17 |
| 11 | Акмолинская | 12 | 13 | 25 |
| 15 | Актюбинская | 12 | 12 | 24 |
| 19 | Алматинская | 6 | 10 | 16 |
| 23 | Атырауская | 10 | 7 | 17 |
| 27 | Западно-Казахстанская | 8 | 14 | 22 |
| 31 | Жамбылская | 9 | 12 | 21 |
| 33 | Жетісу | 8 | 8 | 16 |
| 35 | Карагандинская | 13 | 8 | 21 |
| 39 | Костанайская | 13 | 13 | 26 |
| 43 | Кызылординская | 8 | 12 | 20 |
| 47 | Мангистауская | 11 | 8 | 19 |
| 55 | Павлодарская | 12 | 14 | 26 |
| 59 | Северо-Казахстанская | 9 | 13 | 22 |
| 61 | Туркестанская | 7 | 13 | 20 |
| 62 | Ұлытау | 7 | 6 | 13 |
| 63 | Восточно-Казахстанская | 8 | 8 | 16 |
| 71 | г. Астана | 19 | ~~-~~ | 19 |
| 75 | г. Алматы | 24 | ~~-~~ | 24 |
| 79 | г. Шымкент | 16 |  | 16 |
|  | Всего | 220 | 180 | 400 |