

Об утверждении Методики по проведению выборочных обследований домашних хозяйств

Приказ Председателя Комитета по статистике Министерства национальной экономики Республики Казахстан от 11 ноября 2016 года № 266. Зарегистрирован в Министерстве юстиции Республики Казахстан 13 декабря 2016 года № 14515.

В соответствии с подпунктом 5) статьи 12 Закона Республики Казахстан от 19 марта 2010 года "О государственной статистике" и подпунктом 258) пункта 17 Положения о Министерстве национальной экономики Республики Казахстан, утвержденного постановлением Правительства Республики Казахстан от 24 сентября 2014 года № 1011 , **ПРИКАЗЫВАЮ:**

1. Утвердить прилагаемую Методику по проведению выборочных обследований домашних хозяйств.

2. Управлению статистических регистров и классификаций совместно с Юридическим управлением Комитета по статистике Министерства национальной экономики Республики Казахстан обеспечить в установленном законодательством порядке:

1) государственную регистрацию настоящего приказа в Министерстве юстиции Республики Казахстан;

2) в течение десяти календарных дней после государственной регистрации настоящего приказа направление его копии на официальное опубликование в периодических печатных изданиях и информационно-правовой системе "Эділет";

3) направление копии настоящего приказа в Республиканское государственное предприятие на праве хозяйственного ведения "Республиканский центр правовой информации" для включения в Эталонный контрольный банк нормативных правовых актов Республики Казахстан в бумажном и электронном виде в течение десяти календарных дней со дня государственной регистрации;

4) размещение настоящего приказа на интернет-ресурсе Комитета по статистике Министерства национальной экономики Республики Казахстан.

3. Управлению статистических регистров и классификаций Комитета по статистике Министерства национальной экономики Республики Казахстан довести настоящий приказ до структурных подразделений и территориальных органов Комитета по статистике Министерства национальной экономики Республики Казахстан для руководства и использования в работе.

4. Контроль за исполнением настоящего приказа возложить на курирующего заместителя Председателя Комитета по статистике Министерства национальной экономики Республики Казахстан (Батанов А.С.).

5. Настоящий приказ вводится в действие по истечении десяти календарных дней после дня его первого официального опубликования.

*Председатель Комитета по
Статистике Министерства
Национальной экономики
Республики Казахстан*

Н. Айдапкелов

Утверждена
приказом Председателя
Комитета по статистике
Министерства национальной
экономики Республики Казахстан
от 11 ноября 2016 года № 266

Методика по проведению выборочного обследования домашних хозяйств

Глава 1. Общие положения

1. Методика по проведению выборочного обследования домашних хозяйств (далее - Методика) относится к статистической методологии, формируемой в соответствии с международными стандартами и утверждаемой в соответствии с Законом Республики Казахстан от 19 марта 2010 года "О государственной статистике" (далее - Закон).

2. Методика устанавливает основные аспекты и методы анализа выборочной и генеральной совокупности домашних хозяйств и предназначена для применения структурными подразделениями Комитета по статистике Министерства национальной экономики Республики Казахстан (далее - Комитет).

3. Для проведения выборочных обследований отбирается часть домашних хозяйств, и в рамках этого контингента проводится наблюдение или собираются данные. Полученные результаты экстраполируются (распространяются) на все население в целом.

4. Основными преимуществами применения выборочного метода в современной статистике являются:

- 1) сокращение сроков проведения статистических наблюдений (обследований);
- 2) уменьшение информационной нагрузки на респондентов;
- 3) значительная экономия затрат труда, материальных и финансовых ресурсов при проведении обследования;
- 4) значительно ускоренного получения результатов исследования по сравнению со сплошным обследованием.

5. В настоящей Методике используются понятия в значениях определенных в Законе, а также следующие определения:

1) **панельный метод наблюдения** - способ сбора информации, при котором в течении относительно длительного времени периодически опрашивается определенная группа единиц анализа, причем предмет исследования остается постоянным;

2) генеральная совокупность - полная группа всех единиц анализа, чьи характеристики подлежат оценке;

3) репрезентативность - соответствие характеристик выборки характеристикам популяции или генеральной совокупности;

4) математическое ожидание - среднее значение отдельной характеристики во всех возможных выборках, также средневзвешенное значение всех возможных результатов с весом вероятностей, отражающих возможность появления в каждом результате;

5) параметр - величина, рассчитанная из всех значений в наборе генеральной совокупности, то есть описательное измерение генеральной совокупности;

6) страта - деление на специальные слои единиц (респондентов) обладающих одинаковыми или схожими показателями;

7) план выборки - набор спецификаций, определяющих генеральную совокупность и единицы выборки, а также степени вероятности возможных выборок;

8) выборочная совокупность (выборка) - множество случаев (испытуемых, объектов, событий, образцов), с помощью определенной процедуры выбранных из генеральной совокупности для участия в исследовании;

9) размер выборки - общее число единиц наблюдения в выборочной совокупности.

Глава 2. Планирование и процесс формирования выборки

6. При планировании обследования необходимо определить охватываемые им географические районы и исследуемое население.

7. При определении статистической совокупности необходимо выявить группу населения, из которой формируется выборка. Из основы выборки удаляются отдаленные районы с небольшим количеством домохозяйств или жителей, поскольку их охват связан со слишком высокими расходами. Они представляют собой лишь небольшую долю населения, их влияние на показатели численности населения весьма незначительны. В отчете по результатам такого обследования четко указан факт исключения этих районов.

8. Процесс формирования выборки для проведения обследования состоит из нескольких этапов:

определение генеральной совокупности;

создание основы выборки;

выбор между вероятностным и невероятностным методами отбора;

определение плана выборки;

определение объема выборки;

непосредственное формирование выборки согласно плану.

Глава 3. Определение генеральной совокупности и основа выборки

9. Данные переписи населения или информационная система статистический регистр жилищного фонда (далее - ИС СРЖФ) является основным источником для формирования выборочной совокупности для обследования домашних хозяйств в Республике Казахстан. Данные переписи населения служат средством предоставления информации о численности, составе и географическом распределении населения в дополнение к социально-экономическим и демографическим характеристикам. В рамках переписи населения ведется сбор информации по каждому отдельному лицу в домашнем хозяйстве и по каждому комплексу жилых помещений на всей территории. Во избежание случаев неполучения данных от респондентов для формирования выборки используется ИС СРЖФ. ИС СРЖФ создан с целью формирования и накопления данных по жилым домам и жилым помещениям для статистики жилищного фонда и формирования выборок для обследования домашних хозяйств.

10. Единицами учета в ИС СРЖФ являются все жилые дома и жилые помещения (квартиры), находящиеся на территории Республики Казахстан.

К ним относятся:

- жилое помещение (квартира);
- одноквартирный (индивидуальный) дом;
- двухквартирный дом;
- трех и более квартирный дом.

Каждый дом и квартира имеет идентификационный номер (далее - ID).

Кроме того, в ИС СРЖФ содержатся следующие данные: ID квартиры, классификатор административно территориальных объектов (КАТО), улица, номер дома, номер квартиры, общая площадь, жилая площадь. Актуализация данных, содержащихся в ИС СРЖФ, производится ежедневно.

Глава 4. Стратегия и методы формирования выборки

11. Для отбора элементов из генеральной совокупности используются следующие методы вероятностного отбора:

- простая случайная выборка (далее - ПСВ);
- систематическая случайная выборка (пошаговая выборка);
- выборка с вероятностью пропорционально размеру (далее - ВПР).

12. Простой случайный отбор (выборка) обеспечивает равную вероятность быть отобранным для каждого элемента генеральной совокупности. Встречаются следующие разновидности данного метода:

- повторный случайный отбор;

бесповторный случайный отбор.

13. Бесповторный случайный отбор дает более точные результаты выборочного наблюдения по сравнению с повторным, так как при одном и том же объеме выборки наблюдение охватывает больше единиц генеральной совокупности. В случаях, когда бесповторный отбор провести нельзя, используется повторная выборка.

14. Сущность систематической случайной выборки заключается в отборе из основы элемента, начиная с первого элемента, который отбирается случайно.

Например, при формировании систематической выборки с размером 500 элементов из генеральной совокупности 15 000 сотрудников организации.

Вначале определяется случайный старт, затем шаг отбора. ($15\ 000 / 500 = 30$, шаг отбора равен 30).

15. Метод отбора ВПР позволяет повысить точность оценивания, если используемая для определения вероятностей вспомогательная переменная размера приблизительно пропорциональна изучаемым признакам. При использовании метода ВПР существует больше вероятности, что в выборку попадут единицы с крупными признаками. Метод отбора ВПР часто применяется в обследованиях домохозяйств при отборе территориальных единиц, при этом вероятность включения элементов в выборку пропорциональна численности проживающего населения в территориальных единицах отбора.

Параграф 1. Стратифицированная выборка

16. При планировании обследования домашних хозяйств, широко применяемым методом, является стратификация для обследования совокупности до формирования выборки. Она служит для целей классификации совокупности в подсовкупностях на основе дополнительной информации, которая известна в отношении генеральной совокупности. Например, территориальные признаки или половозрастные категории, тип местности, количество проживающих, тип или вид строения, здания. Основным принципом формирования страт (расслоение) является - разнородность между стратами и однородность внутри страт. Городские и сельские районы формируются в качестве двух отдельных страт для обследования домашних хозяйств. Городское и сельское население отличаются друг от друга по многим аспектам (вид занятости, источник и размер дохода, средний размер домохозяйства, уровень рождаемости) в то время как лица, относящиеся к одной из этих подгрупп, обладают аналогичными характеристиками. Вероятность отбора при стратифицированной выборке с применением бесповторного случайного отбора рассчитывается по следующей формуле:

где:

n_i

- размер выборочной совокупности в страте;

- размер генеральной совокупности в страте.

17. Преимуществами стратифицированной выборки является:

1) уменьшение вариации (разброс переменных) в выборке;

2) обеспечение пропорциональной представленности отобранных единиц в каждом слое;

3) увеличение в итоговой выборке количества элементов из подсовокупностей повысит надежность статистического анализа.

Параграф 2. Кластерная (гнездовая) выборка

18. "Гнездовая выборка" ведется для обозначения планов выборки, в которую входят все члены группы. Сами группы обозначаются как кластеры или гнезда. Формирование кластерной выборки снижает надежность выборки в связи с высокой вероятностью того, что лица, проживающие в одном кластере, склонны к однородности или к обладанию более или менее аналогичными характеристиками (чем больше однородность в выборке, тем ниже ее надежность). Во избежание степени однородности кластеров увеличивается количество кластеров. Например, по обследованию уровня жизни населения используется метод кластерной выборки. Выборочная совокупность составляет 12 000 домашних хозяйств и формируется 400 кластеров по 30 домохозяйств в каждом, чем 200 кластеров по 60 домохозяйств в каждом.

19. В первом случае дизайн эффект выборки (далее -) значительно меньше. Величина снижается, если из всех домохозяйств одного кластера домохозяйства отбираются случайным методом, а не из территориально прилегающих друг к другу подсегментов.

где:

-дисперсия кластера или страты;

-дисперсия при простой случайной выборке.

Этот расчет дает оценку "полного" показателя , включая эффекты стратификации, а также вариантность размеров кластеров, связанного с гнездовой группировкой. Если больше 1, то это показывает, что план выборки менее эффективен (дает большую ошибку для выборки того же объема), чем простая случайная выборка. При этом компенсируется в плане выборки путем соразмерного увеличения размера выборки. меньше 1 свидетельствует об обратном.

Параграф 3. Двухэтапная выборка

20. Для обследований домашних хозяйств одним из часто используемых методов является двухэтапная выборка.

Целью такого двухступенчатого подхода является экономия затрат за счет проведения краткого проверочного опроса на этапе формирования первоначальной крупной выборки. Например, на первом этапе проводится отбор населенных пунктов, а на втором этапе отбор домохозяйств.

21. Для расчета общей вероятности отбора (P) в выборку домохозяйства для кластерной, поэтапной и двухэтапной выборки используется стандартное соотношение с учетом этапных характеров формирования выборки. Ниже приводится формула расчета вероятности отбора P для двухэтапной выборки:

где:

- вероятность включения первичных единиц выборки (далее - ПЕВ) в выборку;
- вероятность включения домохозяйства в выборку на втором этапе вычисляется по формуле.

Вероятность включения ПЕВ в выборку (π) на первом этапе вычисляется по формуле :

где:

- число ПЕВ, подлежащих отбору в g -ом отдельно по городскому и сельскому населению (h);
- количество домохозяйств в i -ой ПЕВ, включенной в выборку на первом этапе (в g -ом отдельно по городскому и сельскому населению (h));
- общее количество домохозяйств по всей совокупности ПЕВ (в g -ом отдельно по городскому и сельскому населению (h)).
- вероятность включения домохозяйства в выборку на втором этапе вычисляется по формуле:

,

где:

- количество домохозяйств, подлежащих отбору в рамках i -ой ПЕВ (в g -ом отдельно по городскому и сельскому населению (h));
- количество домохозяйств в i -ой ПЕВ, включенной в выборку на первом этапе (в g -ом отдельно по городскому и сельскому населению (h)).

Глава 5. Определение и распределение размера выборки

22. Для определения оптимального размера выборки, необходимого для оценки генеральной совокупности с заданной точностью, используется следующая формула:

где:

t - аргумент функции Лапласа ($t = 1,96$ для 95% уровня доверительности);

N - объем генеральной совокупности;

- дисперсия;

D - предельная ошибка выборки.

23. Для определения размера выборки оцениваются следующие параметры генеральной совокупности:

1) Средний арифметический признак (например, доходы и расходы домашних хозяйств, количество проживающих в домашних хозяйствах), вычисляется для всех единиц генеральной совокупности и носит название генеральной средней (\bar{X}) и рассчитывается по следующей формуле.

где:

N_i

- количество элементов генеральной совокупности i -страте;

- сумма показателя i -страты.

2) Дисперсия генеральной совокупности определяется как среднее значение квадратов отклонений всех отдельных наблюдений от их среднего значения.

Дисперсия генеральной совокупности вычисляется по следующей формуле:

Квадратный корень из дисперсии называется стандартным отклонением или среднеквадратичным отклонением и рассчитывается по формуле:

3) Если ошибка выражается как стандартная ошибка (RSE), то для определения размера выборки используется следующая формула:

где:

RSE - относительная стандартная ошибка выборки.

Если не принимать во внимание поправку для конечной совокупности, формула для определения размера выборки будет выглядеть следующим образом:

24. После определения размера выборки следует распределять выборку по стратам, если это стратифицированная выборка или по кластерам в случае кластерной выборки. Распределение выборки производится одинаковым размером выборки в каждой страте (равномерное распределение), или распределяется другими способами. Для определения распределения выборки по различным стратам имеется два важных критерия, которые влияют на способ определения размера выборки в стратах:

Первый критерий - удобство: выбирается способ пропорционального распределения, при котором размер выборки в i -ой страте рассчитывается по формуле:

где:

n_i - размер выборочной совокупности i - страты;

$i = 1, 2, \dots, h$;

N_i - количество домашних хозяйств в i -ой страте, при этом $i = 1, 2, \dots, h$.

Второй критерий - точность: выбирается способ оптимального распределения, дающий наименьшую среднюю квадратическую ошибку (стандартная ошибка) выборки.

25. Там, где затраты на производство выборки из различных страт одинаковы, оптимальная формула распределения называется распределением Неймана. В этом случае размер выборки в i -ой страте определяется по формуле:

где:

h - число страт в совокупности;

$i = 1, 2, \dots, h$;

S_i - стандартное отклонение i -ой страты.

Использование расслоения с оптимальным распределением (определение размера выборки в i -ой страте) возможно только при наличии стандартного отклонения.

Глава 6. Ротация и наложения выборки

26. При панельном выборочном обследовании, в целях недопущения эффекта усталости домохозяйств от участия в обследовании и для целей сглаживания скачков периодических данных обследования, периодически производится ротация (замена) домохозяйств, попавших в выборочную совокупность. Ротация домашних хозяйств, производится 1 раз в год. Например, если будет решено, что размер ротации будет равен 1/3 от общего числа обследуемых домашних хозяйств, то есть, ежегодно 1/3 часть домашних хозяйств будет выводиться из выборки и заменяться на другие, в этом случае, по истечении 3-х лет выборочная совокупность полностью обновится. При этом замена домохозяйств производится тем же методом, который был использован изначально для формирования выборки. При этом замена домашних хозяйств производится на те домашние хозяйства, которые относятся к той же страте, сегменту или кластеру.

Глава 7. Компенсация на неполучение ответов

27. В обследованиях различают два вида отсутствия ответов респондентов - полные, когда информация об объекте наблюдения отсутствует полностью, и частичные, когда отсутствуют ответы на отдельные вопросы программы обследования.

28. Существуют два типа причин полного отсутствия ответов: объективные и субъективные причины.

К объективным причинам отсутствия ответов относятся: если все члены домашнего хозяйства не подходят к целевой группе обследования (по половозрастной категории), дом разрушен, пустая квартира (дом), объединение квартир, иная причина (изменение назначения помещения), дом (адрес) не найден и другие причины.

К субъективным причинам отсутствия ответов относятся: отказ домашнего хозяйства от участия в обследовании.

29. При наличии объективных причин отсутствия ответов в ходе проведения обследования домашних хозяйств, составляются резервные списки. Резервная выборочная совокупность составляется специалистами по формированию выборочной совокупности для возможной замены домашних хозяйств по объективным причинам отсутствия ответов. Резервная выборочная совокупность составляется тем же методом, которым составлена основная выборочная совокупность по конкретному обследованию. Замена допускается только из резервного списка на то домохозяйство, которое конкретно предназначено служить субститутом или заменой неответившего домохозяйства по объективным причинам. Это делается для исключения замены на "удобное" домохозяйство, при которой увеличивается погрешность выборки.

В случаях неполучения ответов в обследованиях домашних хозяйств не предусматривается замена для компенсации, если не существует надлежащего обоснования для ее использования в конкретных целях.

30. Для корректировки отсутствия ответов по субъективным причинам применяется корректировка весовых значений и производится взвешивание. Взвешивание результатов выборочного обследования производится путем присвоения соответствующего веса каждой отдельной единице наблюдения – персоне.

Взвешивание данных в обследованиях домашних хозяйств включает две операции: расчет базисных весов и внесение поправок (корректировка весов) на неполучение ответов.

31. Расчет базисных весов начинается после определения вероятностей отбора единиц выборки. Вероятность отбора единицы выборки зависит от плана выборки, используемого для отбора данной единицы. Базисный вес единицы выборки является обратной величиной вероятности ее отбора для включения в выборку. В математическом выражении, если единица включается в выборку с вероятностью p , тогда ее базисный вес, обозначаемый как (коэффициент распространения), рассчитывается по формуле:

32. Проблема неполучения ответов от единицы выборки по субъективным причинам при обследованиях домашних хозяйств, решается с помощью корректировки весов выборки. Расчет скорректированного веса в случае неполучения ответов для i -ой единицы выборки рассчитывается с помощью следующего уравнения:

,

где:

- это количество фактически отчитавшихся.

Расчет окончательного скорректированного веса в случае неполучения ответов для i -ой единицы выборки рассчитывается с помощью следующего уравнения:

,

где:

- это первоначальный базисный вес;

- скорректированный вес в случае неполучения ответов.

Первоначальные веса корректируются для компенсации в случаях неполучения ответов.

33. Процесс заполнения частично пропущенных значений называют импутацией. Различают два основных подхода к проведению импутации:

методы, основанные на использовании данных донора;

методы вычисления оценок отсутствующих значений.

34. Метод донора основан на использовании данных ответов подобного респондента. В методах донора импутируемое значение для записи-реципиента берется из другой записи донора, выбираемой по определенным правилам. Есть два метода донорской импутации:

1) Случайный выбор донора. Случайный выбор записей, используемых в качестве доноров, обычно осуществляется в некоторых специально определенных классах. Классы представляют собой достаточно однородные по заданным параметрам группы наблюдений.

2) Метод ближайшего соседа. Импутация методом ближайшего соседа предполагает выбор такой записи из числа донорских, которая имеет наименьшее расстояние от импутируемой записи.

3) Методы вычисления оценок отсутствующих значений заключается в подстановке вместо отсутствующего среднего значения или подстановке исторических данных предыдущих обследований.

Глава 8. Оценка параметров выборочной совокупности по результатам обследований

35. По результатам обследований осуществляется оценка параметров выборочной совокупности. Для оценки достоверности результатов выборки вычисляются некоторые статистические характеристики.

Оценка стандартной ошибки выборки. Возможные расхождения между характеристиками выборочной и генеральной совокупности измеряются стандартной ошибкой (средней ошибкой) выборки. Стандартная ошибка выборки определяется по следующей формуле:

где:

- стандартная ошибка;
- генеральная дисперсия;
- объем выборочной совокупности.

36. Стандартная ошибка выборки показывает абсолютные значения ошибки. Для определения оцениваемого значения в долях используется относительная стандартная ошибка (коэффициент вариации). Этот коэффициент выражается в процентах и рассчитывается по формуле:

37. Относительная стандартная ошибка рассчитывается после проведения выборочного обследования. Чем больше значение относительной стандартной ошибки, тем относительно больший разброс и меньшая выравненность исследуемых значений. Если относительная стандартная ошибка по каждой страте меньше 10%, то изменчивость вариационного ряда считается незначительной, от 10% до 20% относится к средней, больше 20% и меньше 33% к значительной, более 33% выборка является нерепрезентативной и принимается решение об увеличении объема выборки. Для уменьшения относительной стандартной ошибки выборки в два раза, размер выборки увеличивается в четыре раза.

38. Для установления предельной ошибки выборки используется стандартная ошибка (σ) выборки. Предельная ошибка выборки по стратам t . Предельную ошибку используют для расчета доверительного интервала. Доверительный интервал показывает, в каком диапазоне расположатся результаты выборочных наблюдений.

39. Коэффициент t определяется задаваемой исследователем вероятностью P ($0 \leq P \leq 1$). Для значений P , приближающихся к единице, практически исключается возможность генеральной средней отличаться от вычисленной выборочной средней больше, чем на D , гарантируемую заданным уровнем доверительности вероятности P . При этом, чем выше уровень доверительности (используются, например, значения 0,90; 0,95; 0,99 и др.), тем выше коэффициент t , а, следовательно, и значение предельной ошибки D .

$t = 1,28$ для 80% уровня доверительности;

$t = 1,64$ для 90% уровня доверительности;

$t = 1,96$ для 95% уровня доверительности;

$t = 2,58$ для 99% уровня доверительности;

Как и сама выборочная характеристика, ошибка выборки является случайной величиной. Теорема Ляпунова, указывает на вероятность того, что ошибка выборки не превысит некоторую заданную величину, то есть что - .

Доверительный интервал генеральной средней определяется на основе неравенств - , из которых следует, что $x - x +$.

Глава 9. Распространение результатов выборки на генеральную совокупность

40. Конечной целью выборочного наблюдения является характеристика генеральной совокупности на основе данных, полученных по выборке. Распространение данных выборочного наблюдения на генеральную совокупность производится по следующей формуле:

или:

- сумма показателя выборочной совокупности $\sum_{i=1}^n x_i$;
- среднее значение показателя выборочной совокупности \bar{x} i-страте.