

Утверждена
приказом Председателя
Комитета по статистике
Министерства национальной
экономики Республики Казахстан
от «30» июня 2016 года № 132

Методика построения выборки домашних хозяйств по обследованию занятости населения

Глава 1. Общие положения

1. Методика построения выборки домашних хозяйств по обследованию занятости населения (далее – Методика) относится к статистической методологии, формируемой в соответствии с Законом Республики Казахстан «О государственной статистике».

2. Настоящую Методику применяет Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан для формирования выборки домашних хозяйств по обследованию занятости населения.

3. Целью настоящей Методики является построение выборки домашних хозяйств для выборочного обследования занятости населения (далее – ВОЗН). Методика определяет основные аспекты и методы построения выборки с учетом районного уровня, включая описание этапов формирования, стратификации и расчета выборочных весов.

4. В Методике используются понятия в значениях, определенных в Законах Республики Казахстан «О государственной статистике», «О занятости населения», а также определения принятые Международными Конференциями Статистиков Труда (далее – МКСТ) и рекомендации Международной Организации Труда (далее – МОТ).

Глава 2. Генеральная совокупность (основа выборки)

5. В качестве генеральной совокупности для формирования выборочной совокупности домашних хозяйств ВОЗН с учетом районного уровня является информационная система «Статистический регистр жилищного фонда», компонент интегрированной информационной системы «е-Статистика» (далее – СРЖФ)

6. Указанная база данных обладает рядом преимуществ:

наличие готовой рамки выборки;

наличие постоянно актуализируемой базы данных;

наличие информации о домашних хозяйствах в территориальном разрезе.

7. Единицами учета СРЖФ являются все жилые дома и жилые помещения, находящиеся на территории Казахстана. К ним относятся

квартиры, многоквартирные (индивидуальные) дома, двухквартирные дома, трех и более квартирные дома.

8. База данных СРЖФ состоит из двух таблиц, дом – основная таблица, содержит информацию в целом по жилому дому, квартира – дополнительная таблица, содержит информацию о каждой квартире жилого дома.

9. Особенности формирования генеральной совокупности:

1) в генеральную совокупность включаются все типы жилых помещений, за исключением общих коммунальных квартир, общежитий, медико-социальных учреждений (организаций) для престарелых, лиц инвалидность, детей с инвалидностью в условиях стационара, детских домов, тюрем, гостиниц, религиозных общин и других аналогичных жилых помещений;

2) в целях снижения нагрузки на интервьюеров при сборе информации из генеральной совокупности исключаются населенные пункты с численностью проживающих менее 100 домашних хозяйств.

Глава 3. Дизайн выборки

10. ВОЗН с учетом районного уровня проводится во всех регионах Казахстана на основе выборочного метода наблюдений с последующим распространением итогов на всю численность населения страны обследуемого возраста (15 лет и старше).

11. Модель организации ВОЗН с учетом районного уровня разработана на основе выборки с наложением, формируется один раз в год и в целях снижения нагрузки на интервьюеров равномерно распределена на двенадцать месяцев.

12. Домашнее хозяйство, попавшее в выборку текущего года, имеет равную вероятность быть отобранным снова при формировании выборки на новый год. Наложение осуществляется в размере части ежегодно рассчитываемой от общего количества домашних хозяйств.

13. Для целей сглаживания скачков ежемесячных данных обследования занятости, предусмотрено использование ежемесячного наложения ежегодно рассчитываемой части домашних хозяйств. Ежемесячно, начиная с февраля, ежегодно рассчитываемая часть домашних хозяйств переходят из предыдущего месяца на следующий месяц (обследуются два месяца подряд), а остальные – новые каждый месяц.

14. Ежемесячное формирование каталога для наложения осуществляется на территориальном уровне автоматически в программном комплексе на основе каталога введенных отчетов за отчетный месяц и заданного количества домашних хозяйств Z (рассчитывается Департаментом статистических регистров и классификаций при формировании выборочной совокупности на текущий год) путем использования шага N_n , где n – месяц в котором используются выбранные домохозяйства. Шаг $N_n = N_{otc}/Z$, где Z – количество домашних хозяйств определенное для наложения по данному региону,

постоянное для всех месяцев отчетного года, $N_{отс}$ – количество домашних хозяйств, фактически отчитавшихся в текущем месяце.

15. Данная модель, обеспечивает:

1) репрезентативные годовые итоги на основе объединения всех двенадцати месячных выборок:

по Республике Казахстан;

по областям;

по районам.

2) репрезентативные квартальные итоги на основе объединения трех месячных выборок, входящих в квартал:

по Республике Казахстан;

по областям.

3) репрезентативные месячные итоги на основе месячной выборки:

по Республике Казахстан.

4) сглаживание скачков ежемесячных данных.

16. Первичной единицей обследования являются жилые дома (индивидуальные дома, многоквартирные дома), в обследовании участвуют все квартиры в доме, попавшие в выборку.

Вторичной единицей обследования являются домашние хозяйства (лица в возрасте от 15 лет и старше, проживающие в них).

Выборочная совокупность домашних хозяйств формируется методом двухступенчатой стратифицированной кластерной выборки.

Формирование выборочной совокупности осуществляется централизованно на республиканском уровне в разрезе регионов Казахстана.

Глава 4. Стратификация генеральной совокупности

17. Стратификация является действенным приемом повышения эффективности выборочного метода. Преимуществами стратификации являются:

большая точность, по сравнению с другими вариантами отбора для тех же данных;

большая точность, при использовании меньшей по размеру выборки, что экономит средства;

обеспечение представительности выборки.

18. Необходимость стратификации генеральной совокупности обусловлена неоднородностью единиц наблюдения по характеристикам.

19. При формировании выборки для ВОЗН с учетом районного уровня используется следующая процедура стратификации:

генеральная совокупность стратифицируется по районам, по типу местности (город, село), по классам размерности зданий (1 квартирные дома, 2–20 квартирные дома, 21–80 квартирные дома, более 80 квартирные дома);

определяется объем генеральной совокупности в каждой страте путем суммирования количества домов.

20. Непосредственный отбор домов разных типов осуществляется с помощью специализированного программного обеспечения с использованием метода случайного отбора. Использование данного программного продукта сокращает объем работ и дает гарантированный качественный результат.

Если в отобранных адресах фактически проживают несколько домашних хозяйств, то случайным методом выбирается одно из них.

После отбора формируются списки домашних хозяйств, попавших в выборку, для каждой области.

Глава 5. Определение размера выборки

21. Размером выборки является общее число единиц наблюдения в выборочной совокупности. Размер выборки зависит от размера относительной ошибки выборки, которая с определенной вероятностью обеспечивает заданную точность результатов наблюдения.

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

$$n = \frac{k^2 * N * \sigma^2}{k^2 * \sigma^2 + E^2 * N} \quad (1)$$

где,

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

N – объем генеральной совокупности;

σ^2 – дисперсия;

E – абсолютная ошибка.

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

Дисперсия генеральной совокупности:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1} \quad (2)$$

Если ошибка выражается как стандартная ошибка (μ), можно воспользоваться следующей формулой:

$$n = \frac{k^2 * N * (RSE)^2}{k^2 * (RSE)^2 + N * \mu^2} \quad (3)$$

где, RSE – относительная стандартная ошибка выборки.

Если не принимать во внимание поправку для конечной совокупности формула будет выглядеть следующим образом:

$$n = \frac{k^2 (RSE)^2}{\mu^2} \quad (4)$$

24. При определении объема выборочной совокупности на текущий год для ВОЗН используется расчет относительной ошибки выборки результатов ВОЗН за предыдущий год. Относительная ошибка выборки обратно пропорциональна объему выборки, при увеличении размера выборки в четыре раза, ошибки уменьшаются вдвое.

25. Основным требованием к результатам ежемесячного обследования является получение репрезентативных данных на уровне района со стандартной ошибкой выборки по показателю «уровень безработицы» на уровне 5–7%.

С учетом требований рассчитан и принят оптимальный размер выборки – 5%.

Глава 6. Компенсация неполученных ответов

26. Недополучение ответов от отобранных домашних хозяйств является серьезным недостатком, искажающим результаты обследования. В этих условиях возникает необходимость компенсации недостающих данных. Данная процедура призвана осуществить досчет показателей, компенсировать недополучение данных выборочного обследования.

27. В ходе обследования имеет место случаи, когда не удается опросить домохозяйства. Существуют два типа причин объективные и субъективные.

Объективные причины отказа:

все члены домашнего хозяйства в возрасте 73 года и старше;

дом разрушен;

уехали, пустая квартира (дом);

объединение квартир;

дом под снос;

дом (адрес) не найден;

иная причина (изменение назначения помещения).

Субъективные причины отказа:

отказ домашнего хозяйства от участия в обследовании.

28. Замена адресов жилых помещений в случае отсутствия членов домашнего хозяйства или отказа участвовать в обследовании не производится.

Интервьюер опрашивает членов домашних хозяйств, проживающих по адресам, только согласно спискам, представленным в выборке.

28-1. 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.

Замена адресов жилых помещений допускается только из резервного списка на то домохозяйство, которое конкретно предназначено служить субститутом или заменой не ответившего домохозяйства по объективным причинам. Это делается для исключения замены на "удобное" домохозяйство, при которой увеличивается погрешность выборки.

29. При обработке случаев не предоставления данных, неполученные ответы по объективным причинам – не компенсируются. Компенсации подлежат только случаи неполученных ответов по субъективным причинам.

Для этих целей обработки применяется метод перевзвешивания, заключающийся в корректировке выборочных весов.

30. В целях получения данных, распространенных на генеральную совокупность, производится статистическое взвешивание итогов обследования.

Взвешивание результатов выборочного обследования производится путем присвоения соответствующего веса каждой отдельной единице наблюдения – персоне.

31. Окончательный индивидуальный вес K представляет собой произведение базового веса F и коэффициентов компенсации и экстраполяции.

Индивидуальный вес, рассчитанный для каждого респондента, вводится в базу индивидуальных данных в качестве множителя распространения и используется при формировании распространенных данных на всю численность населения обследуемого возраста, пола и типа местности по любым показателям программы обследования.

32. Веса для показателей занятости населения рассчитываются ежемесячно. Для расчета весов используются данные СРЖФ о распределении обследуемых домашних хозяйств отдельно по городскому и сельскому населению в региональном разрезе.

33. При расчете индивидуальных весов применяется метод итеративного взвешивания выборки по отношению к общей численности населения в периоде, наиболее приближенному к отчетному периоду (критической неделе обследования).

Процедура заключается в сопоставлении выборочной совокупности, распределенной на группы с учетом половозрастной и региональной характеристик, со всей совокупностью населения, распределенной по этим же характеристикам.

34. При обработке результатов ежемесячного опроса расчет индивидуальных весов взвешивания (коэффициентов экстраполяции) производится последовательно в несколько этапов.

35. В целях выравнивания неизбежных при выборках случайных, а также неслучайных систематических ошибок:

проводится выравнивание известных случаев несостоявшихся опросов – компенсация;

проводится распространение на генеральную совокупность до показателей СРЖФ – адаптация или экстраполяция.

Для этого вычисляются соответствующие факторы, фактор компенсации и фактор экстраполяции.

Параграф 1. Компенсация

36. С помощью данной процедуры осуществляется дорасчет показателей по домашним хозяйствам, с субъективными причинами отказа от обследования, и компенсируются недополученные данные выборочного обследования.

37. Перед осуществлением этой процедуры суммируют количество домашних хозяйств, предполагаемых к обследованию (S_1), количество домашних хозяйств, фактически опрошенных (S_2) и количество неопрошенных домашних хозяйств с указанием причины (Sp).

Численность неопрошенных подразделяют на группы – по объективным причинам (Sp_1) и по субъективным причинам (Sp_2):

объективные причины связаны с невозможностью проведения обследования, вследствие разрушения (сноса) или объединения самого жилого помещения, смерти, отсутствия длительное время или выбытия на новое место подлежащего обследованию члена домашнего хозяйства и других непредвиденных обстоятельств;

к субъективным причинам относят отсутствие на момент обследования жильцов, отказ отдельного члена или всего домашнего хозяйства от опроса.

38. С помощью процедуры компенсации осуществляется дорасчет показателей по домашним хозяйствам, попадающим под субъективные причины отказа от обследования, компенсируются недополученные данные выборочного обследования. Для этого рассчитывается промежуточный компенсирующий фактор (вес) K_1

$$K_1 = 1 + Sp_2 / S_2 \quad (5)$$

где,

K_1 – компенсирующий фактор (коэффициент досчета);

Sp_2 – количество домашних хозяйств, не опрошенных по субъективным причинам отказа от опроса;

S_2 – количество фактически опрошенных домашних хозяйств.

39. Промежуточный компенсирующий фактор вычисляется с точностью до пяти десятичных знаков по каждому району отдельно и присваивается каждому респонденту данного района вне зависимости от его пола и возраста.

Далее производится корректировка базового веса (f) на компенсирующий фактор (K_1).

Скорректированный на базовый вес, компенсирующий фактор используют для определения численности, полученной в результате доведения числа домашних хозяйств до обследуемого (S_1)

$$S_1 = S_2 * K_1 \quad (6)$$

40. Для компенсации случаев полного неполучения данных применяется схема корректировки весов, путем задания больших весов для всех ответивших домохозяйств в данном населенном пункте. Веса всех домашних хозяйств, ответивших на вопросы в данном населенном пункте, увеличиваются на один и тот же коэффициент. Все не ответившие домашние хозяйства исключаются из выборки путем задания для каждого из них фактического веса равного нулю.

Параграф 2. Экстраполяция

41. Распространение выборочных данных обследования основано на присвоении соответствующего индивидуального веса каждой отдельной единице наблюдения – домашнему хозяйству.

Для этого осуществляется сопоставление данных обследования по выборочной совокупности (численности обследованных граждан), стратифицированной с учетом половозрастной и региональной характеристик, с генеральной совокупностью населения по данным текущих демографических расчетов, стратифицированной по этим же характеристикам.

42. В общем виде формула расчета фактора адаптации (веса) имеет вид

$$Vg = Wg * N / Ng \quad (7)$$

где,

Vg – вес по признаку g ;

Wg – доля населения в генеральной совокупности, с характеристикой g ;

N – общее число опрошенных;

Ng – число опрошенных, с характеристикой g .

43. Процедура взвешивания производится на основе данных о структуре населения по сложившейся совокупности, используемой в качестве генеральной, только внутри региональных слоев (страт) по полу и возрасту. Для каждого респондента рассчитывается система факторов адаптации (весов) по следующим признакам:

территориальная структура (район);

тип местности;

пол (мужчины и женщины);

6 возрастных групп (6–10 лет; 11–14 лет; 15–34 года; 35–54 года; 55–71 год; 72 года и выше).

44. Для расчета базового индивидуального веса используется следующая формула расчета фактора адаптации:

$$K_{2k} = S_k / S_{1k} \quad (8)$$

где,

K_2 – фактор адаптации;
 S – численность населения по генеральной совокупности;
 S_1 – численность населения фактически опрошенных домашних хозяйств;
 K – отличительный признак в зависимости от совокупной характеристики лица, для которого рассчитывается фактор.

45. Окончательный индивидуальный вес (или коэффициент экстраполяции) является произведением фактора компенсации и фактора адаптации.

$$K = K_1 * K_{2k}, \quad (9)$$

где,

K – индивидуальный вес (коэффициент экстраполяции);

K_1 – компенсирующий фактор (коэффициент досчета);

K_{2k} – фактор адаптации.

При этом, если фактор компенсации для всех членов заданного домашнего хозяйства имеет одно и тоже значение, то факторы адаптации, как соответственно и окончательный фактор экстраполяции (индивидуальный вес) – различен.

46. Рассчитанные индивидуальные веса в качестве дополнительных переменных записываются в базу индивидуальных данных по каждому опрошенному респонденту и используются при формировании сводных итогов по соответствующему месяцу, кварталу, году на республиканском, областном, районном уровнях.

Распространенные данные в среднем за текущий год получают путем объединения ежемесячных выборок за год, а распространенные квартальные данные – путем объединения ежемесячных выборок за квартал.

Глава 7. Ротация выборки

47. Ротация домашних хозяйств (обновление выборки) проводится ежегодно в размере 100%, ежегодно вся выборка заменяется новыми домашними хозяйствами.

Глава 8. Оценка точности показателей

48. В качестве показателей точности статистического оценивания, используются стандартная ошибка выборки и стандартная относительная ошибка выборки.

49. Стандартной ошибкой выборки является стандартное отклонение значения параметра выборки от выборочного среднего значения этого параметра.

Относительной стандартной ошибкой является отношение оцениваемой статистической величины к его среднему значению.

50. Точность обследования по районам рассчитывается раз в год по показателям численность занятых, численность безработных, численность

рабочей силы, численность лиц, не входящих в состав рабочей силы, доля рабочей силы в численности населения, уровень безработицы.

51. В качестве показателей точности статистического оценивания используются стандартная ошибка выборки и стандартная относительная ошибка выборки.

52. Стандартная ошибка выборки (μ) определяет возможные расхождения между характеристиками выборочной и генеральной совокупности. Значение стандартной ошибки выборки определяются по формуле:

$$\mu = \sqrt{\frac{\sigma^2}{n} (1 - n/N)} \quad (10)$$

где,

σ^2 – генеральная дисперсия;

n – объем выборочной совокупности;

N – объем генеральной совокупности.

53. По итогам ВОЗН статус респондента определяется из ответов респондентов на вопросы (например: да или нет). Переменные, определяющиеся выбором одной из двух альтернативных ответов, являются бинарной. По теории вероятностей для бинарной переменной рассчитывается дисперсия. Для расчета относительной стандартной ошибки ВОЗН, используется долевое значение ответов (да или нет). Определяется дисперсия доли, затем вычисляется стандартная относительная ошибка.

Стандартная ошибка выборочной доли (μ_p) определяется по формуле:

$$\mu_p = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} \quad (11)$$

где:

w - доля ответов (доля «успехов») в процентах;

n – объем выборочной совокупности;

N – объем генеральной совокупности.»;

дополнить пунктом 54 следующего содержания:

«54. Часто требуется рассмотреть не абсолютные значения стандартной ошибки, но ее значение в отношении оцениваемой статистической величины. Относительная стандартная ошибка (далее – ОСО) рассчитывается как отношение оцениваемой статистической величины к его среднему значению.

ОСО определяется по следующей формуле:

$$RSE = SE / \bar{x} \quad (12)$$

где,

RSE – относительная стандартная ошибка выборки;
SE – стандартная ошибка выборки;
 \bar{x} – среднее значение переменной, использованной для оценки величины относительной стандартной ошибки.

$$\bar{x} = \sum x_i / N_i \quad (13)$$

где,

N_i – количество элементов генеральной совокупности i -страте;

x_i – показатель i -страты.».