

Система выборки

В настоящем приложении описаны основные свойства системы выборки. В число этих свойств входят размер целевой выборки, распределение выборки, рамка выборки и формирование списков, выбор областей, этапы формирования выборки, стратификация и расчет выборочных весов.

Основной целью системы выборки для Мультииндикаторного кластерного обследования (МИКС) для Кыргызстана является получение статистически надежных оценок большинства показателей на национальном уровне для городских и сельских районов. Некоторые из показателей рассчитаны для областного уровня. Городские и сельские регионы в каждой из семи областей (Баткенская, Чуйская, Жалалабатская, Ысыккульская, Ошская, Нарынская, Таласская области) и город Бишкек, что составило 15 территориальных единиц, определены как области выборки.

При формировании выборки для обследования использовался метод многоэтапной стратифицированной кластерной выборки.

Размер и распределение выборки

Размер целевой выборки для МИКС в Кыргызстане определен в 5200 домохозяйств. В качестве ключевого показателя для расчета размера выборки принята распространенность задержки роста среди детей в возрасте 0–4 года. Для оценки необходимого размера выборки по этим показателям использована следующая формула:

$$n = \frac{[4(r)(1-r)(f)(1.1)]}{[(0.12r)^2(p) \cdot n_h]}$$

где

- n – необходимый размер выборки, выраженный как число домохозяйств
- 4 – коэффициент, обеспечивающий 95–процентной доверительный уровень
- r – прогнозируемая или ожидаемая распространенность (процент охвата) данного показателя
- $1,1$ – коэффициент, необходимый для увеличения размера выборки на 10 процентов для компенсации неполученных ответов

- f – сокращенное обозначение для $deff$ (design effect – дизайн-эффект)
- $0.12r$ – допустимый предел погрешности, приемлемый при 95–процентном доверительном уровне и определяемый как 12 процентов от r (относительная ошибка выборки для r)
- p – доля в генеральной совокупности, на которой основан показатель r
- n_h – средний размер домохозяйства.

При расчетах значение r (распространенность задержки роста) было принято равным 0.25 (25 процентам). Для $deff$ (дизайн-эффекта) принималось значение 1,25 исходя из оценок, полученных по предыдущим обследованиям, величина p (доля детей в возрасте 0–4 года в генеральной совокупности) принималась равной 9.74 процентам, а n_h (средний размер домохозяйства) принят равным 4.6 домохозяйств.

Количество домохозяйств, полученное в результате расчета, равно 2557, что соответствует размеру выборки, необходимой в каждом районе (сельском/городском), таким образом, всего получится 5115 домохозяйств. По ряду соображений, в частности, с учетом бюджета обследования и времени, которое потребуется одной бригаде, проводящей обследование, для обработки одного кластера, общее число домохозяйств определен равным 5200, а число домохозяйств в кластере 13 домохозяйствам. Разделив общее число домохозяйств (5200) на число домохозяйств в кластере (13), получаем, что в каждом регионе следует сформировать 400 кластеров.

Было необходимо обеспечить равное распределение всей выборки между восемью регионами (7 областей и город Бишкек). В связи с этим в каждом регионе было сформировано 48 кластеров с дополнительным числом 16 для города Ош, что составило всего 400 кластеров. В каждом регионе кластеры (первичные единицы выборки) равномерно распределялись между городскими и сельскими областями в данном регионе. В таблице, приведенной ниже, дается распределение кластеров по областям выборки.

Таблица SD.1, Распределение кластеров (первичных единиц выборки) по областям выборки

Область	Население (по оценке 2005 г.)			Число кластеров		
	Всего	Городское	Сельское	Всего	Городское	Сельское
Баткенская	413722	105167	308555	48	24	24
Жалалабадская	947608	225361	722247	48	24	24
Ысыккульская	426406	122175	304231	48	24	24
Нарынская	265515	47375	218140	48	24	24

Ошская	1283918	338755	945163	64	40	24
Таласская	212098	35493	176605	48	24	24
Чуйская	752804	155473	597331	48	24	24
г. Бишкек	790731	790731	-	48	48	-
Всего	5092802	1820530	3272272	400	232	168

Рамка выборки и формирование кластеров

Для формирования кластеров использовалась рамка переписи 1999 года. Участки переписи принимались в качестве первичных единиц выборки (ПЕВ), при этом они выбирались из каждой области выборки с помощью систематических выборочных процедур с вероятностью, пропорциональной размеру (ВПР), на основе оценок размеров участков для переписи населения 1999 года. Первый этап формирования выборки завершился выбором необходимого числа участков переписи в каждой из 8 областей отдельно для городских и сельских районов.

Работы по формированию списков

Учитывая, что рамка выборки (результаты переписи 1999 года) устарела, до момента отбора домохозяйств необходимо было обновить их списки во всех выбранных участках переписи. Для этого были сформированы группы по составлению списков, которые посещали каждый участок переписи и переписывали все населенные домохозяйства. Формирование списков было осуществлено исполнительным агенством в августе-сентябре 2005 года. Для составления списков дополнительно была разработана анкета, в которую вносились записи о всех жилых единицах в первичных единицах выборки.

Мультииндикаторное кластерное обследование		Нацстаткомитет Кыргызской Республики	
		Номер пачки	
Наименование области _____		<div style="text-align: right; margin-bottom: 5px;">Код СОАТО</div> <div style="border: 1px solid black; width: 100%; height: 20px;"></div>	
Наименование района _____			
Наименование айылного кенеша _____			
Наименование населенного пункта _____			
Список жилых помещений на территории населенного пункта			
№ п/п	Для населенных пунктов с существующей адресной частью	Для населенных пунктов без адресной части	

Член f_h – выборочная доля в страте h – представляет собой произведение вероятностей отбора на каждом этапе в каждой области выборки:

$$f_h = P_{1h} * P_{2h} * P_{3h}$$

где P_{ih} – вероятность отбора данной выборочной единицы на этапе i для выборочной области h .

Оценки числа домохозяйств на каждом участке переписи, полученные до первого этапа отбора (отбора первичных единиц выборки), и уточненное число домохозяйств на каждом таком участке различны, поэтому для каждого участка переписи (кластера) рассчитывались индивидуальные выборочные доли домохозяйств. В связи с этим выборочные доли для домохозяйств на каждом участке переписи (кластере) включали вероятность отбора данного участка переписи в данной области выборки и вероятность отбора домохозяйства на данном участке переписи (кластере).

Второй составляющей, которую необходимо учитывать при расчете выборочных весов, является уровень неполученных ответов для домохозяйств и отдельных лиц, опрошенных в ходе обследования. Корректирующий коэффициент для неполученных ответов для домохозяйств равен обратной величине от

$$RR = \text{Число опрошенных домохозяйств} / \text{Число населенных домохозяйств, включенных в список}$$

После завершения работ на местах для каждой области выборки рассчитывались доли ответивших. Полученные величины использовались для корректировки выборочных весов, рассчитанных для каждого кластера.

Аналогичным образом, корректирующий коэффициент для учета неполученных ответов на индивидуальном уровне (женщины и дети в возрасте до 5 лет) определялся как величина, обратная:

$$RR = \text{Заполненные вопросники для женщин (детей до 5 лет)} / \text{Женщины (или дети до 5 лет), входящие в обследуемую категорию}$$

Число женщин и число детей в возрасте до 5 лет, входящих в обследуемую категорию, были получены из списков тех домохозяйств в Вопроснике для домохозяйств, где опросы уже были проведены.

Нескорректированные веса для домохозяйств рассчитывались путем умножения на полученные выше коэффициенты для каждого участка переписи. Затем эти веса нормировались, в частности, для того чтобы сделать

сумму опрошенных выборочных единиц равной общему размеру выборки на национальном уровне. Нормирование выполнялось с помощью умножения нескорректированных весов на отношение числа обследованных домохозяйств к общему нескорректированному взвешенному числу домохозяйств. Аналогичная нормировочная процедура применялась для получения нормированных весов для вопросников для женщин и о детях в возрасте до 5 лет. Скорректированные (нормированные) веса варьировались в 15 (из которых были отобраны 13) за исключением одного очень малого участка переписи (кластера) в пределах от 0.037 до 4.167.

Выборочные веса были определены для всех наборов данных, после чего на основе полученных весов проводился анализ с использованием взвешивания каждого домохозяйства, женщины или ребенка в возрасте до 5 лет с таким выборочным весом.