

## МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К ПРОВЕДЕНИЮ ВЫБОРОЧНЫХ ОБСЛЕДОВАНИЙ МИКРООРГАНИЗАЦИЙ

Н. Ч. Бокун

*В статье обосновывается методология проведения выборочных обследований микроорганизаций по видам деятельности, проблемы внедрения выборочных обследований предприятий в практику официальной статистики (на примере стран СНГ). Проанализированы возможности применения случайных выборок без расслоения, с простым, пропорциональным, оптимальным расслоением, механических, серийных выборок. Обобщены подходы к построению многомерных выборок: типизированные и множественные основы выборки, расслоение по комбинированному признаку, комбинированные методы. Рассмотрен механизм формирования выборочной совокупности микроорганизаций, включая алгоритм отбора предприятий, определение оптимального объема выборки по видам деятельности и регионам. Рекомендуется статистическое оценивание трех видов: HT-, GREG- и SYN- оценки. Предложено использовать комбинацию одномерных и многомерных методов, многовариантное статистическое взвешивание.*

*Ключевые слова:* выборка, взвешивание, оценивание, микроорганизации, многомерная выборка, оптимальное расслоение, расслоенный отбор.

*JEL:* C83, O12.

Выборочное наблюдение как метод статистического исследования обладает рядом преимуществ: относительно небольшие (по сравнению со сплошным обследованием) материальные, трудовые, стоимостные затраты, оперативность получения данных, достаточно высокая достоверность, широкая область применения. Одним из важных направлений использования выборки выступают обследования деятельности предприятий, в частности малого и среднего бизнеса.

В последние десятилетия в Республике Беларусь, как и в других странах СНГ, значительно возросло число малых организаций. В течение 2001–2013 гг. оно увеличилось в 3,2 раза, несколько снизилось в 2016 г. (на 3,3%), охватывая 19% занятых (при этом доля микроорганизаций составляет почти 90%<sup>1</sup>). Возникает необходимость детального исследования границ, структуры и объемов производства мелких экономических единиц. Сбор статистической информации по полному кругу организаций представляется технически сложным и дорогостоящим; усиливается актуальность выборочных методов.

Отечественная и зарубежная практика применения отраслевых выборок показала наличие

таких общеметодологических проблем, как локализация выборки, малые области статистического оценивания, региональные выборки, невыборочные ошибки, изменение структуры совокупности, адекватная экстраполяция и т. д. Возможные направления совершенствования теории выборки могут быть связаны с детализацией различных видов отбора, усложнением схем экстраполяции. В данной статье автор анализирует возможности организации репрезентативного мониторинга малого предпринимательства, одномерных и многомерных выборок и на этой основе предлагает методологию проведения выборочных обследований микроорганизаций, базирующуюся на применении комбинации различных видов отбора и многовариантном статистическом оценивании.

***Проблемы проведения выборочных обследований в сфере малого бизнеса.*** Изучение малого предпринимательства в странах СНГ исключительно методами выборочных обследований находится в начальной стадии; информацию о малом бизнесе получают, основываясь на сочетании сплошного наблюдения и выборки.

Бокун Наталья Чеславовна (nataliabokun@rambler.ru) – канд. экон. наук, доцент кафедры статистики Белорусского государственного экономического университета (г. Минск, Республика Беларусь).

<sup>1</sup> Республика Беларусь: статистический ежегодник. Минск, 2016. С. 258–259; Малое и среднее предпринимательство в Республике Беларусь. 2009–2013: стат. сб., Минск, 2014; Малое и среднее предпринимательство в Республике Беларусь: стат. сб., Минск, 2017.

Наиболее разработан и отлажен механизм наблюдения за деятельностью малых предприятий в России, где с 1996 г. проводятся ежеквартальные выборочные обследования малых предприятий. С 2003 г. используется режим сплошного годового обследования (сначала один раз в два года, затем один раз в пять лет) и годовых выборочных наблюдений. Источниками для основы выборки выступают: статистический регистр предприятий, база данных по государственной бухгалтерской отчетности и генеральная совокупность объектов статистического наблюдения. Используемый метод отбора - расслоенная случайная выборка; результат - оценки суммарных показателей с характеристиками их точности. При необходимости осуществляется дорасчет показателей на ненаблюдаемую часть исследуемой совокупности. Процесс формирования выборки и распространения ее результатов на исследуемую совокупность производится на региональном уровне с использованием программного продукта. Сводные итоги по России формируются на федеральном уровне. Кроме того, начиная с 2000 г. проводится выборочное обследование индивидуальных предпринимателей в розничной торговле с применением двух основ выборки, механического и случайного отбора [1].

На Украине используется полугодовая расслоенная выборка малых предприятий; в Казахстане - ежеквартальная выборка в сочетании с годовым обследованием. В Беларуси с 2004 по 2008 г. проводились ежеквартальные обследования малых предприятий. Выборки строились по областям в разрезе отраслей, применялось простое, пропорциональное и оптимальное случайное расслоение, с 2005-2006 гг. внедрялись методы одномерного и многомерного отбора [2, 3]. Но в 2008 г. квартальная отчетность малых предприятий была отменена, по этой причине выборочные обследования прекратились. Тем не менее в связи с ростом числа микроорганизаций, составивших в 2011 г. 84%, в 2013 г. - уже 87% от суммарного числа малых предприятий, осложнились процессы организации их годового сплошного наблюдения. От каждого мелкого хозяйствующего субъекта становится экономически нецелесообразно и практически невозможно требовать предоставления статистической отчетности. В результате в 2014 г. Национальный статистический комитет Республики Беларусь и кафедра статистики БГЭУ начали разрабатывать методологию и алгоритмы специализированной версии программного обе-

спечения годового выборочного обследования микроорганизаций.

Опыт проведения в Беларуси ежеквартальных обследований малых предприятий, анализ массивов показателей деятельности микроорганизаций за 2011-2013 гг., отчитывающихся по форме № 1-МП (микро), позволил выделить следующие проблемы, которые возникают при разработке методологического и программного обеспечения выборок:

1. Неответы респондентов. Совокупность микроорганизаций является чрезвычайно динамичной - постоянно происходят демографические процессы создания новых, ликвидации, изменения видов деятельности и масштабов предприятий. Так как база выборки строится на основе отчетности предыдущего года, то в выборку попадают не ответившие предприятия (ликвидированные, временно приостановившие деятельность, изменившие вид деятельности или не представившие отчет).

2. Нетипичные единицы, то есть наличие в основе выборки нетипичных единиц, включение (или исключение) которых в выборочную совокупность сильно влияет на итоговое значение получаемых оценок показателей и в конечном счете отражается на качестве результатов обследования. К нетипичным можно отнести единицы, имеющие экстремальные значения показателей или сложную структуру.

3. Выборки малого объема. Построение отраслевых выборок малых и микроорганизаций в разрезе областей и г. Минска в ряде случаев связано с разбиением исследуемой совокупности на мелкие подсовокупности и вызывает необходимость использовать малые выборки, в которых доли отбора становятся неприемлемо высокими (50-60%), а возможности обеспечения допустимой стандартной ошибки выборки проблематичными.

4. Раздробление выборок на группы. Стремление получить низкую допустимую ошибку выборки может привести к раздроблению небольшой по объему совокупности на значительное число групп (7-8), что, в свою очередь, ведет либо к сплошному обследованию, либо к высокой доле отбора в этих группах, а в результате - к общей высокой доле отбора.

5. Проблема компромисса, который возникает между требованием точности по различным признакам, представительством единиц наблюдения от различных групп по результатам расслоения

генеральной совокупности и ограничениями по объему выборок.

6. Экстраполяция. Проблема экстраполяции выборочных данных на генеральную совокупность сохраняется даже при построении одномерной расслоенной выборки с приемлемыми стандартной ошибкой и долей отбора. Веса организаций, в том числе групповые, позволяют достаточно точно экстраполировать значения признака, положенного в основу отбора, но искажают другие варьирующие показатели, число которых по форме № 1-МП (квартальная) составляло 30, по форме № 1-МП (микро) - свыше 80.

7. Проблемы программного обеспечения, связанные как со сложностью математического аппарата методов несплошного наблюдения, так и с необходимостью интегрирования программ, реализующих алгоритмы этих методов в общую систему сбора и обработки статистических данных.

8. При построении выборки, расслоенной по нескольким признакам, или многомерной, возникают такие специфические проблемы, как сложность оптимального способа многомерного отбора, сложность выбора ведущего признака, техническая невозможность построения многомерной выборки в рамках большой генеральной совокупности (свыше 350-400 единиц). В условиях неразработанности теории многомерной выборки возможности использования различных видов многомерного отбора в определенной мере ограничены: использование комбинационных таблиц ведет к формированию большого количества групп слабой заполненности, отбор по решетке возможен при небольшом числе признаков (2-4), многомерный (композиционный) признак во многом условен и не применим для качественных признаков. Отсутствуют общепринятые методы расчета множественных ошибок, оценивания, экстраполяции. Кроме того, следует учитывать, что если при одномерном отборе можно достичь минимальной ошибки по одному из признаков (до 1%), по другим она может составлять 50-100%, что недопустимо, то при многомерной выборке ошибки по всей совокупности показателей будут находиться в допустимых пределах (до 10%), но окажутся значительно выше, чем по признаку отбора в одномерной выборке (примерно 4-6%). Если число единиц в генеральной совокупности превышает 400 единиц, процесс кластерного анализа или построение комбинационных таблиц представляется чрезвычайно трудоемким, требует высоких затрат, что делает его неэффективным.

Решение проблем «неответов респондентов» и «нетипичных единиц совокупности» связано с изменением структуры базовой совокупности предприятий. В процессе построения выборочной совокупности основу выборки целесообразно разделить на три массива: 1) нетипичные предприятия, которые по значениям одного или нескольких показателей выделяются на фоне других единиц совокупности; 2) предприятия, по которым получены данные о признаках отбора; 3) предприятия, по которым данные о признаках отбора отсутствуют. Первый массив полностью включается в выборку; из второго массива осуществляется случайный или механический отбор; третий массив либо исключается из обследования, либо отбор единиц из него осуществляется по иному признаку. При наличии неответов, превышающих 20% единиц совокупности, можно осуществлять процедуру перевзвешивания. Новый коэффициент пространства равен отношению объема слоя в генеральной совокупности к числу ответивших единиц данного слоя. Если число неответов менее 20%, то они замещаются данными случайно отобранных единиц-доноров из соответствующего класса замещения в сформированном резервном (донорском) фонде.

Первые две проблемы - «неответы респондентов» и «нетипичные единицы совокупности» - могут быть решены в рамках традиционной одномерной выборки. Для позитивного разрешения остальных возможно три направления: 1) использование комбинации различных видов одномерной выборки; 2) построение многомерных выборок, которые дают возможность получить репрезентативные выборки небольшого объема, характеризующие объект наблюдения по большому количеству сильно варьируемых показателей и позволяющие адекватно экстраполировать выборочные данные на генеральную совокупность; 3) усложнение процедур оценивания путем использования различных схем взвешивания, GREG-оценок, калибрации.

**Одномерные выборки.** Механизм методов одномерного отбора детально разработан, способы оценивания и расчетов ошибок выборки охватывают весь спектр возможных подходов, используемые формулы приобрели классический характер. Наиболее часто применяются одномерные вероятностные выборки: случайный отбор без расслоения, с простым, пропорциональным

и оптимальным расслоением, а также серийная и механическая выборки.

*Случайная выборка без расслоения* обеспечивает строгую объективность отбора, составляет основу большинства более сложных методов отбора. Но возможности ее применения в чистом виде ограничены в связи с отсутствием равномерного распределения единиц по совокупности.

*Расслоенная случайная выборка*, при которой совокупность предварительно делится на группы (слои) по какому-либо признаку, а затем из каждой группы определенным образом производится отбор единиц, более точно отражает колеблемость признака в генеральной совокупности, позволяет строить относительно однородные выборочные группы, воспроизводит структуру совокупности, позволяет уменьшить объем выборки при сохранении необходимой точности, дает дополнительные преимущества в тех случаях, когда проблемы отбора в разных частях совокупности сильно различаются.

Каждый из видов распределения объема выборки по группам имеет свои особенности. При простом и непропорциональном размещении недостаточно точно воспроизводится структура совокупности, при пропорциональном существует разница между рабочей и точной долей отбора. Отбор с учетом колеблемости признака (оптимальное расслоение) дает наименьшую величину ошибки, но его практическое применение затрудняется отсутствием данных о величине  $\sigma_p$ , которую можно определить с помощью проведения пробных обследований. Кроме того, при строгом соблюдении принципов оптимального размещения возможна ситуация, когда теоретический объем выборочной группы превышает число единиц в слое генеральной совокупности ( $n_i > N_i$ ), что приводит к исчерпывающему охвату данного слоя, но не соответствует теоретической доле отбора.

*Механические выборки* удобно намечать и извлекать. В ряде случаев они дают более точные результаты, чем расслоенная выборка. Но точность механической выборки может быть низкой, если существует неожиданная периодичность признака в совокупности. Использование механического отбора целесообразно, если расслоение в исходной совокупности намечено очень слабо; возможен независимый механический отбор в каждой страте.

*Серийные выборки* основаны на случайном отборе из генеральной совокупности серий единиц

(гнезд, кластеров), внутри которых проводится или сплошное обследование, или случайная выборка. Чаще всего используются в тех случаях, когда генеральная совокупность состоит из разрозненных групп единиц (города, административные районы).

Выбор оптимального способа отбора для проведения определенного выборочного обследования зависит от двух укрупненных факторов - цели исследования и характера исходной информации, а именно: степени однородности изучаемой совокупности, размера генеральной совокупности, наличия естественных обособленных групп в совокупности, наличия дополнительной информации.

Для получения наиболее репрезентативной одномерной выборки целесообразно:

- применять комбинацию одномерных выборов различного вида; при наличии программного обеспечения для формирования исходной выборочной совокупности отбор единиц из регистра или массива данных сплошной отчетности целесообразно производить несколькими методами, используя случайный отбор без расслоения, простое случайное расслоение, пропорциональное расслоение, оптимальное расслоение, механический отбор. Сравнение ошибок (стандартных, предельных, фактических) и объема выборки, полученных при разных способах отбора, позволяет отобрать оптимальный вариант выборочной совокупности;

- строить взаимопроникающие выборки, позволяющие получить независимые оценки значений параметров генеральной совокупности.

**Многомерные выборки.** В отличие от одномерных методов отбора теория и практика многомерной выборки находится в начальной стадии разработки. Вопросы множественного расслоения, отбора по решетке исследуются в зарубежной экономической литературе начиная с 50 - 80-х годов прошлого века (Р.Д. Джессен, У.Г. Кокрен, П.Ч. Махаланобис, А. Паттерсон, Т. Далениус) [4-10]. В 1990-е годы появляются работы по многомерной выборке (Т. Чернышева, С. Степанов, Н. Бокун) [11-14].

Предлагаются разные способы многомерного отбора, то есть отбора единиц из генеральной совокупности по группе атрибутивных и количественных признаков, отличающиеся методом построения основы выборки, формой реализации принципа многомерности, способом построения

многомерного показателя. В зависимости от характера учета многомерности их можно объединить в три агрегированные группы:

1. Расслоение в независимых признаках;
2. Расслоение по композитному признаку;
3. Комбинированные методы (см. рисунок).



Рисунок. Методы многомерного отбора

*Расслоение в независимых признаках.* Признаки единицы наблюдения считаются абсолютно независимыми, расслоение по каждому из них проводится независимо, конечные слои определяются с учетом всех полученных независимых границ. Данный подход порождает большие количества конечных слоев слабой заполненности. При желании отразить в количестве слоев по размерности известные данные о структуре объекта исследователь сталкивается с проблемой ограничения объема выборки. Количество слоев, превышающее 7-8, фактически не дает эффекта сокращения ошибки, достигнув точки насыщения.

Для оптимизации многомерного расслоения и получения представительных выборок по системе независимых показателей целесообразно использовать методы типизации, оптимизации и комбинаторного анализа, что реализуется при отборе из типизированных основ выборки. При этом типизация включает стратификацию единиц отбора по ряду признаков, классификацию территориальных единиц и выделение количественно однородных структурных частей изучаемой совокупности. Методы оптимизации используются при отборе объектов наблюдения и размещении

объема выборки, в процессе которого: а) требуется представить все выделенные структурные части; б) обеспечить достаточность числа отобранных единиц для получения результатов в пределах заданной точности по совокупности в целом и по отдельным областям изучения. Методы комбинаторного анализа применяются для выделения типичных непересекающихся объектов и поиска оптимального варианта выборки.

Составление типизированных основ предполагает формирование одной или нескольких комбинационных таблиц, позволяющих получить упорядоченное размещение номеров единиц наблюдения по совокупности образованных графоклеток, отражающих структурные соотношения показателей и их призначной части. При построении выборочной совокупности учитывается число образованных блоков, заданная доля отбора, число групп по каждому из анализируемых признаков. Можно использовать: случайный, механический отбор, серийную выборку.

Отбор из двумерной или множественной основы выборки (Р.Д. Джессен, У.Г. Кокрен, П.Ч. Махаланобис) применяется при наличии двух и более основ выборки (домохозяйства и кварталы

города; фермерские хозяйства, справочник видов деятельности и предприятий). Предполагается организация исследуемой совокупности в две и более основы, причем каждый элемент данной совокупности связан хотя бы с одной единицей в каждой из возможных основ выборки [4, 6-9].

**Отбор по решетке** (Т. Далениус, Ф. Йейтс, Р.Д. Джессен, А. Паттерсон) используется, если генеральная совокупность может быть разделена на группы по нескольким показателям таким образом, что каждая «клетка» занята только одним элементом или гнездом. Отбор осуществляется по схеме «решетки»: если мы имеем квадрат со стороной  $p$ , разделенный на  $p^2$  единичных квадратов, то можно извлечь выборку объемом в  $p$  единичных квадратов так, чтобы каждый ряд и каждый столбец большого квадрата содержал один из отбираемых единичных квадратов. Ряды и столбцы большого квадрата могут соответствовать любой группировке по двум признакам, где число групп по каждому признаку одинаковое, а каждая подгруппа содержит по одной единице. Подобные схемы возможны и при группировке по трем и более признакам. Обычно выделяются три типа отбора по решетке: 1) случайные квадратные решетки; 2) прямоугольные решетки; 3) кубические случайные решетки [4, 5, 9].

**Расслоение по композитному признаку.** Строится дополнительный обобщающий показатель, учитывающий исходные признаки единиц наблюдения. По данному показателю осуществляется одномерное расслоение. Возможно два подхода к построению композитного признака: 1) использование эконометрической модели в виде определенной функции; 2) определение нормированного значения многомерного показателя. В первом случае существенным моментом является выбор вида функции, определение состава учитываемых исходных признаков и формы их участия в композитной формуле. Определение нормированного значения многомерного показателя в большей мере стандартизировано; при этом сам показатель может рассчитываться следующим образом:

$$\bar{P}_j = \frac{\sum P_j}{k}, \quad (1)$$

где  $\bar{P}_j$  - многомерный признак (признаки отбора) по  $i$ -ой единице наблюдения;  $P_j$  - нормированное значение  $j$ -го признака  $i$ -й единицы наблюдения (способы нормирования - см. таблицу 1);  $P_j = \frac{y_j}{y_j}$  - один из способов

нормирования;  $k$  - общее количество признаков;  $y_{ij}$  - индивидуальное значение  $j$ -го признака у  $i$ -й единицы наблюдения;  $y_j$  - среднее значение  $j$ -го признака.

Таблица 1

**Способы нормирования индивидуальных значений признака**

Варианты нормирования	Алгоритм расчета
1	$P_{ij1} = y_{ij} / \bar{y}_j$
2	$P_{ij2} = (y_{ij} - \bar{y}_j) / \sigma_{y_j}$ , наиболее часто используется, так как $\bar{p} = 0, D(p) = 1$
3	$P_{ij3} = y_{ij} / y'_j$
4	$P_{ij4} = y_{ij} / y_{jmax}$
5	$P_{ij5} = (y_{ij} - \bar{y}_j) / (y_{jmax} - y_{jmin})$

*Примечание:*  $\sigma$  - среднее квадратическое отклонение признака;  $y_j$  - нормированное (эталонное) значение признака;  $y_{jmax}, y_{jmin}$  - максимальное и минимальное значения признака.

Применение отбора по композитному признаку дает возможность использовать методы одномерной выборки, минуя методологические и организационные сложности многомерного отбора. К недостаткам расслоения по композитному признаку следует отнести невозможность одновременно учесть и числовые, и атрибутивные параметры. Кроме того, построение оптимизирующей функции применимо только к составу относительно однородных признаков (например, финансовых). Тем не менее использование данного метода представляется эффективным: в значении композитного признака «смешиваются» различные признаки, превращаясь в один агрегированный показатель - «вес предприятия».

Комбинированные методы сочетают в себе приемы многомерности, используемые как при расслоении в независимых признаках, так и при расслоении по композитному признаку. Модель многомерной выборки в виде специализированной нейронной сети рассматривает исследуемую совокупность в виде структурной модели групп - случайных величин - абстрактных типических единиц наблюдения, которым присущи количественные (численность занятых, доход, объем производства и т. д.) и атрибутивные признаки (вид деятельности, форма собственности и т. д.). Значения признаков конкретного предприятия колеблются в определенных границах, не охватывая весь спектр величин от минимума по всей совокупности до максимума. Критерии формирования элементов подобной сети, то есть нейронов, отражают сочетание объективных характеристик

исследуемых объектов и субъективных (предпочтения наблюдателя). В качестве данных для обучающих ситуаций используется информация об объектах наблюдения, взятая из статистического регистра. Любая из выборок, построенная по предлагаемой модели, будет предназначена для проведения наблюдения и оценки показателей генеральной совокупности в момент времени, отдаленный от момента получения исходной информации. Для отбора единиц в выборку может быть использован метод имитационного моделирования. Практическое использование модели нейронной сети осложняется двумя факторами: а) потенциальной неадекватностью обучающих ситуаций тем ситуациям, в которых создаваемая нейронная сеть будет действовать; б) необходимостью интегрирования нейронной сети и существующего программного обеспечения обработки статистических данных в одну систему, что в настоящее время не представляется возможным вследствие чрезмерной сложности и несовместимости разнородных программных продуктов [13]. Более простой вариант комбинационного подхода к формированию многомерной выборки - ее построение с помощью кластерного анализа [2, 3, 15], когда исследуемая совокупность делится с помощью иерархического агломеративного метода или метода  $k$ -средних на однородные группы. Внутри каждой из них выделяется основной (ведущий) признак, по которому осуществляется последующий случайный или механический отбор единиц в выборку.

Анализ существующих методов формирования многомерной выборки позволяет сделать вывод, что применение расслоения по независимым признакам может привести к чрезмерно большому числу групп в исследуемой совокупности. Расслоение по композитному признаку свободно от данного недостатка, но отсутствует гарантия того, что динамика колеблемости композитного признака и исходных показателей будут пропорциональны. Комбинированные методы достаточно сложны и трудоемки. Оптимальной представляется разработка такой модели выборки, которая предусматривала бы возможность выбора исследователем метода построения одномерной или многомерной выборки в зависимости от объема совокупности, числа и характера рассматриваемых признаков. В составе одномерных выборок имеют приоритет по эффективности оптимальное и простое случайное расслоение; из многомерных методов наиболее приемлемыми по степени надежности и доступ-

ности для пользователя представляются методы кластерного анализа.

**Методология проведения выборочного обследования микроорганизаций.** С учетом рассмотренных проблем, анализа возможностей применения одномерных и многомерных выборок автором разработаны принципы формирования выборочной совокупности микроорганизаций и статистического оценивания:

- определение совокупности количественных признаков для отбора, наиболее адекватно отражающих деятельность микроорганизаций;
- апробация нескольких способов отбора для решения одной и той же задачи, критерий наиболее точных оценок параметров генеральной совокупности - минимальная величина предельной и стандартной ошибки при фиксированном объеме выборки;
- использование сочетания одномерных и многомерных методов отбора;
- приоритет оптимального, пропорционального и простого случайного расслоения в составе одномерных выборок, а среди многомерных - формируемым на основе кластерного анализа;
- актуализация основы выборки путем перевзвешивания или импутации (замещения) единиц наблюдения;
- формирование резервного (донорского) фонда единиц наблюдения, имеющего структуру выборочной совокупности;
- использование в процессе экстраполяции нескольких вариантов оценивания в зависимости от объема выборки, числа слоев, колеблемости данных.

В соответствии с рассматриваемыми принципами под руководством автора в течение 2014-2015 гг. разработана методология проведения выборочного обследования микроорганизаций, включая формирование основы выборки, состава и структуры выборочной совокупности, статистическое оценивание [15-20].

**Формирование основы выборки.** В качестве основы выборки рассматривается совокупность респондентов по форме № 1-МП (микро), составляющая свыше 80 тыс. единиц, а также фермерские хозяйства (свыше 2 тыс. единиц). При формировании генеральной совокупности учитывалась структура ответивших и не ответивших респондентов.

В составе организаций, не отчитавшихся за базисные годы (2011-2013 гг.), выделены две группы:

Доверительные вероятности и значения  $t$ -критерия  
( $n \geq 30$ )

Параметры	Наиболее часто употребляемое значение				
	1,00	1,96	2,00	2,58	3,00
$t$ -критерий кратности ошибок	1,00	1,96	2,00	2,58	3,00
Доверительная вероятность ( $P = \Phi(t)$ )	0,683	0,95	0,954	0,990	0,997

1) ответившие, но не отчитавшиеся, потому что не вели в отчетном периоде экономическую деятельность; практически все показатели по форме № 1-МП (микро) могут быть приравнены к нулю; 2) действительно не ответившие единицы; возможные причины неответов: находятся в стадии ликвидации, адресат не найден, форс-мажорные обстоятельства и т. д. В результате процент неответов по республике в целом - 9-13%, по областям - от 6 до 12%, по г. Минску - значительно выше (18-20%, в 2013 г. - 11%) [15, с. 21-23].

Генеральная совокупность формируется по каждой области и включает в себя по два массива данных: респонденты формы № 1-МП (микро) и фермерские хозяйства соответственно с переменной и постоянной (100%) долями отбора. Из полученных генеральных совокупностей может производиться отбор как в целом, так и по видам деятельности.

*Формирование выборочной совокупности* предполагает применение нескольких методов отбора, сравнение результатов и обоснование наиболее оптимального вида выборки (собственно-случайный отбор, простое, пропорциональное, оптимальное расслоение, многомерная выборка). Предусматривается следующая последовательность действий:

1. Выделяется генеральная совокупность. Предприятия, у которых значение признака равно нулю, удаляются из совокупности. Определяется число единиц генеральной совокупности  $N$ .

2. Из совокупности показателей ( $X_1, X_2, \dots, X_k$ ) пользователем выделяется признак отбора  $x_j$ . По отобранному показателю генеральная совокупность делится на однородные группы (строится ранжированный ряд, пользователь самостоятельно выбирает и задает число и границы групп).

3. По каждому имеющемуся экономическому показателю ( $X_1, X_2, \dots, X_k$ ), включая признак отбора, в целом по генеральной совокупности и в разрезе групп рассчитываются средние, суммарные значения и показатели вариации (дисперсия, коэффициенты вариации).

4. Пользователь задает допустимую ошибку выборки в %  $h_g$  (2; 3; 4; 5; 6; 7; 8; 9; 10; 12; 15%); задает значение  $t$ -критерия; с этой целью отбирается величина вероятности  $P$  и соответствующая ей величина коэффициента доверия (см. таблицу 2). Так, если  $P = 0,95$ ,  $t = 1,96$ , то величина предельной ошибки устанавливается с вероятностью, равной 0,95.

Для выборок небольшого объема ( $n < 30$ ) значения  $t$  и соответствующие вероятности  $P = \Phi(t)$  при разном числе степеней свободы приводятся в таблицах распределения Стьюдента.

5. Задается начальный объем выборки ( $n = 0,05N$ : предельная величина численности выборки  $n = 0,8N$ ). В случае расслоенных выборок заданный объем выборки  $n$  распределяется по группам  $n_i$  пропорционально численности групп в генеральной совокупности (пропорциональное расслоение), или пропорционально колеблемости (оптимальное расслоение):

$$n_i = \frac{N_i}{N} n; \quad (2)$$

$$n_i = \frac{\sigma^2_i N_i}{\sum \sigma^2_i N_i} n, \quad (3)$$

где  $n_i$  - численность  $i$ -й группы в выборке;  $N_i$  - численность  $i$ -й группы в генеральной совокупности;  $\sigma_i$  - среднее квадратическое отклонение в  $i$ -й группе генеральной совокупности.

Затем из соответствующей группы генеральной совокупности отбирается случайным способом число организаций, равное  $n_i$ .

При простом расслоении последовательно по каждой из выделенных групп генеральной совокупности задается допустимая ошибка в процентах (2; 5; 10; 12; 15; 20; 30), обозначаемая через  $h_{gi}$  и начальный объем выборки; случайным методом из группы  $N_i$  отбирается начальное число предприятий ( $n_i$ ), равное начальному объему выборки, исчисляются предельные (или стандартные) ошибки выборки и сравниваются с допустимой ( $h_{gi}$ ). Если предельная ошибка выборки превышает допустимую, то или меняется шкала и состав групп, или объем выборки увеличивается, а процесс поиска оптимальной выборки по группе генеральной совокупности продолжается. Кроме того, необходимо предусмотреть случай, когда пользователь выбирает вариант выборки, где ошибка в какой-либо группе превышает допустимую (например, в многочисленной группе ошибка является допустимой, в малочисленной -

превышает допустимую), то есть осуществляется просмотр промежуточных результатов расчета.

Для полученной выборочной совокупности исчисляется предельная (стандартная или фактическая) ошибка в абсолютном ( $\Delta x_i$ ) и относительном ( $h_{x_i}$ ) выражении:

$$\Delta x_i = \pm t \frac{1}{N} \sqrt{\sum \frac{\sigma_{x_i}^2}{n_i} N_i^2 \left(1 - \frac{n_i}{N_i}\right)} \text{ или} \\ \Delta x_i = \pm t \sqrt{\sum \left(\frac{N_i}{N}\right)^2 \left(1 - \frac{n_i}{N_i}\right) \frac{\sigma_i^2}{n_i}}; \quad (4)$$

$$h_{x_i} = \pm t \frac{1}{N} \sqrt{\sum \frac{V_{x_i}^2}{n_i} N_i^2 \left(1 - \frac{n_i}{N_i}\right)} \text{ или} \\ h_{x_i} = \pm t \sqrt{\sum \left(\frac{N_i}{N}\right)^2 \left(1 - \frac{n_i}{N_i}\right) \frac{V_i^2}{n_i}}. \quad (5)$$

Ошибки суммарных значений исчисляются по формулам (6) и (7):

$$\mu_{\Sigma X} = \sqrt{\sum \frac{N_i}{n_i} (N_i - n_i) \cdot \sigma_i^2}; \quad (6)$$

$$h_{\Sigma X} = \frac{\sqrt{\sum \frac{N_i}{n_i} (N_i - n_i) \cdot \sigma_i^2}}{\sum X_i}. \quad (7)$$

Фактические ошибки находятся как разность генерального и экстраполированного выборочного показателя. При расчете итоговой величины экстраполированного суммарного признака по совокупности в целом расчет производится путем суммирования групповых экстраполированных значений:

$$\sum X_s = \sum \sum X_{si} = \frac{N_i}{n_i} \sum x_i, \quad (8)$$

где  $\frac{N_i}{n_i}$  - групповой вес (коэффициент распространения).

Полученные ошибки (предельные, стандартные, фактические) сравниваются с допустимой. Если предельная (фактическая, стандартная) ошибка не выше допустимой, то выборка признается приемлемой, если выше - продолжается поиск оптимального объема выборки.

Критериями предпочтительного применения расслоенных одномерных выборок могут служить: а) относительная вариация показателей (коэффициенты вариации  $V_{x_j}$ ) не превышает 50-60%; б) в системе изучаемых показателей встречается один высоковарируемый показатель ( $V_{x_j}$  превышает 150-200%), что позволяет применить по нему

типизацию; в) в системе изучаемых показателей имеется один, корреляционно связанный с рядом других, что позволяет применить его для формирования выборки с целью получения репрезентативных данных по другим показателям; г) имеется дополнительная информация по ряду признаков, корреляционно связанных с изучаемыми, и ее можно использовать для повышения точности оценок статистических показателей. В противном случае применяется многомерная выборка (МВ).

По результатам пробных обследований и опыту проведения отраслевых выборок малых предприятий можно сделать вывод, что чаще всего МВ используется при наличии одного из трех условий: а) показатели вариации превышают 100% ( $V_{x_j} > 100\%$ ); б) объекты наблюдения неоднородны по многим показателям (высокие  $V_{x_j}$ ); в) небольшой размер генеральной совокупности (верхний предел - 450-500, нижний оценить трудно).

Многомерный отбор предлагается осуществлять с помощью методов кластерного анализа: исследуемая совокупность делится с помощью агломеративного иерархического метода на однородные группы (кластеры); внутри каждой из них пользователем выделяется ведущий признак, по которому осуществляется последующий случайный отбор единиц в выборку. Если по ведущему показателю коэффициент вариации ( $V_{x_j}$ ) превышает 50%, то возможно дополнительное расслоение внутри кластера. По каждому признаку исчисляется стандартная (предельная, фактическая) ошибка выборки. Если она превышает допустимые границы, то возможно три способа ее снижения: 1) увеличение объема выборки в кластере; 2) дополнительное расслоение предприятий в кластере по ведущему признаку; 3) повторение процесса кластеризации, но с увеличением числа шагов.

**Статистическое оценивание.** Полученные в результате годового выборочного обследования показатели по форме № 1-МП (микро), в состав которых входят  $k$  переменных ( $x_1, \dots, x_j, \dots, x_k$ ): средняя численность работающих; фонд заработной платы; объем подрядных работ, выполненных своими силами; использовано инвестиций в основной капитал; выручка от реализации и т. д. - предложено экстраполировать на генеральную совокупность с помощью одного из трех методов: 1) традиционные оценки Горвица-Томпсона (групповые веса, или коэффициенты распространения); 2) калибрация (обобщенные

регрессионные оценки GREG); 3) калибрация (синтетические оценки SYN).

Оценки Горвица-Томпсона (HT-оценка) выступают основным методом взвешивания. Расчет веса микроорганизаций осуществляется по данным предварительно сформированной выборочной совокупности и на стадии формирования выборки. Выборочные совокупности строятся в пределах областей и г. Минска отдельно по каждому виду деятельности.

В качестве веса предприятия, или коэффициента досчета, выступает показатель, обратный групповой доле отбора предприятий  $l$ -го вида деятельности  $m$ -й области:

$$w_i = \frac{N_{ilm}}{n_{ilm}}, \quad (9)$$

где  $N_{ilm}$  - число предприятий  $i$ -й группы (по ведущему признаку отбора, например выручка от реализации)  $l$ -го вида деятельности  $m$ -й области;  $n_{ilm}$  - число предприятий  $i$ -й выборочной группы (по признаку отбора)  $l$ -го вида деятельности  $m$ -й области.

Для многомерной выборки вес предприятия имеет вид:

$$w_i = \frac{N_{hrim}}{n_{hrim}}, \quad (10)$$

где  $h$  - число групп в  $r$ -м кластере;  $r$  - число кластеров по  $l$ -му виду деятельности в  $m$ -м регионе.

**Калибрация (обобщенные регрессионные оценки GREG).** При условии наличия высоких стандартных или предельных ошибок оценки Горвица-Томпсона могут быть уточнены с помощью специальных техник взвешивания (постстратификация, итеративное взвешивание, калибрация и т. д.). В последние десятилетия особое внимание уделяется процедурам калибрации (calibration, Ж.К. Девиль, К.Э. Сэндел), направленным на расчет скорректированных (калиброванных) весов выборки при соблюдении условий: 1) сумма весов первоначального плана выборки и сумма калиброванных весов равны; 2) калибрация должна способствовать минимизации отклонений между базовым выборочным весом  $w_i$  ( $w_i = k_p \cdot k_i = \frac{N_i}{n_i}$ ) и корректировочным весом  $d_i$  ( $|w_i - d_i| \rightarrow \min$ ), чаще всего минимизируется функция расстояния типа хи-квадрат ( $\sum \frac{(w_i - d_i)^2}{d_i \cdot q_i}$ , где  $q_i$  - неотрицательные константы, с помощью которых задаются веса объектов, не связанные с выборочным планом); 3) наличие дополнительной информации по вспомо-

гательным переменным, известной для всей генеральной совокупности [21-23].

Наиболее распространенные оценки для весов Горвица-Томпсона - обобщенная регрессионная оценка (GREG) и синтетическая оценка (SYN).

Обобщенная регрессионная оценка GREG суммарного значения  $x_j$  по генеральной совокупности имеет вид:

$$\hat{X}_{jGREG} = \hat{X}_{jHT} + \hat{\beta}(Y - \hat{Y}_{HT}), \quad (11)$$

где  $\hat{X}_{jHT}$  - оценка суммарного значения  $x_j$  на основе базового веса предприятия (оценка Горвица-Томпсона);  $Y$  - суммарное значение вспомогательной переменной по генеральной совокупности;  $\hat{Y}_{HT}$  - оценка суммарного значения вспомогательной переменной  $Y$  на основе базового веса (оценка Горвица-Томпсона);  $\hat{\beta}$  - коэффициент регрессии,  $\hat{\beta} = \frac{\sum w_i q_i x_{ij} y_i}{\sum w_i q_i y_i^2}$ .

Подставляя в выражение (11) значения  $\hat{X}_{jHT}$  и  $\hat{Y}_{HT}$ , получаем:

$$\hat{X}_{jGREG} = \sum w_i x_{ij} + \hat{\beta}(Y - \sum w_i y_i). \quad (12)$$

**Калибрация (синтетическая оценка SYN)** основана на замещении признака  $x_j$  вспомогательной переменной  $Y$ , прогнозированием  $x_j$  на основе уравнения взаимосвязи:

$$\hat{X}_{jSYN} = \sum \hat{x}_{ij}, \quad (13)$$

где  $\hat{x}_{ij} = f(y_{ij} \hat{\beta})$ ;  $\hat{x}_{ij} = \sum \hat{\beta} \cdot y_i$ ,  $\hat{\beta} = \left( \sum \frac{y_i y_i'}{\sigma_y^2 k_i} \right)^{-1} \cdot \sum \frac{y_i x_i}{\sigma_y^2 k_i}$ ,  $k_i = w_i$  - вес по оценке Горвица-Томпсона.

Оценку SYN можно применять в сочетании с HT-оценкой для малых совокупностей.

По результатам расчета трех оценок по каждой из обследованных переменных выбирают ту, которая дает наименьшую оценку стандартной (предельной) ошибки.

В качестве иллюстрации применения различных методов оценивания рассмотрены показатели расслоенной выборки по виду деятельности «Текстильное и швейное производство», Брестская область (см. таблицу 3).

По результатам расчетов HT-оценки получили:

$$\hat{X}_{HT} = \sum_{i=1}^h k_i x_i = \frac{31}{7} \cdot 133 + \frac{54}{11} \cdot 1239 + \frac{75}{15} \cdot 6578 + \frac{14}{3} \cdot 4640 + \frac{6}{2} \cdot 11472 = 95895,79 \text{ млн рублей.}$$

Фактическая ошибка в абсолютном выражении:  $\Delta \Phi_{HT} = |X - \hat{X}_{HT}| = |88715 - 95895,79| = 7180$  млн рублей.

Показатели выборочной совокупности «Швейное и текстильное производство». Брестская область, 2013 г.  
 Расслоенная выборка (пропорциональное расслоение)

Показатель	Группы по выручке от реализации (X), млн рублей	Объем совокупности, единиц		Доля отбора $d = \frac{n}{N}$	Суммарное значение признака		Дисперсия		Коэффициент вариации, %	
		генеральной, N	выборочной, n		генеральная совокупность X	выборочная совокупность $\sum x_i$	генеральная совокупность $\sigma_x^2$	выборочная совокупность $S_x^2$	генеральная совокупность $V_x$	выборочная совокупность $v_x$
Выручка от реализации (X)	до 50	31	7	0,225	712	133	252,4	174,6	69,2	69,5
	50-200	54	11	0,20	6640	1293	1742,9	1561,2	33,95	33,62
	200-1000	65	15	0,20	34435	6578	40959,2	30380,5	44,1	39,8
	1000-2000	14	3	0,21	18353	4640	39401,9	27671,3	15,14	10,8
	2000 и более	6	2	0,33	28575	11472	7417021,58	13286025	57,2	63,5
	Итого	180	38	0,21	88715	24116	1007191	2311616	203,6	239,6

Фактическая ошибка в относительном выражении:

$$\Delta\Phi_{\text{н.о.}} = \left| \frac{X - \hat{X}_{\text{н.т.}}}{X} \right| = \left| \frac{88715 - 95895,79}{88715} \right| = \frac{7180,8}{88715} = 0,081 \text{ (8,1\%)}$$

GREG-оценка. В качестве вспомогательной переменной избран показатель средней численности работников (y), как характеризующейся

меньшей колеблемостью, чем стоимостные показатели (см. таблицу 4).

Калибровочные веса имеют вид:

$$d_i = k_i + \frac{k_i y_i}{\sum k_i \cdot y_i^2} \cdot (Y - \sum k_i y_i) \quad (14)$$

Например, для предприятия №1 калибровочный вес составит:

$$d_1 = 3_i + \frac{3 \cdot 8}{15900,06} \cdot (1270 - 1374,4) = 2,842415.$$

Таблица 4

Исходные и расчетные данные для экстраполяции путем использования GREG- и SYN-оценок.  
 (Расчет по 38 отобраным организациям). Пропорциональное расслоение

№ п/п	Выручка от реализации, $x_i$	Среднесписочная численность работающих, $y_i$	Базовый вес, $k$	$k_i x_i$	$k_i x_i y_i$	$k_i y_i^2$	$k_i y_i$	$y_i^2$	$x_i^2$
1	9381	8	3	28143	225144	192	24	64	88003161
2	24	2	4,4285714	106,2857	212,5714	17,71429	8,857143	4	576
3	626	21	5	3130	65730	2205	105	441	391876
4	390	12	5	1950	23400	720	60	144	152100
5	411	12	5	2055	24660	720	60	144	168921
6	109	3	4,9090909	535,0909	1605,273	44,18182	14,72727	9	11881
7	476	14	5	2380	33320	980	70	196	226576
8	7	0	4,4285714	31	0	0	0	0	49
9	150	4	4,9090909	736,3636	2945,455	78,54545	19,63636	16	22500
10	1612	13	4,6666666	7522,667	97794,67	788,6667	60,66667	169	2598544
11	219	10	5	1095	10950	500	50	100	47961
12	465	11	5	2325	25575	605	55	121	216225
...	...	...	...	...	...	...	...	...	...
38	143	4	4,9090909	702	2808	78,54545	19,63636	16	20449
Итого	24116	285		95895,79	1005171	15717,06	1369,437	3239	103149334

Оценка суммарного значения выручки составляет:

$$\hat{X}_{GREG} = \sum_{i=1}^h k_i x_i + \frac{\sum_{i=1}^h k_i x_i y_i}{\sum_{i=1}^h k_i y_i^2} \cdot (Y - \sum_{i=1}^h k_i y_i) =$$

$$= 95895,79 + \frac{1005171}{15717,06} (1270 - 1369,4) = 89538,75$$

млн рублей.

Фактическая ошибка:

$$\Delta\Phi_{GREG} = \left| X - \hat{X}_{GREG} \right| = \left| 88715 - 89538,75 \right| = 823,75$$

млн рублей, или 0,93%.

SYN-оценка:

$$\hat{X}_{SYN} = \sum_{i=1}^h k_i y_i \cdot \frac{\sum_{i=1}^h k_i x_i y_i}{\sum_{i=1}^h k_i y_i^2} = 1369 \cdot \frac{1005171}{15717,06} = 87553,2$$

млн рублей.

Фактическая ошибка:

$$\Delta\Phi_{SYN} = |X - \hat{X}_{SYN}| = |88715 - 87553,2| = 1161,8$$

млн рублей, или 1,3%.

Очевидно, что наилучшие оценки дал метод GREG-оценивания (фактическая ошибка составила 0,9%).

В многомерной выборке оценивание в большинстве случаев проводится на основе метода НТ-оценок. В кластерах без расслоения в качестве аналога групп выступают кластеры; в кластерах с расслоением учитывается еще дополнительно вес группы:  $k_h = \frac{N_h}{n_h}$ , где  $N_h$  - число организаций в  $h$ -й группе  $k$ -го кластера по генеральной совокупности;  $n_h$  - по выборочной совокупности.

В соответствии с предложенной методологией Национальным статистическим комитетом Республики Беларусь в ноябре 2014 г. было проведено пробное выборочное обследование микроорганизаций, а начиная с 2015 г. оно проводится уже на регулярной основе. Полученные результаты апробации первой версии методологического и программного обеспечения выборки микроорганизаций показали:

- основные проблемы проведения обследований связаны с использованием различных схем взвешивания, необходимостью оценки группы индикаторов, наличием малых областей оценивания, локализацией выборки;

- примерная доля отбора - 15-20%; относительная ошибка выборки по регионам - 2-3%, по видам деятельности - 4-5%, по отдельным видам деятельности с небольшим объемом исследуемой совокупности - 8-9%; общая доля отбора по Республике Беларусь - 21%, по г. Минску - 13%; по таким видам деятельности, как торговля, транспорт - 10%, строительство - 5-10%, некоторые виды промышленного производства - до 30-40%.

**Заключение.** Предлагаемая методология выборочных обследований, основанная на применении комбинации одномерных и многомерных выборок, различных методов оценивания, включая GREG- и SYN-оценки, позволяет получить достаточно надежную информацию по группе показателей, отражающих деятельность

микроорганизаций: занятость, фонд заработной платы, прибыль, выручку от реализации, объем производства. Уточненные и экстраполированные показатели выборочной совокупности дают возможность полнее оценить ненаблюдаемые виды экономической деятельности в части неучитываемых мелких субъектов хозяйствования. Тем не менее стандартные ошибки по отдельным индикаторам (например, выручке, инвестициям) в разрезе видов деятельности и областей могут быть относительно высокими (до 10-16%). С целью повышения их репрезентативности целесообразно усложнить процессы оценивания путем применения дополнительных методов калибровки, усовершенствовать процедуры ежегодной корректировки состава выборочной совокупности.

## Литература

1. Организация системы статистического наблюдения состояния и развития малого предпринимательства: Доклад Росстата. М., 2004.
2. Разработать методологическое и программное обеспечение построения многомерной выборочной совокупности малых предприятий: Отчет по НИР, № гос.регр. 2005920. Мн.: НИИ статистики, 2005.
3. **Бокун Н.Ч.** Проблемы построения многомерной выборки в сфере розничной торговли // Вопросы статистики. 2010. № 3. С. 52-60.
4. **Джессен Р.Д.** Методы статистических обследований. М.: Финансы и статистика, 1985. 478 с.
5. **Йейтс Ф.** Выборочный метод в переписях и обследованиях. М.: Статистика, 1965. 435 с.
6. **Кокрен У.Г.** Методы выборочного обследования. М.: Статистика, 1976. 440 с.
7. **Махаланобис П.Ч.** Выборочные обследования в Индии. М.: Госстатиздат, 1958. 82 с.
8. **Cochran W.G.** Sampling Techniques. (2nd ed.). New York, John Wiley and Sons, Inc. 1953, 1963.
9. **Dalenius T.** Sampling in Sweden: Contributions to the method and theories of sample survey practice. Stockholm: Almqvist and Wiksell. 1957.
10. **Yates F.** Sampling Methods for Censuses and Surveys. London, Chares Griffin and Sons. 1971.
11. **Чернышева Т.** Выборочные наблюдения в социально-экономических исследованиях и основные этапы их подготовки // Вопросы статистики. 1996. № 11. С. 12-19.
12. **Чернышева Т., Муханова О.** Проблемы формирования выборочной совокупности домашних хозяйств, подлежащих бюджетному обследованию // Вопросы статистики. 1997. № 5. С. 16-26.
13. **Степанов С.В.** Нейронная интерпретация выборки для структурного статистического моделирования совокупностей предприятий // Вопросы статистики. 2004. № 6. С. 21-32.

14. **Бокун Н.Ч., Чернышева Т.М.** Методы выборочных обследований: Уч.-справ. пособие. Минск: Минстат, НИИ статистики, 1997. 416 с.
15. Алгоритм формирования выборочной совокупности микроорганизаций: отчет по НИР; рук. темы Н.Ч. Бокун. Минск, 2014.
16. Алгоритм экстраполяции выборочных показателей на генеральную совокупность микроорганизаций по Республике Беларусь: отчет по НИР; рук. темы Н.Ч. Бокун. Минск, 2014.
17. **Бокун Н.Ч.** Выборочные обследования малого бизнеса в Беларуси // Бухгалтерский учет и анализ. 2015. № 10. С. 36-47.
18. **Bokun N.** Micro-entities sample survey: problems of design, formation and usage // Workshop of Baltic-Nordic-Ukrainian network on Survey Statistics, Tallinn, Estonia, August 25-28, 2014. P. 25-31.
19. **Bokun N.** Micro-entities and Small Enterprises surveys in Belarus // Computer Data Analysis and Modelling: Theoretical and Applied Stochastics: Proc. of the Eleventh. Intern. Conf.; Minsk, Sept. 6-10, 2016. P. 240-245.
20. **Bokun N.** Small business surveys in Belarus // Workshop on Survey Statistics and Methodology, Vilnius, Lithuania, August, 21-24, 2017. P. 120-132.
21. **Devill J.C., Sarndal C.E.** Calibration estimators in survey sampling // Journal of the American Statistical Association. 1992. Vol. 87. Jun. P. 376-382.
22. **Sarndal C.E., Swensson B., Wretman J.** Model assisted survey sampling. New York: Springer-Verlag, 1992.
23. **Sarndal C.E., Lundstrom S.** Estimation in Surveys with Nonresponse. New York: John Wiley, 2005.

## MICRO-ENTITIES SAMPLE SURVEY METHODOLOGY

*Natalia Ch. Bokun*

*Author affiliation:* Belarus State Economic University (Minsk, Republic of Belarus). E-mail: nataliabokun@rambler.ru.

This article presents sampling methodology for micro-entities survey by economic activity, problems associated with introducing sample surveys of enterprises into official statistical practice (on the example of the CIS countries). There is an analysis of the potential use of random samples without stratification, simple, proportional, optimal allocation, systematic, cluster samples. The following approaches to formation of multivariate sample design are summarized: typified, multiple sampling frames, stratification by a composite variable, combined methods. The author considers the sample of micro-enterprises that includes selection algorithm for enterprises, determining optimal sample size by economic activities and regions. HT-, GREG-, and SIN- estimators are recommended for statistical weighing. It is proposed to use a combination of univariate and multivariate samples, multivariant statistical weighting.

*Keywords:* sample, weighting, estimation, micro-enterprises, multivariate sample, optimal allocation, stratified selection.

*JEL:* C83, O12.

## References

1. Organizatsiya sistemy statisticheskogo nablyudeniya sostoyaniya i razvitiya malogo predprinimatel'stva: Doklad Rosstata [Organization of the system of statistical observation of the state and development of small business: Report of Rosstat]. Moscow, 2004. (In Russ.).
2. Razrabotat' metodologicheskoe i programmnoe obespechenie postroeniya mnogomernoi vyborochnoi sovokupnosti malykh predpriyatii: Otchet po NIR, № gos. registr. 2005920 [To develop methodological and software support for constructing a multidimensional sample of small enterprises: Report on R&D work, state registration number 2005920]. Minsk, Research Institute of Statistics, 2005. (In Russ.).
3. **Bokun N.Ch.** Problemy postroeniya mnogomernoi vyborki v sfere roznichnoi trgovli [Problems of multidimensional sample design in retail trade]. *Voprosy statistiki*, 2010, no.3, pp. 52-60. (In Russ.).
4. **Jessen R.D.** Statistical survey techniques. Wiley, 1978. 520 p. (Russ. ed.: Dzhessen R. Metody statisticheskikh obsledovaniy. Moscow, Finansy i statistika Publ., 1985, 478 p.)

5. **Yates F.** Sampling methods for censuses and surveys. C. Griffin, 1949. 318 p. (Russ. ed.: Ieits F. Vyborochnyi metod v perepisyakh i obsledovaniyakh. Moscow, Statistika Publ., 1965, 435 p.)
6. **Cochran W.G.** Sampling techniques, John Wiley & Sons Ltd; 2nd edition, December 1963. (Russ. ed.: Kokren U. Metody vyborochnogo obsledovaniya. Moscow, Statistika Publ., 1976, 440 p.)
7. **Makhalanobis P.Ch.** Vyborochnye obsledovaniya v Indii [Sample surveys in India]. Moscow, Gosstatizdat Publ., 1958, 82 p. (In Russ.).
8. **Cochran W.G.** Sampling Techniques. (2nd ed.). New York, John Wiley and Sons, Inc. 1953, 1963.
9. **Dalenius T.** Sampling in Sweden: Contributions to the method and theories of sample survey practice. Stockholm: Almqvist and Wiksell. 1957.
10. **Yates F.** Sampling Methods for Censuses and Surveys. London, Chares Griffin and Sons. 1971.
11. **Chernysheva T.** Vyborochnye nablyudeniya v sotsial'no-ekonomicheskikh issledovaniyakh i osnovnye etapy ikh podgotovki [Sample surveys in socio-economic studies and the main stages of survey preparation]. *Voprosy statistiki*, 1996, no. 11. pp. 12-19. (In Russ.).
12. **Chernysheva T., Mukhanova O.** Problemy formirovaniya vyborochnoi sovokupnosti domashnikh khozyaistv, podlezhashchikh byudzhethnomu obsledovaniyu [Problems of making a sample of households that are subject to the budget survey]. *Voprosy statistiki*, 1997, no. 5, pp. 16-26. (In Russ.).
13. **Stepanov S.V.** Neironnaya interpretatsiya vyborki dlya strukturnogo statisticheskogo modelirovaniya sovokupnostei predpriyatii [Neuronic interpretation of a sample for structural statistical modeling of populations of enterprises]. *Voprosy statistiki*. 2004, no. 6, pp. 21-32. (In Russ.).
14. **Bokun N.Ch.** Metody vyborochnykh obsledovaniy: Uch.-sprav. Posobie [Sample survey methods. Training and reference manual]. Minsk, Minstat, Research Institute of Statistics, 1997. 416 p. (In Russ.).
15. Algoritm formirovaniya vyborochnoi sovokupnosti mikroorganizatsii: otchet po NIR; ruk. temy N.Ch. Bokun [Algorithm for the formation of a sample of micro-entities. Report on R&D work; project coordinator Bokun N.Ch.]. Minsk, 2014. (In Russ.).
16. Algoritm ekstrapolyatsii vyborochnykh pokazatelei na general'nyu sovokupnost' mikroorganizatsii po Respublike Belarus': otchet po NIR; ruk. temy N.Ch. Bokun [Extrapolation algorithm of sample indicators to the general set of micro-entities in the Republic of Belarus. Report on R&D work; project coordinator Bokun N.Ch.]. Minsk, 2014. (In Russ.).
17. **Bokun N.Ch.** Vyborochnye obsledovaniya malogo biznesa v Belarusi. [Sample surveys of small business in Belarus]. *Accounting and analysis*, 2015, no. 10, pp. 36-47. (In Russ.).
18. **Bokun N.** Micro-entities sample survey: problems of design, formation and usage. *Workshop of Baltic-Nordic-Ukrainian network on Survey Statistics, Tallinn*, Estonia, August 25-28, 2014. P. 25-31.
19. **Bokun N.** Micro-entities and Small Enterprises surveys in Belarus. Computer Data Analysis and Modelling: Theoretical and Applied Stochastics: Proc. of the Eleventh. Intern. Conf.; Minsk, Sept. 6-10, 2016. P. 240-245.
20. **Bokun N.** Small business surveys in Belarus. *Workshop on Survey Statistics and Methodology*, Vilnius, Lithuania, August, 21-24, 2017. P. 120-132.
21. **Devill J.C., Sarndal C.E.** Calibration estimators in survey sampling. *Journal of the American Statistical Association*. Vol. 87. Jun. 1992, pp. 376-382.
22. **Sarndal C.E.** Model assisted survey sampling / C.E. Sarndal, B. Swensson, J. Wretman. New York: Springer-Verlag, 1992.
23. **Sarndal C.E., Lundstrom S.** Estimation in Surveys with Nonresponse. New York: John Wiley, 2005.